

(11) Número de Publicação: **PT 975755 E**

(51) Classificação Internacional:

C12N 15/12 (2006.01) **C07K 14/47** (2006.01)
C07K 16/18 (2006.01) **G01N 33/50** (2006.01)
G01N 33/53 (2006.01) **C12Q 1/68** (2006.01)

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

(22) Data de pedido: **1998.04.16**

(30) Prioridade(s): **1997.04.16 US 843704**
1997.04.17 US 842898
1998.01.15 US 71589 P
1998.01.20 US 9802

(43) Data de publicação do pedido: **2000.02.02**

(45) Data e BPI da concessão: **2007.02.14**
023/2007

(73) Titular(es):

MILLENNIUM PHARMACEUTICALS, INC.
40 LANDSDOWNE STREET CAMBRIDGE, MA
02139 US

(72) Inventor(es):

SEAN A. MCCARTHY US

(74) Mandatário:

ANTÓNIO JOÃO COIMBRA DA CUNHA FERREIRA
RUA DAS FLORES, Nº 74, 4º AND 1249-235 LISBOA PT

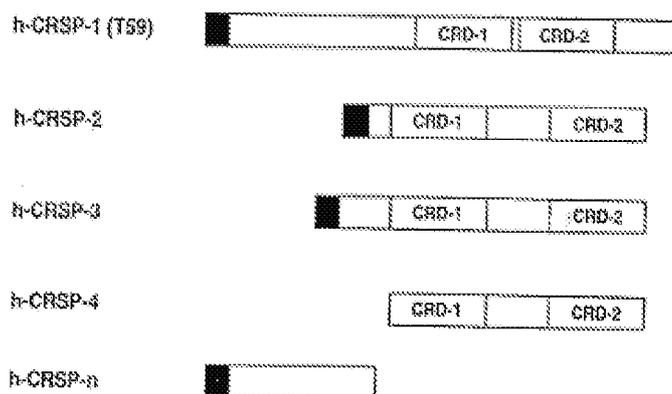
(54) Epígrafe: **PROTEÍNAS CRSP (PROTEÍNAS SEGREGADAS RICAS EM CISTEÍNAS),
MOLÉCULAS DE ÁCIDO NUCLEICO QUE AS CODIFICAM E SUAS UTILIZAÇÕES**

(57) Resumo:

RESUMO

**"Proteínas CRSP (proteínas segregadas ricas em cisteínas),
moléculas de ácido nucleico que as codificam e suas
utilizações"**

A Família da CRSP Humana



São revelados polipéptidos, proteínas e moléculas de ácido nucleico de CRSP (Proteínas Segregadas Ricas em cisteínas). Em adição às proteínas CRSP de comprimento completo isoladas, a invenção proporciona ainda proteínas de fusão de CRSP isoladas, péptidos antigénicos e anticorpos anti-CRSP. A invenção proporciona também moléculas de ácido nucleico de CRSP, vectores de expressão recombinantes contendo uma molécula de ácido nucleico da invenção, células hospedeiras nas quais os vectores de expressão foram introduzidos e animais transgénicos não humanos nos quais um gene de CRSP foi introduzido ou corrompido. São igualmente proporcionados métodos de diagnóstico, de rastreio e terapêuticos que utilizam composições da invenção.

DESCRIÇÃO

"Proteínas CRSP (proteínas segregadas ricas em cisteínas), moléculas de ácido nucleico que as codificam e suas utilizações"

Antecedentes da invenção

As proteínas segregadas desempenham um papel integral na formação, diferenciação e manutenção das células nos organismos multicelulares. Por exemplo, sabe-se na especialidade que as proteínas secretórias estão envolvidas na sinalização entre células que estão em contacto directo. Estas moléculas de sinalização segregadas são particularmente importantes no desenvolvimento dos tecidos dos vertebrados durante a embriogénese assim como na manutenção do estado diferenciado dos tecidos adultos. Por exemplo, as interacções de indução que ocorrem entre camadas de células e tecidos vizinhas no embrião em desenvolvimento estão grandemente dependentes da existência e da regulação de moléculas de sinalização segregadas. Nas interacções de indução, os sinais bioquímicos segregados por uma população de células influenciam o destino de desenvolvimento de uma segunda população de células, tipicamente por alteração do destino da segunda população de células. Por exemplo, as proteínas Wnt são agora reconhecidas como uma das principais famílias de moléculas de sinalização importantes no desenvolvimento em organismos que vão desde a *Drosophila* ao ratinho. Para uma revisão veja-se Cadigan, K.M. e R. Nusse (1997) *Genes & Development*, 11:3286-3305.

É hoje igualmente bem reconhecido que importantes famílias de moléculas de sinalização possuem um papel duplo a desempenhar tanto no desenvolvimento de um organismo como na promoção ou manutenção do estado diferenciado de tecidos no animal adulto. Adicionalmente, importantes famílias de moléculas de sinalização foram implicadas no controlo da proliferação de células em tecido adulto maduro, por exemplo, durante o normal *turnover* celular no organismo adulto assim como na regeneração de tecidos activada em resultado de danos no tecido adulto.

Sumário da invenção

A presente invenção baseia-se, pelo menos em parte, na identificação de moléculas de ácido nucleico que codificam uma nova família de proteínas segregadas, aqui referida como as Proteínas Segregadas Ricas em cisteínas ("proteínas CRSP" ou "CRSP", do inglês *Cysteine-Rich Secreted Proteins*). As moléculas de CRSP da presente invenção são úteis como agentes moduladores na regulação de uma variedade de processos celulares. Assim, num aspecto, a presente invenção proporciona moléculas de ácido nucleico isoladas que codificam proteínas CRSP ou suas porções biologicamente activas, assim como fragmentos de ácido nucleico adequados como iniciadores ou sondas de hibridação para a detecção de ácidos nucleicos que codificam CRSP.

Numa concretização, a invenção proporciona uma molécula de ácido nucleico de CRSP tal como definida nas reivindicações anexas.

Outro aspecto da invenção proporciona um vector compreendendo uma molécula de ácido nucleico de CRSP. Em certas concretizações, o vector é um vector de expressão recombinante. Noutra concretização, a invenção proporciona uma célula hospedeira contendo um vector da invenção. A invenção proporciona também um método para a produção de uma proteína CRSP através de cultura, num meio adequado, de uma célula hospedeira da invenção contendo um vector de expressão recombinante de modo a produzir uma proteína CRSP.

Outro aspecto da presente invenção caracteriza proteínas e polipéptidos de CRSP isolados ou recombinantes. Numa concretização, uma proteína CRSP isolada possui uma sequência de sinal e pelo menos um domínio rico em cisteínas, preferivelmente uma região rica em cisteínas, e é segregada. Noutra concretização, uma proteína CRSP isolada possui uma sequência de aminoácidos tal como definida nas reivindicações anexas.

As proteínas CRSP da presente invenção, ou suas porções biologicamente activas, podem estar operativamente ligadas a um polipéptido que não de CRSP formando proteínas de fusão de

CRSP. A invenção caracteriza ainda anticorpos que se ligam especificamente a proteínas CRSP, tais como anticorpos monoclonais ou policlonais. Em adição, as proteínas CRSP ou suas porções biologicamente activas podem estar incorporadas em composições farmacêuticas, que opcionalmente incluem transportadores farmacêuticamente aceitáveis.

Noutro aspecto, a presente invenção proporciona um método para a detecção de expressão de CRSP numa amostra biológica através de contacto da amostra biológica com um agente capaz de detectar uma molécula de ácido nucleico, uma proteína ou um polipéptido de CRSP, de modo que a presença de uma molécula de ácido nucleico, uma proteína ou um polipéptido de CRSP é detectada na amostra biológica.

Noutro aspecto, a presente invenção proporciona um método para a detecção da presença de actividade de CRSP numa amostra biológica através de contacto da amostra biológica com um agente capaz de detectar um indicador de actividade de CRSP de modo que a presença de actividade de CRSP é detectada na amostra biológica.

Noutro aspecto, a invenção proporciona um método para a modulação da actividade de CRSP compreendendo o contacto da célula com um agente que modula a actividade de CRSP de modo que a actividade de CRSP na célula é modulada. Numa concretização, o agente inibe a actividade de CRSP. Noutra concretização, o agente estimula a actividade de CRSP. Numa concretização, o agente é um anticorpo que se liga especificamente a uma proteína CRSP. Noutra concretização, o agente modula a expressão de CRSP através da modulação da transcrição de um gene de CRSP ou da tradução de um ARNm de CRSP. Ainda noutra concretização, o agente é uma molécula de ácido nucleico possuindo uma sequência nucleotídica que é anti-sentido relativamente à cadeia de codificação de um ARNm de CRSP ou um gene de CRSP.

Os métodos baseados na presente invenção podem ser utilizados para tratar um indivíduo possuindo uma desordem caracterizada por uma expressão ou actividade aberrantes de uma proteína CRSP ou um ácido nucleico de CRSP, por administração ao indivíduo de um agente que é um modulador de

CRSP. Numa concretização, o modulador de CRSP é uma proteína CRSP. Noutra concretização o modulador de CRSP é uma molécula de ácido nucleico de CRSP. Ainda noutra concretização, o modulador de CRSP é um péptido, um peptidomimético ou outra molécula pequena. Numa concretização preferida, a desordem caracterizada por expressão aberrante de proteína CRSP ou ácido nucleico de CRSP é uma desordem de desenvolvimento, de diferenciação ou de proliferação.

A presente invenção proporciona também um ensaio de diagnóstico para a identificação da presença ou ausência de uma alteração genética caracterizada por pelo menos uma de (i) modificação aberrante ou mutação de um gene que codifica uma proteína CRSP; (ii) má-regulação do referido gene; e (iii) modificação pós-tradução aberrante de uma proteína CRSP, em que uma forma de tipo selvagem do referido gene codifica uma proteína com uma actividade de CRSP.

Noutro aspecto, a invenção proporciona um método para a identificação de um composto que se liga a, ou modula a actividade de, uma proteína CRSP, proporcionando uma composição indicadora compreendendo uma proteína CRSP possuindo actividade de CRSP, colocando em contacto a composição indicadora com um composto de teste e determinando o efeito do composto de teste sobre a actividade de CRSP na composição indicadora para identificar um composto que modula a actividade de uma proteína CRSP.

Outras características e vantagens da invenção serão evidentes a partir da seguinte descrição detalhada e das reivindicações.

Breve Descrição dos Desenhos

A *Figura 1* representa a sequência de ADNc e a sequência de aminoácidos prevista de CRSP-1 humana. A sequência nucleotídica corresponde aos ácidos nucleicos 1 a 2479 de SEQ ID NO:1. A sequência de aminoácidos corresponde aos aminoácidos 1 a 350 de SEQ ID NO:2.

A *Figura 2* representa a sequência de ADNc e a sequência de aminoácidos prevista de CRSP-2 humana. A sequência

nucleotídica corresponde aos ácidos nucleicos 1 a 848 de SEQ ID NO:4. A sequência de aminoácidos corresponde aos aminoácidos 1 a 224 de SEQ ID NO:5.

A *Figura 3* representa a sequência de ADNc e a sequência de aminoácidos prevista de CRSP-3 humana. A sequência nucleotídica corresponde aos ácidos nucleicos 1 a 1529 de SEQ ID NO:7. A sequência de aminoácidos corresponde aos aminoácidos 1 a 266 de SEQ ID NO:8.

A *Figura 4* representa a sequência de ADNc e a sequência de aminoácidos prevista de CRSP-4 humana. A sequência nucleotídica corresponde aos ácidos nucleicos 1 a 702 de SEQ ID NO:10. A sequência de aminoácidos corresponde aos aminoácidos 1 a 179 de SEQ ID NO:11.

A *Figura 5* representa a sequência de ADNc e a sequência de aminoácidos prevista de CRSP-like-N humana. A sequência nucleotídica corresponde aos ácidos nucleicos 1 a 928 de SEQ ID NO:13. A sequência de aminoácidos corresponde aos aminoácidos 1 a 242 de SEQ ID NO:14.

A *Figura 6* representa as sequências de aminoácidos de CRSP-1 humana, CRSP-2 humana, CRSP-3 humana e CRSP-4 humana, assim como a sequência de consenso gerada a partir do alinhamento destas sequências. A figura detalha a relação entre as proteínas CRSP da presente invenção.

A *Figura 7* é um diagrama esquemático que representa a relação entre as proteínas CRSP da presente invenção. A figura representa os domínios biológico e funcional das CRSP humanas. Os péptidos de sinal estão indicados por caixas a cheio. Os domínios ricos em cisteínas da região rica em cisteínas de CRSP estão representados como CRD-1 e CRD-2.

A *Figura 8* representa a sequência de ADNc e a sequência de aminoácidos prevista de CRSP-1 murina. A sequência nucleotídica corresponde aos ácidos nucleicos 1 a 928 de SEQ ID NO:16. A sequência de aminoácidos corresponde aos aminoácidos 1 a 349 de SEQ ID NO:17.

A *Figura 9* é um diagrama esquemático que representa a relação entre a sequência nucleotídica de CRSP-1 e as de RIG e RIG-like 7-1 (N.^{os} de acesso U32331 e AF034208, respectivamente).

Descrição Detalhada da Invenção

A presente invenção baseia-se na identificação de novas moléculas, aqui referidas como proteína CRSP e moléculas de ácido nucleico de CRSP, que constituem uma família de moléculas possuindo certas características estruturais e funcionais conservadas. O termo "família", quando referido à proteína e a moléculas de ácido nucleico da invenção, significa duas ou mais proteínas ou moléculas de ácido nucleico possuindo um domínio estrutural comum e possuindo suficiente homologia de sequência nucleotídica ou de aminoácidos como aqui definido. Os membros desta família podem ser de ocorrência natural e podem ser da mesma espécie ou de espécies diferentes. Por exemplo, uma família pode conter uma primeira proteína de origem humana, assim como outras proteínas, distintas, de origem humana, ou alternativamente, pode conter homólogos de origem não humana. Os membros de uma família podem também possuir características funcionais comuns.

Numa concretização, um membro da família CRSP é identificado com base na presença de pelo menos um "domínio rico em cisteínas" na proteína ou na correspondente molécula de ácido nucleico. Tal como aqui definido, um "domínio rico em cisteínas" refere-se a uma porção de uma proteína CRSP (e.g., CRSP-1) que é rica em resíduos de cisteína, i.e., pelo menos cerca de 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ou 20 aminoácidos são resíduos de cisteína. Preferivelmente, um "domínio rico em cisteínas" é um domínio de proteína possuindo uma sequência de aminoácidos de cerca de 30-100 aminoácidos, preferivelmente cerca de 35-95 aminoácidos, mais preferivelmente cerca de 40-90 aminoácidos, mais preferivelmente cerca de 45-85 aminoácidos, ainda mais preferivelmente cerca de 50-80 aminoácidos, e ainda mais preferivelmente cerca de 55-75, 60-70, ou 65 aminoácidos, dos quais pelo menos cerca de 3-20, preferivelmente cerca de 5-15, ou mais preferivelmente cerca de 6, 7, 8, 9, 10, 11, 12, 13,

14, 15, 16, 17, 18, 19, ou 20 aminoácidos são resíduos de cisteína.

Um domínio rico em cisteínas preferido possui pelo menos cerca de 10, pelo menos cerca de 15, 16, 17, 18, 19, ou preferivelmente 20 resíduos de cisteína localizados na mesma posição relativa de aminoácido que os resíduos de cisteína na CRSP-1 humana possuindo SEQ ID NO:2. Por exemplo, como mostrado na Figura 6, a CRSP-2 possui pelo menos cerca de 10 resíduos de cisteína localizados na mesma posição relativa de aminoácido que os resíduos de cisteína na CRSP-1 humana possuindo SEQ ID NO:2 (e.g., cys151 em CRSP-2, SEQ ID NO:5, está localizado na mesma posição relativa de aminoácido que cys214 em CRSP-1, SEQ ID NO:2; cys156 em CRSP-2, SEQ ID NO:5, está localizado na mesma posição relativa de aminoácido que cys219 em CRSP-1, SEQ ID NO:2; e cys157 em CRSP-2, SEQ ID NO:5, está localizado na mesma posição relativa de aminoácido que cys220 em CRSP-1, SEQ ID NO:2). Similarmente, como mostrado na Figura 6, a CRSP-3 possui pelo menos cerca de 10 resíduos de cisteína localizados na mesma posição relativa de aminoácido que os resíduos de cisteína na CRSP-1 humana possuindo SEQ ID NO:2. Como também mostrado na Figura 6, a CRSP-4 possui pelo menos cerca de 10 resíduos de cisteína localizados na mesma posição relativa de aminoácido que os resíduos de cisteína na CRSP-1 humana possuindo SEQ ID NO:2.

Uma proteína CRSP preferida da presente invenção é uma proteína humana (e.g., codificada por uma sequência nucleotídica correspondente a um gene humano de ocorrência natural). Por exemplo, numa concretização, uma proteína CRSP-1 (CRSP-1 humana) contém um primeiro domínio rico em cisteínas contendo cerca dos aminoácidos 147-195 de SEQ ID NO:2, possuindo 10 resíduos de cisteína, e um segundo domínio rico em cisteínas contendo cerca dos aminoácidos 201-284 de SEQ ID NO:2, possuindo 10 resíduos de cisteína (as posições dos resíduos de cisteína estão representadas na Figura 6). Noutra concretização, uma proteína CRSP-2 (CRSP-2 humana) contém um primeiro domínio rico em cisteínas contendo cerca dos aminoácidos 41-90 de SEQ ID NO:5, possuindo 10 resíduos de cisteína, e um segundo domínio rico em cisteínas contendo cerca dos aminoácidos 41-218 de SEQ ID NO:5, possuindo 10 resíduos de cisteína (as posições dos resíduos de cisteína

estão representadas na Figura 6). Noutra concretização, uma proteína CRSP-3 (CRSP-3 humana) contém um primeiro domínio rico em cisteínas contendo cerca dos aminoácidos 85-138 de SEQ ID NO:8, possuindo 10 resíduos de cisteína, e um segundo domínio rico em cisteínas contendo cerca dos aminoácidos 182-263 de SEQ ID NO:8, possuindo 10 resíduos de cisteína (as posições dos resíduos de cisteína estão representadas na Figura 6). Noutra concretização, uma proteína CRSP-4 (CRSP-4 humana) contém um primeiro domínio rico em cisteínas contendo cerca dos aminoácidos 1-47 de SEQ ID NO:11, possuindo 8 resíduos de cisteína, e um segundo domínio rico em cisteínas contendo cerca dos aminoácidos 96-176 de SEQ ID NO:11, possuindo 10 resíduos de cisteína (as posições dos resíduos de cisteína estão representadas na Figura 6).

As proteínas CRSP preferidas possuem mais de um domínio rico em cisteínas, mais preferivelmente possuem pelo menos dois domínios ricos em cisteínas e, assim, possuem uma região rica em cisteínas. Tal como aqui se utiliza, a expressão "região rica em cisteínas" refere-se a um domínio de proteína que inclui pelo menos dois domínios ricos em cisteínas e possui uma sequência de aminoácidos de cerca de 120-200, preferivelmente cerca de 130-190, mais preferivelmente cerca de 140-180 resíduos de aminoácido, e ainda mais preferivelmente pelo menos cerca de 135-175 aminoácidos dos quais pelo menos cerca de 10-30, preferivelmente cerca de 15-20, e mais preferivelmente cerca de 16, 17, 18, ou 19 dos aminoácidos são resíduos de cisteína. Numa concretização preferida, uma região rica em cisteínas está localizada na região C-terminal de uma proteína CRSP. Por exemplo, numa concretização, uma proteína CRSP-1 contém uma região rica em cisteínas contendo cerca dos aminoácidos 147-284 de SEQ ID NO:2, possuindo 20 resíduos de cisteína nas posições indicadas na Figura 6. Noutra concretização, uma proteína CRSP-2 contém uma região rica em cisteínas contendo cerca dos aminoácidos 41-218 de SEQ ID NO:5, possuindo 20 resíduos de cisteína nas posições indicadas na Figura 6. Noutra concretização, uma proteína CRSP-3 contém uma região rica em cisteínas contendo cerca dos aminoácidos 85-263 de SEQ ID NO:8, possuindo 20 resíduos de cisteína nas posições indicadas na Figura 6. Noutra concretização, uma proteína CRSP-4 contém uma região rica em cisteínas contendo cerca dos

aminoácidos 1-176 de SEQ ID NO:2, possuindo 18 resíduos de cisteína nas posições indicadas na Figura 6.

Noutra concretização, em adição aos domínios ricos em cisteínas, a região rica em cisteínas contém uma região espaçadora que separa os primeiro e segundo domínios ricos em cisteínas. Tal como aqui se utiliza, a "região espaçadora" refere-se a resíduos de aminoácido que estão localizados entre os primeiro e segundo domínios ricos em cisteínas de uma região rica em cisteínas e inclui resíduos de aminoácido localizados em posição C-terminal relativamente ao primeiro domínio rico em cisteínas e N-terminal relativamente ao segundo domínio rico em cisteínas. Tal como aqui definido, uma "região espaçadora" refere-se a um domínio de proteína de cerca de 5-70 aminoácidos, preferivelmente cerca de 10-65 aminoácidos, mais preferivelmente cerca de 15-60 aminoácidos, ainda mais preferivelmente cerca de 20-55 aminoácidos, e ainda mais preferivelmente cerca de 25-50, 30-45 ou 35-40 aminoácidos. Por exemplo, a proteína CRSP-1 contém uma região espaçadora de cerca dos aminoácidos 196-200 de SEQ ID NO:2; a proteína CRSP-2 contém uma região espaçadora de cerca dos aminoácidos 91-137 de SEQ ID NO:5; a proteína CRSP-3 contém uma região espaçadora de cerca dos aminoácidos 139-181 de SEQ ID NO:8; e a proteína CRSP-4 contém uma região espaçadora de cerca dos aminoácidos 48-95 de SEQ ID NO:11.

Noutra concretização da invenção, a proteína CRSP possui pelo menos um domínio rico em cisteínas, preferivelmente uma região rica em cisteínas, e uma sequência de sinal. Tal como aqui se utiliza, uma "sequência de sinal" refere-se a um péptido contendo cerca de 20 aminoácidos que ocorre no terminal N de proteínas de membrana integrais e secretórias e que contém um grande número de resíduos de aminoácido hidrófobos. Por exemplo, uma sequência de sinal contém pelo menos cerca de 14-28 resíduos de aminoácido, preferivelmente cerca de 16-26 resíduos de aminoácido, mais preferivelmente cerca de 18-24 resíduos de aminoácido, e mais preferivelmente cerca de 20-22 resíduos de aminoácido, e possui pelo menos cerca de 40-70%, preferivelmente cerca de 50-65%, e mais preferivelmente cerca de 55-60% resíduos de aminoácido hidrófobos (e.g., Alanina, Valina, Leucina, Isoleucina, Fenilalanina, Tirosina, Triptofano ou Prolina). Esta

"sequência de sinal", também referida na especialidade por "péptido de sinal", serve para dirigir uma proteína contendo uma destas sequências para uma bicamada lipídica. Por exemplo, numa concretização, uma proteína CRSP-1 contém uma sequência de sinal de cerca dos aminoácidos 1-23 de SEQ ID NO:2. Noutra concretização, uma proteína CRSP-2 contém uma sequência de sinal de cerca dos aminoácidos 1-19 de SEQ ID NO:5. Noutra concretização, uma proteína CRSP-3 contém uma sequência de sinal de cerca dos aminoácidos 1-20 de SEQ ID NO:8.

As proteínas CRSP da presente invenção podem ser utilizadas para identificar membros adicionais da família CRSP. Por exemplo, uma proteína possuindo homologia com CRSP-1 foi identificada utilizando a sequência nucleotídica que codifica a região N-terminal única de CRSP-1 para rastrear uma base de dados de sequências de proteínas. Esta proteína, referida aqui como "CRSP-like-N" está representada na Figura 5. A sequência nucleotídica de CRSP-like-N (SEQ ID NO:13) codifica uma proteína possuindo 242 aminoácidos (SEQ ID NO:14). A sequência nucleotídica de CRSP-like-N inclui uma região a 5' não traduzida contendo os nucleótidos 1-74 de SEQ ID NO:13, uma região codificante contendo os nucleótidos 75-800 de SEQ ID NO:13 (correspondentes aos nucleótidos 1-726 de SEQ ID NO:15), e uma região a 3' não traduzida contendo os nucleótidos 801-928 de SEQ ID NO:13.

Deste modo, uma concretização da invenção caracteriza uma proteína CRSP possuindo pelo menos um domínio rico em cisteínas, preferivelmente pelo menos uma região rica em cisteínas. Outra concretização caracteriza uma proteína CRSP possuindo pelo menos uma região rica em cisteínas, em que a região rica em cisteínas inclui pelo menos um domínio rico em cisteínas. Outra concretização caracteriza uma proteína CRSP possuindo pelo menos uma região rica em cisteínas, em que a região rica em cisteínas inclui pelo menos dois domínios ricos em cisteínas. Outra concretização caracteriza uma proteína possuindo 20, 30, 40, 50, 60, 70, 80, 90, 95 ou 99% de homologia com um domínio rico em cisteínas de uma proteína CRSP da invenção (e.g., CRSP-1, CRSP-2, CRSP-3 ou CRSP-4).

Ainda outra concretização da invenção caracteriza uma proteína CRSP possuindo pelo menos um domínio rico em

cisteínas, preferivelmente pelo menos uma região rica em cisteínas e um péptido de sinal. Outra concretização caracteriza uma proteína CRSP possuindo pelo menos um domínio rico em cisteínas, preferivelmente pelo menos uma região rica em cisteínas e um péptido de sinal, em que a região rica em cisteínas inclui pelo menos dois domínios ricos em cisteínas. Outra concretização caracteriza uma proteína CRSP possuindo pelo menos um domínio rico em cisteínas, preferivelmente pelo menos uma região rica em cisteínas e um péptido de sinal, em que a região rica em cisteínas inclui pelo menos dois domínios ricos em cisteínas e um espaçador.

As moléculas de CRSP preferidas da presente invenção estão definidas nas reivindicações anexas. Tal como aqui se utiliza, a expressão "suficientemente homóloga" refere-se a uma primeira sequência nucleotídica ou de aminoácidos que contém um número suficiente ou mínimo de resíduos de aminoácido ou nucleótidos idênticos ou equivalentes (e.g., um resíduo de aminoácido que possui uma cadeia lateral similar) a uma segunda sequência nucleotídica ou de aminoácidos, de modo que a primeira e a segunda sequências nucleotídicas ou de aminoácidos partilham domínios estruturais comuns e/ou uma actividade funcional comum. Por exemplo, as sequências nucleotídicas ou de aminoácidos que partilham domínios estruturais comuns, possuem pelo menos cerca de 40% de homologia, preferivelmente 50% de homologia, mais preferivelmente 60%-70% de homologia ao longo das sequências de aminoácidos dos domínios e contêm pelo menos um, preferivelmente dois, mais preferivelmente três, e ainda mais preferivelmente quatro, cinco ou seis domínios estruturais, são aqui definidas como suficientemente homólogas. Adicionalmente, as sequências nucleotídicas ou de aminoácidos que partilham pelo menos 40%, preferivelmente 50%, mais preferivelmente 60, 70, ou 80% de homologia e partilham uma actividade funcional comum são aqui definidas como suficientemente homólogas.

Tal como aqui se utilizam indiferentemente, uma "actividade de CRSP", "actividade biológica de CRSP" ou "actividade funcional de CRSP", referem-se a uma actividade exercida por uma proteína, um polipéptido ou uma molécula de ácido nucleico de CRSP, sobre uma célula responsiva a CRSP,

determinada *in vivo* ou *in vitro*, de acordo com técnicas *standard*. Numa concretização, uma actividade de CRSP é uma actividade directa, tal como uma associação com uma molécula alvo de CRSP. Tal como aqui se utiliza, uma "molécula alvo" é uma molécula com a qual uma proteína CRSP se liga ou interage na natureza, de tal modo que é conseguida uma função mediada por CRSP. Uma molécula alvo de CRSP pode ser uma molécula que não uma CRSP ou uma proteína ou um polipéptido de CRSP da presente invenção. Numa concretização exemplificativa, uma molécula alvo de CRSP é uma proteína ligada à membrana (*e.g.*, um receptor de superfície celular ou "receptor de CRSP") ou uma forma modificada de uma destas proteínas que foi alterada para que a proteína ficasse solúvel (*e.g.*, produzida recombinantemente de modo a que a proteína não expresse um domínio de ligação à membrana). Noutra concretização, um alvo de CRSP é uma segunda molécula de proteína solúvel (*e.g.*, um "parceiro de ligação de CRSP" ou "substrato de CRSP"). Nesta concretização exemplificativa, um parceiro de ligação de CRSP pode ser uma segunda proteína solúvel que não uma CRSP, ou uma segunda molécula de proteína CRSP da presente invenção. Alternativamente, uma actividade de CRSP é uma actividade indirecta, tal como uma actividade de sinalização celular mediada por interacção da proteína CRSP com uma segunda proteína (*e.g.*, um receptor de CRSP). Tal como aqui se utiliza, a expressão "receptor de CRSP" refere-se a uma proteína ou um complexo de proteína, a que uma proteína CRSP, *e.g.*, CRSP humana, se pode ligar. Um receptor pode ser um receptor de superfície celular, *e.g.*, um receptor de uma hormona nuclear. Os receptores de CRSP podem ser isolados por métodos conhecidos na especialidade e estão adicionalmente aqui descritos. A interacção de uma proteína CRSP com um receptor de CRSP pode resultar na transdução de um sinal da superfície celular para o núcleo. O sinal transduzido pode ser um aumento do cálcio intracelular, um aumento do fosfatidilinositol ou outra molécula, e pode resultar, *e.g.*, na fosforilação de proteínas específicas, numa modulação da transcrição génica e quaisquer das outras actividades biológicas aqui definidas.

Numa concretização preferida, uma actividade de CRSP é pelo menos uma ou mais das seguintes actividades: (i) interacção de uma proteína CRSP com, e/ou ligação a, uma

segunda molécula, (e.g., uma proteína, tal como um receptor de CRSP, uma forma solúvel de um receptor de CRSP, um receptor para um membro da família *wnt* de proteínas de sinalização, ou uma molécula de sinalização que não CRSP); (ii) interacção de uma proteína CRSP com uma proteína intracelular por via de um receptor de CRSP ligado a membranas; (iii) formação de complexos entre uma proteína CRSP solúvel e um segundo parceiro de ligação de CRSP solúvel (e.g., uma molécula que não uma proteína CRSP ou uma segunda molécula de proteína CRSP); (iv) interacção com outras proteínas extracelulares (e.g., regulação de adesão celular dependente de *wnt* a componentes da matriz extracelular); (v) ligação a, e eliminação de, uma molécula indesejável (e.g., uma actividade de desintoxicação ou função de defesa); e/ou (vi) uma actividade enzimática. Em ainda outra concretização preferida, uma actividade de CRSP é pelo menos uma ou mais das seguintes actividades: (1) modulação da transdução de sinais celulares, quer *in vitro* quer *in vivo* (e.g., antagonismo da actividade de membros da família *wnt* de proteínas segregadas ou supressão da transdução de sinal dependente de *wnt*); (2) regulação da comunicação entre células (e.g., regulação de interacções célula-célula dependentes de *wnt*); (3) regulação da expressão de genes cuja expressão é modulada por ligação de CRSP (e.g., CRSP-1) a um receptor; (4) regulação da transcrição de genes numa célula envolvida no desenvolvimento ou na diferenciação, quer *in vitro* quer *in vivo* (e.g., indução de diferenciação celular); (5) regulação da transcrição de genes numa célula envolvida no desenvolvimento ou na diferenciação, em que pelo menos um gene codifica uma proteína específica da diferenciação; (6) regulação da transcrição de genes numa célula envolvida no desenvolvimento ou na diferenciação, em que pelo menos um gene codifica uma segunda proteína segregada; (7) regulação da transcrição de genes numa célula envolvida no desenvolvimento ou na diferenciação, em que pelo menos um gene codifica uma molécula de transdução de sinal; (8) regulação da proliferação celular, quer *in vitro* quer *in vivo* (e.g., indução de proliferação celular ou inibição da proliferação como no caso de supressão da tumorigénese (e.g., supressão da formação de glioblastomas)); (9) formação e manutenção de arranjos espaciais ordenados de tecidos diferenciados em vertebrados, tanto adultos como embrionários (e.g., indução da formação da cabeça durante o desenvolvimento

dos vertebrados ou manutenção das células progenitoras hematopoiéticas); (10) modulação da morte celular, tal como a estimulação da sobrevivência das células; (11) regulação da migração celular; e/ou (12) modulação imunitária.

Como aqui referido, "proteínas específicas da diferenciação" incluem proteínas envolvidas na transição de uma célula do fenótipo indiferenciado para o diferenciado. Por exemplo, estas proteínas podem ser proteínas estruturais específicas da diferenciação ou factores de transcrição específicos da diferenciação. Estas proteínas específicas da diferenciação são geralmente expressas em níveis mais elevados em células que estão a fazer a transição do fenótipo indiferenciado para o diferenciado (e.g., durante o desenvolvimento embrionário ou durante a regeneração do tecido maduro no animal adulto), ou são expressas em níveis superiores em células completamente diferenciadas ou terminalmente diferenciadas em comparação com as suas contrapartidas indiferenciadas. Igualmente, como aqui referido, "genes específicos da diferenciação" incluem moléculas de ácido nucleico que codificam proteínas específicas da diferenciação.

Deste modo, outra concretização da invenção caracteriza proteínas e polipéptidos de CRSP isolados possuindo uma actividade de CRSP. As proteínas CRSP preferidas possuem pelo menos uma região rica em cisteínas e uma actividade de CRSP. Noutra concretização preferida, a proteína CRSP possui pelo menos uma região rica em cisteínas, em que a região rica em cisteínas compreende pelo menos um domínio rico em cisteínas, e uma actividade de CRSP. Noutra concretização preferida, a proteína CRSP possui pelo menos uma região rica em cisteínas, em que a região rica em cisteínas compreende pelo menos dois domínios ricos em cisteínas, e uma actividade de CRSP. Em ainda outra concretização preferida, uma proteína CRSP compreende ainda uma sequência de sinal. Ainda noutra concretização preferida, uma proteína CRSP possui uma região rica em cisteínas, uma actividade de CRSP, e uma sequência de aminoácidos tal como definida nas reivindicações anexas.

O ADNc da CRSP-1 humana, que tem aproximadamente 2479 nucleótidos de comprimento, codifica uma proteína que tem

aproximadamente 350 resíduos de aminoácido de comprimento. A proteína CRSP humana contém uma sequência de sinal N-terminal e uma região rica em cisteínas compreendendo dois domínios ricos em cisteínas. Uma região rica em cisteínas da CRSP pode ser encontrada pelo menos, por exemplo, a cerca dos aminoácidos 147-284 de SEQ ID NO:2. A região rica em cisteínas de CRSP-1 compreende um primeiro domínio rico em cisteínas a cerca dos aminoácidos 147-195 de SEQ ID NO:2 e um segundo domínio rico em cisteínas a cerca dos aminoácidos 201-284 de SEQ ID NO:2. A proteína CRSP-1 humana é uma proteína segregada que contém ainda uma sequência de sinal a cerca dos aminoácidos 1-23 de SEQ ID NO:2. A previsão deste péptido de sinal pode ser feita, por exemplo, utilizando o algoritmo de computador SIGNALP (Henrik, et al. (1997) *Protein Engineering* 10:1-6).

O ADNc da CRSP-2 humana, que tem aproximadamente 848 nucleótidos de comprimento, codifica uma proteína que tem aproximadamente 224 resíduos de aminoácido de comprimento. A proteína CRSP humana contém uma sequência de sinal N-terminal e uma região rica em cisteínas compreendendo dois domínios ricos em cisteínas. Uma região rica em cisteínas de CRSP pode ser encontrada pelo menos, por exemplo, a cerca dos aminoácidos 41-218 de SEQ ID NO:5. A região rica em cisteínas de CRSP-2 compreende um primeiro domínio rico em cisteínas a cerca dos aminoácidos 41-90 de SEQ ID NO:5 e um segundo domínio rico em cisteínas a cerca dos aminoácidos 138-218 de SEQ ID NO:5. A proteína CRSP-2 humana é uma proteína segregada que contém ainda uma sequência de sinal a cerca dos aminoácidos 1-19 de SEQ ID NO:5.

O ADNc da CRSP-3 humana, que tem aproximadamente 1529 nucleótidos de comprimento, codifica uma proteína que tem aproximadamente 266 resíduos de aminoácido de comprimento. A proteína CRSP humana contém uma sequência de sinal N-terminal e uma região rica em cisteínas compreendendo dois domínios ricos em cisteínas. Uma região rica em cisteínas de CRSP pode ser encontrada pelo menos, por exemplo, a cerca dos aminoácidos 85-263 de SEQ ID NO:8. A região rica em cisteínas de CRSP-3 compreende um primeiro domínio rico em cisteínas a cerca dos aminoácidos 85-138 de SEQ ID NO:8 e um segundo domínio rico em cisteínas a cerca dos aminoácidos 182-263 de

SEQ ID NO:2. A proteína CRSP-3 humana é uma proteína segregada que contém ainda uma sequência de sinal a cerca dos aminoácidos 1-20 de SEQ ID NO:8.

O ADNc da CRSP-4 humana, que tem aproximadamente 702 nucleótidos de comprimento, codifica uma proteína que tem aproximadamente 179 resíduos de aminoácido de comprimento. A proteína CRSP humana contém uma região rica em cisteínas compreendendo dois domínios ricos em cisteínas. Uma região rica em cisteínas de CRSP pode ser encontrada pelo menos, por exemplo, a cerca dos aminoácidos 1-176 de SEQ ID NO:11. A região rica em cisteínas de CRSP-4 compreende um primeiro domínio rico em cisteínas a cerca dos aminoácidos 1-47 de SEQ ID NO:11 e um segundo domínio rico em cisteínas a cerca dos aminoácidos 96-176 de SEQ ID NO:11.

Vários aspectos da invenção estão descritos com mais detalhes nas subsecções que se seguem:

I. Moléculas de Ácido Nucleico Isoladas

Um aspecto da invenção refere-se a moléculas de ácido nucleico isoladas que codificam proteínas CRSP ou suas porções biologicamente activas, assim como fragmentos de ácido nucleico suficientes para utilização como sondas de hibridação para identificar ácidos nucleicos que codificam CRSP (e.g., ARNm de CRSP) e fragmentos para utilização como iniciadores de PCR para a amplificação ou mutação de moléculas de ácido nucleico de CRSP. Tal como aqui se utiliza, a expressão "molécula de ácido nucleico" inclui moléculas de ADN (e.g., ADNc ou ADN genómico) e moléculas de ARN (e.g., ARNm) e análogos do ADN ou ARN gerados utilizando análogos nucleotídicos. A molécula de ácido nucleico pode ser de cadeia simples ou de cadeia dupla, mas preferivelmente é ADN de cadeia dupla.

Uma molécula de ácido nucleico "isolada" é uma molécula que está separada de outras moléculas de ácido nucleico que estão presentes na fonte natural do ácido nucleico. Preferivelmente, um ácido nucleico "isolado" está isento de sequências que flanqueiam naturalmente o ácido nucleico (i.e., sequências localizadas nas extremidades 5' e 3' do ácido

nucleico) no ADN genómico do organismo do qual o ácido nucleico é derivado. Por exemplo, em várias concretizações, a molécula de ácido nucleico de CRSP isolada pode conter menos de cerca de 5 kb, 4kb, 3kb, 2kb, 1 kb, 0,5 kb ou 0,1 kb de sequências nucleotídicas que flanqueiam naturalmente a molécula de ácido nucleico no ADN genómico da célula da qual o ácido nucleico é derivado. Adicionalmente, uma molécula de ácido nucleico "isolada", tal como uma molécula de ADNc, pode estar substancialmente isenta de outro material celular, ou meio de cultura quando produzido por técnicas recombinantes, ou substancialmente isenta de precursores químicos ou outros químicos quando sintetizada quimicamente.

Uma molécula de ácido nucleico da presente invenção, e.g., uma molécula de ácido nucleico possuindo a sequência nucleotídica tal como definida nas reivindicações anexas, pode ser isolada utilizando técnicas *standard* de biologia molecular e a informação sobre a sequência aqui proporcionada. Utilizando a totalidade ou uma porção da sequência de ácido nucleico de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, ou SEQ ID NO:10, ou a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452 como sonda de hibridação, podem ser isoladas moléculas de ácido nucleico de CRSP utilizando técnicas *standard* de hibridação e clonagem (e.g., como descrito em Sambrook, J., Fritsh, E.F., e Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Adicionalmente, uma molécula de ácido nucleico abrangendo a totalidade ou uma porção de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7 ou SEQ ID NO:10 ou a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452, pode ser isolada através da reacção em cadeia com polimerase (PCR)

utilizando iniciadores oligonucleotídicos sintéticos desenhados com base na sequência de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7 ou SEQ ID NO:10 ou na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452.

Um ácido nucleico da invenção pode ser amplificado utilizando ADNc, ARNm ou alternativamente, ADN genómico, como molde, e iniciadores oligonucleotídicos apropriados de acordo com técnicas de amplificação por PCR *standard*. O ácido nucleico assim amplificado pode ser clonado num vector apropriado e caracterizado por análise da sequência de ADN. Adicionalmente, podem preparar-se oligonucleótidos correspondentes a sequências nucleotídicas de CRSP através de técnicas de síntese *standard*, e.g., utilizando um sintetizador automático de ADN.

Numa concretização preferida, uma molécula de ácido nucleico isolada da invenção compreende uma sequência nucleotídica tal como definida nas reivindicações.

A sequência de SEQ ID NO:1 corresponde ao ADNc da CRSP-1 humana. Este ADNc compreende sequências que codificam a proteína CRSP-1 humana (*i.e.*, "a região codificante", dos nucleótidos 38-1087), assim como sequências não traduzidas a 5' (nucleótidos 1 a 37) e sequências não traduzidas a 3' (nucleótidos 1088 a 2479). Alternativamente, a molécula de ácido nucleico pode compreender apenas a região codificante de SEQ ID NO:1, (e.g., nucleótidos 38 a 1087, correspondentes a SEQ ID NO:3). Um plasmídeo contendo a sequência nucleotídica de comprimento completo que codifica CRSP-1 foi depositada na American Type Culture Collection (ATCC), Manassas, Virginia, E.U.A. 20110-2209 em 11 de Junho, 1997 e foi-lhe atribuído o Número de Acesso 98452. Um plasmídeo contendo a sequência nucleotídica de comprimento completo que codifica CRSP-1 foi depositado na American Type Culture Collection (ATCC), Manassas, Virginia, E.U.A. 20110-2209 em 16 de Janeiro, 1998 e foi-lhe atribuído o Número de Acesso 98634.

A sequência de SEQ ID NO:4 corresponde ao ADNc da CRSP-2 humana. Este ADNc compreende sequências que codificam a proteína CRSP-2 humana (*i.e.*, "a região codificante", dos nucleótidos 126-796), assim como sequências não traduzidas a 5' (nucleótidos 1 a 125) e sequências não traduzidas a 3' (nucleótidos 797 a 848). Alternativamente, a molécula de ácido nucleico pode compreender apenas a região codificante de SEQ ID NO:4 (*e.g.*, nucleótidos 126 a 796, correspondentes a SEQ ID NO:6).

A SEQ ID NO:7 corresponde ao ADNc da CRSP-3 humana. Este ADNc compreende sequências que codificam a proteína CRSP-3 humana (*i.e.*, "a região codificante", dos nucleótidos 93-890), assim como sequências não traduzidas a 5' (nucleótidos 1 a 92) e sequências não traduzidas a 3' (nucleótidos 891 a 1529). Alternativamente, a molécula de ácido nucleico pode compreender apenas a região codificante de SEQ ID NO:7 (*e.g.*, nucleótidos 93 a 890 correspondentes a SEQ ID NO:9). Um plasmídeo contendo a sequência nucleotídica de comprimento completo que codifica CRSP-3 foi depositado na American Type Culture Collection (ATCC), Manassas, Virginia, E.U.A. 20110-2209, em 16 de Janeiro, 1998 e foi-lhe atribuído o Número de Acesso 98633.

A sequência de SEQ ID NO:10 corresponde ao ADNc da CRSP-4 humana. Este ADNc compreende sequências que codificam a proteína CRSP-4 humana (*i.e.*, "a região codificante", dos nucleótidos 1-537), assim como sequências não traduzidas a 3' (nucleótidos 538 a 702). Alternativamente, a molécula de ácido nucleico pode compreender apenas a região codificante de SEQ ID NO:10 (*e.g.*, nucleótidos 1 a 537, correspondentes a SEQ ID NO:12).

Noutra concretização preferida, uma molécula de ácido nucleico isolada da invenção compreende uma molécula de ácido nucleico que é um complemento da sequência nucleotídica reivindicada. Uma molécula de ácido nucleico que é complementar à sequência nucleica reivindicada é uma molécula que é suficientemente complementar à sequência nucleotídica reivindicada de modo a que possa hibridar com a sequência nucleotídica reivindicada e desse modo formar uma dúplex estável.

As moléculas de ácido nucleico isoladas úteis compreendem uma sequência nucleotídica que é pelo menos cerca de 30-35%, preferivelmente cerca de 40-45%; mais preferivelmente cerca de 50-55%, ainda mais preferivelmente cerca de 60-65%, e ainda mais preferivelmente pelo menos cerca de 70-75%, 80-85%, 90-95% ou mais, homóloga às sequências nucleotídicas reivindicadas, ou uma porção de qualquer destas sequências nucleotídicas.

Adicionalmente, uma molécula de ácido nucleico útil pode compreender apenas uma porção da sequência de ácido nucleico de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452, por exemplo um fragmento que pode ser utilizado como sonda ou iniciador ou um fragmento que codifica uma porção biologicamente activa de uma proteína CRSP. A sequência nucleotídica determinada a partir da clonagem dos genes da CRSP humana permite a criação de sondas e iniciadores desenhados para utilização na identificação e/ou clonagem de homólogos de CRSP em outros tipos de células, e.g., de outros tecidos, assim como homólogos de CRSP de outros mamíferos. A sonda/iniciador tipicamente compreendem um oligonucleótido substancialmente purificado. O oligonucleótido tipicamente compreende uma região de sequência nucleotídica que hibrida sob condições rigorosas com pelo menos cerca de 12, preferivelmente cerca de 25, mais preferivelmente cerca de 40, 50 ou 75 nucleótidos consecutivos de uma sequência de sentido directo de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452, de uma sequência anti-sentido de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, da sequência nucleotídica da inserção de ADN

do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452, ou de um mutante de ocorrência natural de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452. Numa concretização exemplificativa, uma molécula de ácido nucleico útil compreende uma sequência nucleotídica que hibrida sob condições de hibridação rigorosas com uma molécula de ácido nucleico compreendendo os nucleótidos 470-2479 de SEQ ID NO:1 ou com uma molécula de ácido nucleico compreendendo os nucleótidos 1-475 de SEQ ID NO:4.

As sondas baseadas na sequência nucleotídica da CRSP humana podem ser utilizadas para detectar transcritos ou sequências genómicas que codificam as mesmas proteínas ou proteínas homólogas. Por exemplo os iniciadores baseados no ácido nucleico representado em SEQ ID NO:1 ou 3 podem ser utilizados em reacções PCR para clonar homólogos de CRSP (e.g., homólogos de CRSP-1). Numa concretização preferida da invenção, os homólogos de CRSP são clonados por amplificação por PCR (e.g., RT-PCR) utilizando iniciadores que hibridam com uma porção da sequência nucleotídica que codifica o domínio rico em cisteínas da CRSP. Do mesmo modo, as sondas baseadas nas sequências de CRSP em questão podem ser utilizadas para detectar transcritos ou sequências genómicas que codificam as mesmas proteínas ou proteínas homólogas. Em concretizações preferidas, a sonda compreende ainda um grupo marcador ligado, e.g., um grupo marcador pode ser um radioisótopo, um composto fluorescente, uma enzima ou um cofactor de enzima. Estas sondas podem ser utilizadas como parte de um *kit* de teste de diagnóstico para a identificação de células ou tecidos que expressam erradamente uma proteína CRSP, tal como por medição de um nível de um ácido nucleico que codifica CRSP numa amostra de células de um indivíduo e.g., detectando os níveis de ARNm de CRSP ou determinando se um gene de CRSP genómico foi mutado ou eliminado.

Um fragmento de ácido nucleico que codifica uma "porção biologicamente activa de uma proteína CRSP" pode ser preparado por isolamento de uma porção de uma sequência definida nas reivindicações anexas, que codifica um polipéptido possuindo uma actividade biológica de CRSP (as actividades biológicas das proteínas CRSP foram previamente descritas), expressão da porção codificada da proteína CRSP (e.g., por expressão recombinante *in vitro*) e avaliação da actividade da porção codificada da proteína CRSP.

A invenção abrange ainda moléculas de ácido nucleico que diferem da sequência nucleotídica apresentada em SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452, devido a degenerescência do código genético e assim codificam as mesmas proteínas CRSP que as que são codificadas pela sequência nucleotídica definida nas reivindicações anexas. Noutra concretização, uma molécula de ácido nucleico isolada da invenção possui uma sequência nucleotídica que codifica uma proteína possuindo uma sequência de aminoácidos definida nas reivindicações anexas.

Em adição às sequências nucleotídicas de CRSP humana apresentadas nas reivindicações, os peritos na especialidade entenderão que podem existir polimorfismos das sequências de ADN que conduzem a alterações nas sequências de aminoácidos das proteínas CRSP numa população (e.g., a população humana). Este polimorfismo genético nos genes de CRSP pode existir entre indivíduos numa população devido à variação alélica natural. Tal como aqui se utilizam, os termos "gene" e "gene recombinante" referem-se a moléculas de ácido nucleico compreendendo um quadro de leitura aberto que codifica uma proteína CRSP, preferivelmente uma proteína CRSP de mamífero. Estas variações alélicas naturais podem tipicamente resultar numa variância de 1-5% na sequência nucleotídica de um gene de CRSP. Qualquer uma e todas estas variações de nucleótidos e resultantes polimorfismos de aminoácidos nos genes de CRSP que são o resultado de variação alélica natural e que não alteram

a actividade funcional de uma proteína CRSP devem entender-se no âmbito da invenção.

Adicionalmente, as moléculas de ácido nucleico que codificam proteínas CRSP de outras espécies, e que portanto possuem uma sequência nucleotídica que difere da sequência de ácido nucleico humana, devem entender-se no âmbito da invenção. Por exemplo, foi identificado um ADNc de CRSP-1 murina com base na sequência nucleotídica da CRSP-1 humana. A sequência nucleotídica da CRSP-1 murina (SEQ ID NO:16) codifica uma proteína CRSP-1 possuindo 349 aminoácidos. As sequências de nucleótidos e de aminoácidos da CRSP-1 murina estão representadas na Figura 8. A região codificante da CRSP-1 murina está representada em SEQ ID NO:18.

Moléculas de ácido nucleico correspondentes a variantes alélicas naturais e homólogos dos ADNc de CRSP da invenção podem ser isoladas com base na sua homologia com os ácidos nucleicos da CRSP humana aqui revelados utilizando o ADNc humano, ou uma sua porção, como sonda de hibridação de acordo com técnicas *standard* de hibridação sob condições de hibridação rigorosas. Os exemplos de tecidos e/ou bibliotecas adequados para isolamento dos ácidos nucleicos em questão incluem timo, nódulos linfáticos e tecido inflamatório. O ADNc que codifica uma proteína CRSP (e.g., uma proteína CRSP-1) pode ser obtido por isolamento de ARNm total de uma célula, e.g., uma célula de vertebrado, uma célula de mamífero ou uma célula humana, incluindo células embrionárias. Os ADNc de cadeia dupla podem então ser preparados a partir do ARNm total, e subsequentemente inseridos num vector plasmídico ou bacteriofágico adequados utilizando qualquer uma de várias das técnicas conhecidas. O gene que codifica uma proteína CRSP-1 pode também ser clonado utilizando técnicas de reacção em cadeia com polimerase estabelecidas, de acordo com a informação sobre a sequência nucleotídica proporcionada pela invenção. O ácido nucleico da invenção pode ser ADN ou ARN ou seus análogos.

Assim, uma molécula de ácido nucleico isolada útil tem pelo menos 15 nucleótidos de comprimento e hibrida sob condições rigorosas com a molécula de ácido nucleico compreendendo a sequência nucleotídica de SEQ ID NO:1,

SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452. Noutra concretização, o ácido nucleico tem pelo menos 30, 50, 100, 250 ou 500 nucleótidos de comprimento. Tal como aqui se utiliza, a expressão "hibrida sob condições rigorosas" pretende descrever condições para hibridação e lavagem sob as quais as sequências nucleotídicas pelo menos 60% homólogas umas das outras tipicamente permanecem hibridadas umas com as outras. Preferivelmente, as condições são tais que sequências pelo menos cerca de 70%, mais preferivelmente pelo menos cerca de 80%, ainda mais preferivelmente pelo menos cerca de 85% ou 90% homólogas umas das outras, tipicamente permanecem hibridadas umas com as outras. Estas condições rigorosas são conhecidas dos peritos na especialidade e podem ser encontradas em *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Um exemplo não limitante, preferido, de condições de hibridação rigorosas são hibridação em cloreto de sódio/citrato de sódio (SSC) 6x a cerca de 45°C, seguida de uma ou mais lavagens com SSC 0,2x, SDS a 0,1% a 50-65°C. Preferivelmente, uma molécula de ácido nucleico isolada da invenção que hibrida sob condições rigorosas com a sequência de SEQ ID NO:1 corresponde a uma molécula de ácido nucleico de ocorrência natural. Tal como aqui se utiliza, uma molécula de ácido nucleico "de ocorrência natural" refere-se a uma molécula de ARN ou ADN possuindo uma sequência nucleotídica que ocorre na natureza (e.g., codifica uma proteína natural).

Em adição a variantes alélicas de ocorrência natural das sequências de CRSP que podem existir na população, o perito na especialidade notará ainda que podem ser introduzidas alterações por mutação nas sequências nucleotídicas de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou na sequência nucleotídica da inserção de ADN do plasmídeo depositado na

ATCC com o Número de Acesso 98452, que conduzem a alterações na sequência de aminoácidos das proteínas CRSP codificadas, sem alteração da capacidade funcional das proteínas CRSP. Por exemplo, substituições de nucleótidos que conduzem a substituições de aminoácidos (particularmente substituições conservativas de aminoácidos) em resíduos de aminoácido "não essenciais" podem ser feitas na sequência de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452. Um resíduo de aminoácido "não essencial" é um resíduo que pode ser alterado relativamente à sequência de CRSP de tipo selvagem (e.g., a sequência de SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8 ou SEQ ID NO:11) sem alteração da actividade biológica, enquanto um resíduo de aminoácido "essencial" é necessário para a actividade biológica. Por exemplo, os resíduos de aminoácido que são conservados entre as proteínas CRSP da presente invenção (e.g., resíduos de cisteína dentro dos domínios ricos em cisteínas), prevêm-se particularmente insusceptíveis de alteração. Adicionalmente, os resíduos de aminoácido que são conservados entre a proteína CRSP e outras proteínas possuindo domínios ricos em cisteínas provavelmente são insusceptíveis de alteração.

Deste modo, a invenção refere-se a moléculas de ácido nucleico que codificam proteínas CRSP que contêm alterações em resíduos de aminoácido que não são essenciais para a actividade. Estas proteínas CRSP diferem na sequência de aminoácidos de SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8 ou SEQ ID NO:11, mas retêm a actividade biológica. Uma molécula de ácido nucleico isolada útil compreende uma sequência nucleotídica que codifica uma proteína, em que a proteína compreende uma sequência de aminoácidos pelo menos cerca de 60% homóloga à sequência de aminoácidos de SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, ou SEQ ID NO:11. Uma proteína útil codificada pela molécula de ácido nucleico é pelo menos cerca de 65-70% homóloga a SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8 ou SEQ ID NO:11, mais preferivelmente pelo menos cerca de 75-80% homóloga a SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8 ou

SEQ ID NO:11, ainda mais preferivelmente pelo menos cerca de 85-90% homóloga a SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8 ou SEQ ID NO:11, e mais preferivelmente pelo menos cerca de 95% homóloga a SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8 ou SEQ ID NO:11.

Uma molécula de ácido nucleico isolada que codifica uma proteína CRSP homóloga à proteína de SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8 ou SEQ ID NO:11, pode ser criada por introdução de uma ou mais substituições, adições ou deleções de nucleótidos na sequência nucleotídica de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452, de modo a que sejam introduzidas uma ou mais substituições, adições ou deleções de aminoácidos na proteína codificada. As mutações podem ser introduzidas em SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou na sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452, por técnicas *standard*, tais como mutagénesse dirigida a um local e mutagénesse mediada por PCR. Preferivelmente, são feitas substituições conservativas de aminoácidos num ou mais resíduos de aminoácido não essenciais previstos. Uma "substituição conservativa de aminoácidos" é uma substituição em que o resíduo de aminoácido é substituído por um resíduo de aminoácido possuindo uma cadeia lateral similar. As famílias de resíduos de aminoácido possuindo cadeias laterais similares foram definidas na especialidade. Estas famílias incluem aminoácidos com cadeias laterais básicas (e.g., lisina, arginina, histidina), cadeias laterais ácidas (e.g., ácido aspártico, ácido glutâmico), cadeias laterais polares não carregadas (e.g., glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadeias laterais não polares (e.g., alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptofano), cadeias laterais

ramificadas em beta (e.g., treonina, valina, isoleucina) e cadeias laterais aromáticas (e.g., tirosina, fenilalanina, triptofano, histidina). Assim, um resíduo de aminoácido não essencial previsto numa proteína CRSP (e.g., um que não está localizado num domínio rico em cisteínas) é preferivelmente substituído por outro resíduo de aminoácido da mesma família de cadeias laterais. Alternativamente, noutra concretização, as mutações podem ser introduzidas aleatoriamente ao longo da totalidade ou de parte de uma sequência de codificação de CRSP, tal como por mutagénesse de saturação, e os mutantes resultantes podem ser rastreados quanto a actividade biológica de CRSP para identificar os mutantes que retêm a actividade. Após a mutagénesse de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634, da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633, ou da sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98452, a proteína codificada pode ser expressa de forma recombinante e a actividade da proteína pode ser determinada.

Uma proteína CRSP mutante pode ser ensaiada quanto ao cálcio intracelular, um aumento de fosfatidilinositol ou outra molécula, e pode resultar, e.g., em fosforilação de proteínas específicas, uma modulação da transcrição génica e qualquer das outras actividades biológicas atrás estabelecidas.

Uma proteína CRSP mutante pode também ser ensaiada quanto à capacidade de (1) modular a transdução de sinal celular, quer *in vitro* quer *in vivo*; (2) regular a comunicação entre células; (3) regular a expressão de genes cuja expressão é modulada por ligação de CRSP (e.g., CRSP-1) a um receptor; (4) regular a transcrição génica numa célula envolvida no desenvolvimento ou na diferenciação, quer *in vitro* quer *in vivo*; (5) regular a proliferação celular, quer *in vitro* quer *in vivo*; (6) formar e/ou manter arranjos espaciais ordenados de tecidos diferenciados em vertebrados; (7) modular a morte celular (e.g. sobrevivência celular); (8) regular a migração celular; e/ou (9) modular a função do sistema imunitário.

Em adição às moléculas de ácido nucleico que codificam proteínas CRSP atrás descritas, a invenção refere-se a moléculas de ácido nucleico isoladas que são anti-sentido relativamente às mesmas. Um ácido nucleico "anti-sentido" compreende uma sequência nucleotídica que é complementar a um ácido nucleico "de sentido directo" que codifica uma proteína, e.g., complementar à cadeia de codificação de uma molécula de ADNc de cadeia dupla ou complementar a uma sequência de ARNm. Deste modo, um ácido nucleico anti-sentido pode ligar-se por ligações de hidrogénio a um ácido nucleico de sentido directo. O ácido nucleico anti-sentido pode ser complementar a uma cadeia de codificação de CRSP completa, ou a apenas uma sua porção. Numa concretização, uma molécula de ácido nucleico anti-sentido é anti-sentido relativamente a uma "região codificante" da cadeia de codificação de uma sequência nucleotídica que codifica CRSP. A expressão "região codificante" refere-se à região da sequência nucleotídica compreendendo códons que são traduzidos em resíduos de aminoácido (e.g., a região codificante da CRSP-1 humana corresponde a SEQ ID NO:3, a região codificante da CRSP-2 humana corresponde a SEQ ID NO:6, a região codificante da CRSP-3 humana corresponde a SEQ ID NO:9, a região codificante da CRSP-4 humana corresponde a SEQ ID NO:12 e a região codificante da CRSP-like-N humana corresponde a SEQ ID NO:15). Noutra concretização, a molécula de ácido nucleico anti-sentido é anti-sentido relativamente a uma "região não codificante" da cadeia de codificação de uma sequência nucleotídica que codifica CRSP. A expressão "região não codificante" refere-se a sequências a 5' e a 3' que flanqueiam a região codificante que não são traduzidas em aminoácidos (i.e., também referidas como regiões não traduzidas a 5' e a 3').

Dadas as sequências das cadeias de codificação que codificam CRSP aqui reveladas (e.g., SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:9 ou SEQ ID NO:12), podem ser desenhados ácidos nucleicos anti-sentido de acordo com as regras de emparelhamento de bases de Watson e Crick. A molécula de ácido nucleico anti-sentido pode ser complementar a toda a região codificante do ARNm de CRSP, mas mais preferivelmente é um oligonucleótido que é anti-sentido apenas relativamente a uma porção da região codificante ou não codificante do ARNm de

CRSP. Por exemplo, o oligonucleótido anti-sentido pode ser complementar à região que rodeia o local de início da tradução do ARNm de CRSP. Um oligonucleótido anti-sentido pode ter, por exemplo, cerca de 5, 10, 15, 20, 25, 30, 35, 40, 45 ou 50 nucleótidos de comprimento. Um ácido nucleico anti-sentido pode ser construído utilizando síntese química e reacções de ligação enzimáticas utilizando procedimentos conhecidos na especialidade. Por exemplo, um ácido nucleico anti-sentido (e.g., um oligonucleótido anti-sentido) pode ser quimicamente sintetizado utilizando nucleótidos de ocorrência natural ou nucleótidos modificados de várias maneiras, desenhados para aumentar a estabilidade biológica das moléculas ou para aumentar a estabilidade física da dúplex formada entre os ácidos nucleicos anti-sentido e de sentido directo, e.g., podem ser utilizados derivados fosforotioato e nucleótidos substituídos com acridina. Os exemplos de nucleótidos modificados que podem ser utilizados para gerar o ácido nucleico anti-sentido incluem 5-fluorouracilo, 5-bromouracilo, 5-clorouracilo, 5-iodouracilo, hipoxantina, xantina, 4-acetilcitosina, 5-(carboxi-hidroxilmetil)uracilo, 5-carboximetilaminometil-2-tiouridina, 5-carboximetilaminometiluracilo, di-hidrouracilo, beta-D-galactosilqueosina, inosina, N6-isopenteniladenina, 1-metilguanina, 1-metilinosina, 2,2-dimetilguanina, 2-metiladenina, 2-metilguanina, 3-metilcitosina, 5-metilcitosina, N6-adenina, 7-metilguanina, 5-metilaminometiluracilo, 5-metoxiaminometil-2-tiouracilo, beta-D-manosilqueosina, 5'-metoxicarboximetiluracilo, 5-metoxiuracilo, 2-metiltio-N6-isopenteniladenina, ácido uracil-5-oxiacético (v), wibutoxosina, pseudouracilo, queosina, 2-tiocitosina, 5-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, éster metílico do ácido uracil-5-oxiacético, ácido uracil-5-oxiacético (v), 5-metil-2-tiouracilo, 3-(3-amino-3-N-2-carboxipropil)uracilo, (acp3)w, e 2,6-diaminopurina. Alternativamente, o ácido nucleico anti-sentido pode ser produzido biologicamente utilizando um vector de expressão no qual foi subclonado um ácido nucleico numa orientação anti-sentido (i.e., o ARN transcrito a partir do ácido nucleico inserido estará numa orientação anti-sentido relativamente a um ácido nucleico alvo de interesse, adicionalmente descrito na subsecção que se segue).

As moléculas de ácido nucleico anti-sentido são tipicamente administradas a um indivíduo ou geradas *in situ* de modo a que hibridem com, ou se liguem, ao ARNm celular e/ou ao ADN genómico que codifica uma proteína CRSP, para desse modo inibir a expressão da proteína, e.g., por inibição da transcrição e/ou da tradução. A hibridação pode ser por complementaridade convencional de nucleótidos para formar uma dúplex estável, ou, por exemplo, no caso de uma molécula de ácido nucleico anti-sentido que se liga a dúplexes de ADN, através de interacções específicas no sulco principal da hélice dupla. Um exemplo de uma via de administração de moléculas de ácido nucleico anti-sentido inclui a injeção directa num local de um tecido. Alternativamente, as moléculas de ácido nucleico anti-sentido podem ser modificadas para atingir células seleccionadas e depois administradas sistemicamente. Por exemplo, para administração sistémica, as moléculas anti-sentido podem ser modificadas de modo a que se liguem especificamente a receptores ou antigénios expressos na superfície de uma célula seleccionada, e.g., por ligação das moléculas de ácido nucleico anti-sentido a péptidos ou anticorpos que se ligam a receptores ou antigénios de superfície celular. As moléculas de ácido nucleico anti-sentido podem também ser entregues a células utilizando os vectores aqui descritos. Para atingir concentrações intracelulares suficientes das moléculas anti-sentido, preferem-se construções de vectores em que a molécula de ácido nucleico anti-sentido é colocada sob o controlo de um promotor pol II ou pol III forte.

A molécula de ácido nucleico anti-sentido pode ser uma molécula de ácido nucleico α -anomérica. Uma molécula de ácido nucleico α -anomérica forma híbridos de cadeia dupla específicos com ARN complementar em que, ao contrário das usuais unidades β , as cadeias correm paralelas umas às outras (Gaultier *et al.*, (1987) *Nucleic Acids Res.* 15:6625-6641). A molécula de ácido nucleico anti-sentido pode também compreender um 2'-o-metilribonucleótido (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) ou um análogo quimérico de ARN-ADN (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

Ainda noutra concretização, um ácido nucleico anti-sentido da invenção é uma ribozima. As ribozimas são moléculas

de ARN catalíticas com actividade de ribonuclease que são capazes de clivar um ácido nucleico de cadeia simples, tal como um ARNm, com o qual têm uma região complementar. Assim, as ribozimas (e.g., ribozimas "cabeça de martelo" (descritas em Haselhoff e Gerlach (1988) *Nature* 334:585-591)) podem ser utilizadas para clivar cataliticamente transcritos de ARNm de CRSP para desse modo inibir a tradução do ARNm de CRSP. Uma ribozima possuindo especificidade para com um ácido nucleico que codifica CRSP pode ser desenhada com base na sequência nucleotídica de um ADNc de CRSP aqui revelado (i.e., SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98634 ou a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633). Por exemplo, pode ser construído um derivado de um ARN de *Tetrahymena* L-19 IVS em que a sequência nucleotídica do local activo é complementar à sequência nucleotídica a clivar num ARNm que codifica CRSP. Veja-se, e.g., Cech et al. Patente U.S. 4 987 071; e Cech et al. Patente U.S. 5 116 742. Alternativamente, pode ser utilizado ARNm de CRSP para seleccionar um ARN catalítico possuindo uma actividade de ribonuclease específica, a partir de um conjunto de moléculas de ARN. Veja-se, e.g., Bartel, D. e Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternativamente, a expressão do gene de CRSP pode ser inibida direccionando sequências nucleotídicas complementares para a região reguladora da CRSP (e.g., o promotor e/ou potenciadores de CRSP) para formar estruturas helicoidais triplas que impedem a transcrição do gene de CRSP em células alvo. Veja-se genericamente, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann N.Y. Acad. Sci.* 660:27-36; e Maher, L.J. (1992) *Bioassays* 14(12):807-15.

As moléculas de ácido nucleico de CRSP da presente invenção podem ser modificadas na porção de bases, na porção de açúcares ou no esqueleto de fosfatos para melhorar, e.g., a estabilidade, a hibridação ou a solubilidade da molécula. Por exemplo, o esqueleto de desoxirribose-fosfatos das moléculas de ácido nucleico pode ser modificado para gerar ácidos nucleicos peptídicos (veja-se Hyrup B. et al. (1996) *Bioorganic & Medicinal Chemistry* 4(1): 5-23). Tal como aqui se

utilizam, as expressões "ácidos nucleicos peptídicos" ou "PNA" referem-se a mímicas de ácidos nucleicos, *e.g.*, mímicas de ADN, em que o esqueleto de desoxirribose-fosfatos está substituído por um esqueleto de pseudopéptidos e apenas as quatro nucleobases naturais são retidas. Mostrou-se que o esqueleto neutro dos PNA permite a hibridação específica com ADN e ARN sob condições de baixa força iónica. A síntese de oligómeros de PNA pode ser realizada utilizando protocolos *standard* de síntese peptídica em fase sólida como descrito em Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al.* *PNAS* 93: 14670-675.

Os PNA de moléculas de ácido nucleico de CRSP podem ser utilizados em aplicações terapêuticas e de diagnóstico. Por exemplo, os PNA podem ser utilizados como agentes anti-sentido ou antigénicos para modulação específica da sequência da expressão génica através de, por exemplo, indução da paragem da transcrição ou da tradução ou inibição da replicação. Os PNA de moléculas de ácido nucleico de CRSP podem também ser utilizados na análise de mutações de um único par de bases num gene, (*e.g.*, por fixação de PCR dirigida por PNA); como 'enzimas de restrição artificiais' quando utilizados em combinação com outras enzimas, (*e.g.*, nucleases S1 (Hyrup B. (1996) *supra*)); ou como sondas ou iniciadores para sequenciação de ADN ou hibridação (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

Os PNA de CRSP podem ser modificados, (*e.g.*, para aumentar a sua estabilidade ou assimilação celular), através da fixação de grupos lipófilos ou outros grupos auxiliares ao PNA, através da formação de quimeras PNA-ADN, ou por utilização de lipossomas ou outras técnicas de entrega de fármacos conhecidas na especialidade. Por exemplo, podem ser geradas quimeras PNA-ADN de moléculas de ácido nucleico de CRSP que podem combinar as propriedades vantajosas do PNA e do ADN. Estas quimeras permitem que enzimas de reconhecimento de ADN, (*e.g.*, ARNase H e ADN-polimerases), interagem com a porção de ADN enquanto a porção PNA proporcionará elevada afinidade de ligação e especificidade. As quimeras PNA-ADN podem ser ligadas utilizando ligantes de comprimentos apropriados seleccionados em termos de empilhamento de bases, número de ligações entre as nucleobases e orientação (Hyrup B.

(1996) *supra*). A síntese de quimeras PNA-ADN pode ser realizada como descrito em Hyrup B. (1996) *supra* e Finn P.J, *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. Por exemplo, uma cadeia de ADN pode ser sintetizada num suporte sólido utilizando a química *standard* de acoplamento de fosforamidito e podem ser utilizados análogos nucleosídicos modificados, *e.g.*, 5'-(4-metoxitritil)amino-5'-desoxi-timidina-fosforamidito, entre o PNA e a extremidade 5' do ADN (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). Os monómeros de PNA são então acoplados por passos para produzir uma molécula quimérica com um segmento de PNA a 5' e um segmento de ADN a 3' (Finn P.J. *et al.* (1996) *supra*). Alternativamente, as moléculas quiméricas podem ser sintetizadas com um segmento de ADN a 5' e um segmento de PNA a 3' (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

O oligonucleótido pode incluir outros grupos ligados tais como péptidos (*e.g.*, para direcção a receptores da célula hospedeira *in vivo*), ou agentes que facilitam o transporte através da membrana celular (veja-se, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. US.* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; Publicação PCT WO88/09810, publicado em 15 de Dezembro, 1988) ou a barreira hematoencefálica (veja-se, *e.g.*, a Publicação PCT WO89/10134, publicado em 25 de Abril, 1988). Em adição, os oligonucleótidos podem ser modificados com agentes de clivagem desencadeada por hibridação (veja-se, *e.g.*, Krol *et al.* (1988) *BioTechniques* 6:958-976) ou agentes intercalantes. (Veja-se, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). Para este fim, o oligonucleótido pode ser conjugado com outra molécula, (*e.g.*, um péptido, um agente de reticulação desencadeada por hibridação, um agente de transporte ou um agente de clivagem desencadeada por hibridação).

II. Proteínas CRSP Isoladas e Anticorpos Anti-CRSP

Um aspecto da invenção refere-se a proteínas CRSP isoladas, e suas porções biologicamente activas, assim como fragmentos polipeptídicos, adequados para utilização como imunogénios para criar anticorpos anti-CRSP. Numa concretização, proteínas CRSP nativas podem ser isoladas a

partir de fontes de células ou tecidos através de um esquema de purificação apropriado utilizando técnicas *standard* de purificação de proteínas. Noutra concretização, as proteínas CRSP são produzidas por técnicas de ADN recombinante. Em alternativa à expressão recombinante, uma proteína ou um polipéptido de CRSP podem ser sintetizados quimicamente utilizando técnicas *standard* de síntese peptídica.

Uma proteína "isolada" ou "purificada" ou uma sua porção biologicamente activa, está substancialmente isenta de material celular ou outras proteínas contaminantes da fonte de células ou tecidos da qual a proteína CRSP deriva, ou está substancialmente isenta de precursores químicos ou outros químicos quando sintetizada quimicamente. A expressão "substancialmente isenta de material celular" inclui preparações de proteína CRSP em que a proteína está separada de componentes celulares das células das quais é isolada ou produzida recombinantemente. Numa concretização, a expressão "substancialmente isenta de material celular" inclui preparações de proteína CRSP possuindo menos de cerca de 30% (em peso seco) de proteína que não CRSP (também referida aqui como "proteína contaminante"), mais preferivelmente menos de cerca de 20% de proteína que não CRSP, ainda mais preferivelmente menos de cerca de 10% de proteína que não CRSP, e o mais preferivelmente menos de cerca de 5% de proteína que não CRSP. Quando a proteína CRSP ou a sua porção biologicamente activa é produzida recombinantemente, está também preferivelmente substancialmente isenta de meio de cultura, *i.e.*, o meio de cultura representa menos de cerca de 20%, mais preferivelmente menos de cerca de 10%, e o mais preferivelmente menos de cerca de 5% do volume da preparação de proteína.

A expressão "substancialmente isento de precursores químicos ou outros químicos" inclui preparações de proteína CRSP em que a proteína está separada de precursores químicos ou outros químicos que estão envolvidos na síntese da proteína. Numa concretização, a expressão "substancialmente isento de precursores químicos ou outros químicos" inclui preparações de proteína CRSP possuindo menos de cerca de 30% (em peso seco) de precursores químicos ou químicos que não CRSP, mais preferivelmente menos de cerca de 20% de

precursores químicos ou químicos que não CRSP, ainda mais preferivelmente menos de cerca de 10% de precursores químicos ou químicos que não CRSP, e o mais preferivelmente menos de cerca de 5% de precursores químicos ou químicos que não CRSP.

As porções biologicamente activas de uma proteína CRSP incluem péptidos compreendendo sequências de aminoácidos suficientemente homólogas a, ou derivadas da sequência de aminoácidos da proteína CRSP, e.g., a sequência de aminoácidos apresentada em SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8 ou SEQ ID NO:11, que incluem menos aminoácidos do que as proteínas CRSP de comprimento completo, e exibem pelo menos uma actividade de uma proteína CRSP. Tipicamente, as porções biologicamente activas compreendem um domínio ou motivo com pelo menos uma actividade da proteína CRSP. Uma porção biologicamente activa de uma proteína CRSP pode ser um polipéptido que tem, por exemplo, 10, 25, 50, 100 ou mais aminoácidos de comprimento.

Numa concretização, uma porção biologicamente activa de uma proteína CRSP compreende pelo menos uma região rica em cisteínas. Noutra concretização, uma porção biologicamente activa de uma proteína CRSP compreende pelo menos uma região rica em cisteínas, em que a região rica em cisteínas inclui pelo menos um domínio rico em cisteínas. Em ainda outra concretização, uma porção biologicamente activa de uma proteína CRSP compreende pelo menos uma sequência de sinal.

Numa concretização alternativa, uma porção biologicamente activa de uma proteína CRSP compreende uma sequência de aminoácidos de CRSP a que falta uma sequência de sinal.

Deve entender-se que uma porção biologicamente activa preferida de uma proteína CRSP da presente invenção pode conter pelo menos um dos domínios estruturais acima identificados. Uma porção biologicamente activa mais preferida de uma proteína CRSP pode conter pelo menos dois dos domínios estruturais atrás identificados. Uma porção biologicamente activa ainda mais preferida de uma proteína CRSP pode conter pelo menos três dos domínios estruturais atrás identificados. Uma porção biologicamente activa particularmente preferida de uma proteína CRSP da presente invenção pode conter pelo menos quatro dos domínios estruturais atrás identificados.

Adicionalmente, outras porções biologicamente activas, em que outras regiões da proteína estão eliminadas, podem ser preparadas por técnicas recombinantes e avaliadas quanto a uma ou mais das actividades funcionais de uma proteína CRSP nativa.

A proteína CRSP da invenção possui uma sequência de aminoácidos tal como apresentada nas reivindicações anexas.

Para determinar a percentagem de homologia de duas sequências de aminoácidos ou de dois ácidos nucleicos, as sequências são alinhadas para fins de comparação óptima (e.g., podem ser introduzidos hiatos na sequência de uma primeira sequência de aminoácidos ou de ácido nucleico para um alinhamento óptimo com uma segunda sequência de aminoácidos ou de ácido nucleico). Os resíduos de aminoácido ou de nucleótidos nas correspondentes posições de aminoácido ou posições de nucleótido são então comparadas. Quando uma posição na primeira sequência está ocupada pelo mesmo resíduo de aminoácido ou nucleótido que a posição correspondente na segunda sequência, então as moléculas são homólogas nessa posição (i.e., tal como aqui se utiliza, "homologia" de aminoácidos ou ácido nucleico é equivalente a "identidade" de aminoácidos ou ácido nucleico). A percentagem de homologia entre as duas sequências é uma função do número de posições idênticas partilhadas pelas sequências (i.e., % de homologia = n.º de posições idênticas/n.º total de posições x 100). A determinação da percentagem de homologia entre duas sequências pode ser realizada utilizando um algoritmo matemático. Um exemplo preferido, não limitante, de um algoritmo matemático utilizado para a comparação de duas sequências é o algoritmo de Karlin e Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-68, como modificado em Karlin e Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-77. Este algoritmo está incorporado nos programas NBLAST e XBLAST de Altschul, et al. (1990) *J. Mol. Biol.* 215:403-10. Os rastreios de nucleótidos BLAST podem ser realizados com o programa NBLAST, pontuação = 100, comprimento = 12 para obter sequências nucleotídicas homólogas a moléculas de ácido nucleico de CRSP da invenção. Os rastreios de proteínas BLAST podem ser realizados com o programa XBLAST, pontuação = 50, comprimento = 3 para obter sequências de aminoácidos homólogas às moléculas de proteína

CRSP da invenção. Para obter alinhamentos com hiatos para fins de comparação, pode utilizar-se o Gapped BLAST como descrito em Altschul *et al.*, (1997) *Nucleic Acids Research* 25(17):3389-3402. Quando se utilizam os programas BLAST e Gapped BLAST, podem-se utilizar os parâmetros por defeito dos respectivos programas (e.g., XBLAST e NBLAST). Veja-se <http://www.ncbi.nlm.nih.gov>. Outro exemplo preferido, não limitante, de um algoritmo matemático utilizado para a comparação de sequências é o algoritmo de Myers e Miller, CABIOS (1989). Este algoritmo está incorporado no programa ALIGN (versão 2.0) que faz parte do pacote de software de alinhamento de sequências GCG. Quando se utiliza o programa ALIGN para comparação de sequências de aminoácidos, pode utilizar-se uma tabela de resíduos ponderal PAM120, uma penalização por comprimento de hiatos de 12 e uma penalização por hiatos de 4.

A invenção pode ser utilizada para proporcionar proteínas CRSP quiméricas ou de fusão. Tal como aqui se utiliza, uma "proteína quimérica" ou "proteína de fusão" de CRSP compreendem um polipéptido de CRSP operativamente ligado a um polipéptido que não CRSP. Um "polipéptido de CRSP" refere-se a um polipéptido possuindo uma sequência de aminoácidos correspondente a CRSP, enquanto um "polipéptido que não CRSP" refere-se a um polipéptido possuindo uma sequência de aminoácidos correspondente a uma proteína que não é substancialmente homóloga à proteína CRSP, e.g., uma proteína que é diferente da proteína CRSP e que é derivada do mesmo organismo ou de um organismo diferente. Numa proteína de fusão de CRSP o polipéptido de CRSP pode corresponder à totalidade ou a uma porção de uma proteína CRSP. Uma proteína de fusão de CRSP compreende pelo menos uma porção biologicamente activa de uma proteína CRSP. Outra proteína de fusão de CRSP compreende pelo menos duas porções biologicamente activas de uma proteína CRSP. Outra proteína de fusão de CRSP compreende pelo menos três porções biologicamente activas de uma proteína CRSP. Na proteína de fusão, a expressão "operativamente ligado" pretende indicar que o polipéptido de CRSP e o polipéptido que não CRSP estão fundidos em enquadramento um em relação ao outro. O polipéptido que não de CRSP pode ser fundido com o terminal N ou o terminal C do polipéptido de CRSP.

Por exemplo, uma proteína de fusão é uma proteína de fusão GST-CRSP em que as sequências de CRSP estão fundidas ao terminal C das sequências de GST. Estas proteínas de fusão podem facilitar a purificação de CRSP recombinantes.

Outra proteína de fusão é uma proteína CRSP contendo uma sequência de sinal heteróloga no seu terminal N. Por exemplo, a sequência de sinal da CRSP nativa (*i.e.*, cerca dos aminoácidos 1 a 23 de SEQ ID NO:2) pode ser removida e substituída por uma sequência de sinal de outra proteína. Em certas células hospedeiras (*e.g.*, células hospedeiras de mamífero), a expressão e/ou secreção de CRSP pode ser aumentada através da utilização de uma sequência de sinal heteróloga.

Ainda outra proteína de fusão é uma proteína de fusão CRSP-imunoglobulina em que as sequências de CRSP compreendendo principalmente as regiões ricas em cisteínas de CRSP estão fundidas com sequências derivadas de um membro da família de proteínas imunoglobulinas. Foram também preparados derivados solúveis de glicoproteínas da superfície celular na superfamília de genes de imunoglobulinas que consistem num domínio extracelular da glicoproteína da superfície celular fundido com uma região constante (Fc) de imunoglobulina (veja-se *e.g.*, Capon, D.J. *et al.* (1989) *Nature* 337:525-531 e Capon Patentes U.S. 5 116 964 e 5 428 130 [construções CD4-IgG1]; Linsley, P.S. *et al.* (1991) *J. Exp. Med.* 173:721-730 [uma construção CD28-IgG1 e uma construção B7-1-IgG1]; e Linsley, P.S. *et al.* (1991) *J. Exp. Med.* 174:561-569 e Patente U.S. 5 434 131 [uma CTLA4-IgG1]). Estas proteínas de fusão provaram-se úteis para a modulação de interações receptor-ligando. Prepararam-se derivados solúveis de proteínas de superfície celular das proteínas da superfamília dos receptores do factor de necrose tumoral (TNFR) consistindo num domínio extracelular do receptor de superfície celular fundido com uma região constante (Fc) de imunoglobulina (Veja-se por exemplo Moreland *et al.* (1997) *N. Engl. J. Med.* 337(3):141-147; van der Poll *et al.* (1997) *Blood* 89(10):3727-3734; e Ammann *et al.* (1997) *J. Clin. Invest.* 99(7):1699-1703.)

As proteínas de fusão CRSP-imunoglobulina podem ser incorporadas em composições farmacêuticas e administradas a um

indivíduo para inibir uma interacção entre um ligando de CRSP e um receptor de CRSP na superfície de uma célula, para desse modo suprimir a transdução de sinal mediada por CRSP *in vivo*. As proteínas de fusão CRSP-imunoglobulina podem ser utilizadas para afectar a biodisponibilidade de um receptor cognato de CRSP. A inibição da interacção ligando de CRSP/CRSP pode ser terapêuticamente útil tanto para o tratamento de desordens de diferenciação ou proliferação, como para a modulação (e.g., promoção ou inibição) de respostas de desenvolvimento, adesão celular e/ou destino celular. Adicionalmente, as proteínas de fusão CRSP-imunoglobulina da invenção podem ser utilizadas como imunogénios para produzir anticorpos anti-CRSP num indivíduo, para purificar ligandos de CRSP e em ensaios de rastreio para identificar moléculas que inibem a interacção de CRSP com um ligando de CRSP.

Preferivelmente, uma proteína quimérica ou de fusão de CRSP é produzida por técnicas *standard* de ADN recombinante. Por exemplo, fragmentos de ADN que codificam para as diferentes sequências polipeptídicas são ligados mutuamente em enquadramento de acordo com técnicas convencionais, por exemplo empregando terminais de extremidade lisas ou extremidades escalonadas para ligação, digestão com enzimas de restrição para proporcionar terminais apropriados, preenchimento de extremidades coesivas conforme apropriado, tratamento com fosfatase alcalina para evitar uniões indesejáveis e ligação enzimática. O gene da fusão pode ser sintetizado por técnicas convencionais incluindo sintetizadores automáticos de ADN. Alternativamente, pode-se realizar a amplificação por PCR de fragmentos de genes utilizando iniciadores de ancoragem que dão origem a extremidades protuberantes complementares entre dois fragmentos de genes consecutivos que podem subsequentemente ser hibridados e re-amplificados para gerar uma sequência de gene quimérica (veja-se, por exemplo, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Adicionalmente, estão disponíveis no comércio muitos vectores de expressão que codificam já uma porção de fusão (e.g., um polipéptido GST). Um ácido nucleico que codifica CRSP pode ser clonado num desses vectores de expressão de modo a que a porção de fusão fique ligada em enquadramento à proteína CRSP.

A presente invenção refere-se também a variantes das proteínas CRSP que funcionam quer como agonistas (miméticos) de CRSP quer como antagonistas de CRSP. As variantes das proteínas CRSP podem ser geradas por mutagénese, *e.g.*, mutação pontual discreta ou truncagem de uma proteína CRSP. Um agonista das proteínas CRSP pode reter substancialmente as mesmas, ou um subconjunto, das actividades biológicas da forma de ocorrência natural de uma proteína CRSP. Um antagonista de uma proteína CRSP pode inibir uma ou mais das actividades da forma de ocorrência natural da proteína CRSP através de, por exemplo, ligação competitiva, a um membro a jusante ou a montante, de uma cascata de sinalização celular que inclui a proteína CRSP. Assim, os efeitos biológicos específicos podem ser eliciados por tratamento com uma variante de função limitada. Numa concretização, o tratamento de um indivíduo com uma variante possuindo um subconjunto das actividades biológicas da forma de ocorrência natural da proteína possui menos efeitos secundários num indivíduo do que o tratamento com a forma de ocorrência natural da proteína CRSP.

As variantes de uma proteína CRSP que funcionam quer como agonistas (miméticos) de CRSP quer como antagonistas de CRSP, podem ser identificadas através de rastreio de bibliotecas combinatórias de mutantes, *e.g.*, mutantes de truncagem, de uma proteína CRSP quanto a actividade agonista ou antagonista da proteína CRSP. Numa concretização, é gerada uma biblioteca diversificada de variantes de CRSP por mutagénese combinatória ao nível do ácido nucleico e é codificada por uma biblioteca diversificada de genes. Uma biblioteca diversificada de variantes de CRSP pode ser produzida através de, por exemplo, ligação enzimática de uma mistura de oligonucleótidos sintéticos a sequências génicas de modo a que um conjunto degenerado de potenciais sequências de CRSP seja expressável na forma de polipéptidos individuais, ou alternativamente, na forma de um conjunto de proteínas de fusão maiores (*e.g.*, para exibição sobre fagos) contendo o conjunto de sequências de CRSP. Existem vários métodos que se podem utilizar para produzir bibliotecas de potenciais variantes de CRSP a partir de uma sequência oligonucleotídica degenerada. A síntese química de uma sequência de genes degenerada pode ser realizada num sintetizador automático de ADN, e o gene sintético pode ser então ligado a um vector de expressão

apropriado. A utilização de um conjunto degenerado de genes permite proporcionar, numa só mistura, todas as sequências que codificam o conjunto desejado de potenciais sequências de CRSP. Os métodos para a síntese de oligonucleótidos degenerados são conhecidos na especialidade (veja-se, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477.

Em adição, bibliotecas de fragmentos de uma sequência de codificação de uma proteína CRSP podem ser utilizadas para gerar uma população diversificada de fragmentos de CRSP para rastreio e subsequente selecção de variantes de uma proteína CRSP. Numa concretização, uma biblioteca de fragmentos de sequências de codificação pode ser gerada por tratamento de um fragmento de PCR de cadeia dupla de uma sequência de codificação de CRSP com uma nuclease sob condições em que ocorrem cortes ("*nicking*") apenas cerca de uma vez por molécula, desnaturação do ADN de cadeia dupla, renaturação do ADN para formar ADN de cadeia dupla que pode incluir pares de sentido directo/anti-sentido de diferentes produtos cortados ("*nicked*"), remoção das porções em cadeia simples das cópias novamente formadas por tratamento com nuclease SI, e ligação da biblioteca de fragmentos resultante a um vector de expressão. Através deste método, pode ser derivada uma biblioteca de expressão que codifica fragmentos N-terminais, C-terminais e internos de vários tamanhos da proteína CRSP.

São conhecidas na especialidade várias técnicas para o rastreio de produtos génicos de bibliotecas combinatórias preparadas por mutações pontuais ou truncagem, e para rastreio de bibliotecas de ADNC quanto a produtos génicos possuindo uma propriedade seleccionada. Estas técnicas são adaptáveis para um rápido rastreio das bibliotecas de genes geradas através da mutagénese combinatória de proteínas CRSP. As técnicas mais amplamente utilizadas, que são susceptíveis de análise de elevado rendimento, para rastreio de grandes bibliotecas de genes, incluem tipicamente a clonagem da biblioteca de genes em vectores de expressão replicáveis, transformação de células apropriadas com a biblioteca de vectores resultante e expressão dos genes combinatórios sob condições em que a detecção de uma actividade desejada facilita o isolamento do

vector que codifica o gene cujo produto foi detectado. A mutagénese por montagem recursiva (*Recursive Ensemble Mutagenesis* - REM), uma nova técnica que aumenta a frequência de mutantes funcionais nas bibliotecas, pode ser utilizada em combinação com os ensaios de rastreio para identificar variantes de CRSP (Arkin e Yourvan (1992) *PNAS* 89:7811-7815; Delgrave et al., (1993) *Protein Engineering* 6(3):327-331).

Os ensaios baseados em células podem ser explorados para analisar uma biblioteca de CRSP diversificada. Por exemplo, uma biblioteca de vectores de expressão pode ser transfectada numa linha celular que normalmente responde a um ligando particular de uma maneira dependente de CRSP. As células transfectadas são então colocadas em contacto com o ligando e o efeito da expressão do mutante sobre a sinalização pelo ligando pode ser detectado, e.g., medindo qualquer de várias respostas de células imunitárias. O ADN plasmídico pode então ser recuperado das células que pontuam para inibição, ou alternativamente, potenciação da indução de ligandos, e os clones individuais podem ser adicionalmente caracterizados.

Uma proteína CRSP isolada, ou uma sua porção ou um seu fragmento, podem ser utilizados como imunogénios para gerar anticorpos que se ligam a CRSP utilizando técnicas *standard* para a preparação de anticorpos policlonais e monoclonais. Pode ser utilizada uma proteína CRSP de comprimento completo ou, alternativamente, a invenção proporciona fragmentos peptídicos antigénicos de CRSP para utilização como imunogénios. O péptido antigénico de CRSP compreende pelo menos 8 resíduos de aminoácido e abrange um epítipo de CRSP de modo a que um anticorpo criado contra o péptido forme um complexo imunitário específico com CRSP. Preferivelmente, o péptido antigénico compreende pelo menos 10 resíduos de aminoácido, mais preferivelmente pelo menos 15 resíduos de aminoácido, ainda mais preferivelmente pelo menos 20 resíduos de aminoácido, e o mais preferivelmente pelo menos 30 resíduos de aminoácido.

Os epítipos preferidos abrangidos pelo péptido antigénico são regiões de CRSP que estão localizadas sobre a superfície da proteína, e.g., regiões hidrófilas.

Tipicamente é utilizado um imunogénio de CRSP para preparar anticorpos por imunização de um indivíduo adequado, (e.g., coelho, cabra, ratinho ou outro mamífero) com o imunogénio. Um preparação imunogénica apropriada pode conter, por exemplo, proteína CRSP expressa recombinantemente ou um polipéptido de CRSP sintetizado quimicamente. A preparação pode ainda incluir um adjuvante, tal como adjuvante completo ou incompleto de Freund, ou um agente imunoestimulante similar. A imunização de um indivíduo adequado com uma preparação imunogénica de CRSP induz uma resposta de anticorpos policlonais anti-CRSP.

Deste modo, outro aspecto da invenção refere-se a anticorpos anti-CRSP. O termo "anticorpo", tal como aqui se utiliza, refere-se a moléculas de imunoglobulina e a porções imunologicamente activas de moléculas de imunoglobulina, i.e., moléculas que contêm um local de ligação ao antigénio que liga especificamente (imunorreage com) um antigénio, tal como CRSP. Os exemplos de porções imunologicamente activas de moléculas de imunoglobulina incluem fragmentos F(ab) e F(ab')₂ que podem ser gerados através do tratamento do anticorpo com uma enzima tal como pepsina. A invenção proporciona anticorpos policlonais e monoclonais que se ligam a CRSP. As expressões "anticorpo monoclonal" ou "composição de anticorpos monoclonais", tal como aqui se utilizam, referem-se a uma população de moléculas de anticorpo que contêm apenas uma espécie de um local de ligação ao antigénio capaz de imunorreagir com um epítipo particular de CRSP. Uma composição de anticorpos monoclonais apresenta portanto tipicamente uma única afinidade de ligação para com uma proteína CRSP particular com a qual imunorreage.

Os anticorpos policlonais anti-CRSP podem ser preparados como descrito atrás por imunização de um indivíduo adequado com um imunogénio de CRSP. O título de anticorpo anti-CRSP no indivíduo imunizado pode ser monitorado ao longo do tempo por técnicas *standard*, tal como com um ensaio imunossorvente com enzima ligada (ELISA) utilizando CRSP imobilizada. Se desejado, as moléculas de anticorpo dirigido contra CRSP podem ser isoladas a partir do mamífero (e.g., do sangue) e adicionalmente purificadas por técnicas bem conhecidas, tais como cromatografia em proteína A, para obter a fracção de IgG.

Num momento apropriado após a imunização, e.g., quando os títulos de anticorpo anti-CRSP são os mais elevados, podem-se obter células produtoras do anticorpo a partir do indivíduo e utilizá-las para preparar anticorpos monoclonais por técnicas *standard*, tais como a técnica do hibridoma, originalmente descrita por Kohler e Milstein (1975) *Nature* 256:495-497) (veja-se também, Brown et al. (1981) *J. Immunol.* 127:539-46; Brown et al. (1980) *J. Biol. Chem.* 255:4980-83; Yeh et al. (1976) *PNAS* 76:2927-31; e Yeh et al. (1982) *Int. J. Cancer* 29:269-75), a mais recente técnica de hibridoma de células B humanas (Kozbor et al. (1983) *Immunol Today* 4:72), a técnica do hibridoma de EBV (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) ou técnicas de triomas. A tecnologia para a produção de hibridomas de anticorpos monoclonais é bem conhecida (veja-se genericamente R.H. Kenneth, em *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E.A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M.L. Gefter et al. (1977) *Somatic Cell Genet.* 3:231-36). Resumidamente, funde-se uma linha celular imortal (tipicamente um mieloma) com linfócitos (tipicamente esplenócitos) de um mamífero imunizado com um imunogénio de CRSP como descrito atrás, e rastreiam-se os sobrenadantes da cultura das células de hibridoma resultantes para identificar um hibridoma que produz um anticorpo monoclonal que se liga a CRSP.

Qualquer um dos muitos protocolos bem conhecidos utilizados para a fusão de linfócitos e linhas celulares imortalizadas pode ser aplicado para fins de gerar um anticorpo monoclonal anti-CRSP (veja-se, e.g., G. Galfre et al. (1977) *Nature* 266:55052; Gefter et al. *Somatic Cell Genet.*, supracitado; Lerner, *Yale J Biol. Med.*, supracitado; Kenneth, *Monoclonal Antibodies*, supracitado). Adicionalmente, as pessoas competentes na matéria notarão que existem muitas variações destes métodos que também poderão ser úteis. Tipicamente, a linha celular imortal (e.g., uma linha celular de mieloma) é derivada da mesma espécie de mamífero que os linfócitos. Por exemplo, podem ser preparados hibridomas de murino por fusão de linfócitos de um ratinho imunizado com uma preparação imunogénica da presente invenção com uma linha celular de ratinho imortalizada. As linhas celulares imortais preferidas são as linhas celulares de mieloma de ratinho que

são sensíveis a um meio de cultura contendo hipoxantina, aminopterina e timidina ("meio HAT"). Pode utilizar-se qualquer uma de várias linhas celulares de mieloma como parceiro de fusão de acordo com técnicas *standard*, e.g., as linhas de mieloma P3-NS1/1-Ag4-1, P3-x63-Ag8.653 ou Sp2/O-Ag14. Estas linhas de mieloma estão disponíveis na ATCC. Tipicamente, células de mieloma de ratinho sensíveis a HAT são fundidas com esplenócitos de ratinho utilizando polietilenoglicol ("PEG"). As células de hibridoma resultantes da fusão são então seleccionadas utilizando meio HAT, que mata as células de mieloma não fundidas e fundidas de modo não produtivo (os esplenócitos não fundidos morrem após vários dias porque não são transformados). As células de hibridoma que produzem um anticorpo monoclonal da invenção são detectadas por rastreio dos sobrenadantes da cultura do hibridoma quanto a anticorpos que se ligam a CRSP, e.g., utilizando um ensaio ELISA *standard*.

Em alternativa à preparação de hibridomas que segregam anticorpos monoclonais, um anticorpo monoclonal anti-CRSP pode ser identificado e isolado através do rastreio de uma biblioteca combinatória de imunoglobulinas recombinantes (e.g., uma biblioteca de exibição de anticorpos sobre fagos) com CRSP, para desse modo isolar membros da biblioteca de imunoglobulinas que se ligam a CRSP. Estão disponíveis comercialmente *kits* para a criação e rastreio de bibliotecas de exibição sobre fagos (e.g., o *Recombinant Phage Antibody System* da Pharmacia, N.º de Catálogo 27-9400-0 1; e o *SurfZAP™ Phage Display Kit* da Stratagene, N.º de Catálogo 240612). Adicionalmente, exemplos de métodos e reagentes particularmente susceptíveis de utilização na criação e pesquisa de bibliotecas de exibição de anticorpos podem ser encontrados em, por exemplo, Ladner et al. Patente U.S. 5 223 409; Kang et al. Publicação Internacional PCT WO 92/18619; Dower et al. Publicação Internacional PCT WO 91/17271; Winter et al. Publicação Internacional PCT WO 92/20791; Markland et al. Publicação Internacional PCT WO 92/15679; Breitling et al. Publicação Internacional PCT WO 93/01288; McCafferty et al. Publicação Internacional PCT WO 92/01047; Garrard et al. Publicação Internacional PCT WO 92/09690; Ladner et al. Publicação Internacional PCT WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372;

Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734; Hawkins et al. (1992) *J. Mol. Biol.* 226:889-896; Clarkson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *PNAS* 89:3576-3580; Garrad et al. (1991) *Bio/Technology* 9:1373-1377; Hoogenboom et al. (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas et al. (1991) *PNAS* 88:7978-7982; e McCafferty et al. *Nature* (1990) 348:552-554.

Adicionalmente, anticorpos anti-CRSP recombinantes, tais como anticorpos monoclonais quiméricos e humanizados, compreendendo tanto porções humanas como não humanas, que podem ser criados utilizando técnicas *standard* de ADN recombinante, estão dentro do âmbito da invenção. Estes anticorpos monoclonais quiméricos e humanizados podem ser produzidos por técnicas de ADN recombinante conhecidas na especialidade, por exemplo utilizando métodos descritos em Robinson et al. Pedido Internacional PCT/US86/02269; Akira, et al. Pedido de Patente Europeia 184 187; Taniguchi, M., Pedido de Patente Europeia 171 496; Morrison et al. Pedido de Patente Europeia 173 494; Neuberger et al. Publicação Internacional PCT WO 86/01533; Cabilly et al. Patente U.S. 4 816 567; Cabilly et al. Pedido de Patente Europeia 125 023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *PNAS* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *PNAS* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; e Shaw et al. (1988) *J Natl. Cancer Inst.* 80:1553-1559; Morrison, S.L. (1985) *Science* 229:1202-1207; Oi et al. (1986) *BioTechniques* 4:214; Winter Patente U.S. 5 225 539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; e Beidler et al. (1988) *J Immunol.* 141:4053-4060.

Pode utilizar-se um anticorpo anti-CRSP (e.g., anticorpo monoclonal) para isolar CRSP por técnicas *standard*, tais como cromatografia de afinidade ou imunoprecipitação. Um anticorpo anti-CRSP pode facilitar a purificação de CRSP natural a partir de células e de CRSP produzida recombinantemente expressa em células hospedeiras. Adicionalmente, pode utilizar-se um anticorpo anti-CRSP para detectar proteína CRSP (e.g., num lisado celular ou sobrenadante celular) de modo a avaliar a abundância e o padrão de expressão da proteína CRSP.

Os anticorpos anti-CRSP podem ser utilizados como diagnóstico para monitorar os níveis de proteína em tecidos como parte de um procedimento de testes clínicos, e.g., para, por exemplo, determinar a eficiência de um determinado regime de tratamento. A detecção pode ser facilitada por acoplamento (i.e., ligação física) do anticorpo a uma substância detectável. Os exemplos de substâncias detectáveis incluem várias enzimas, grupos prostéticos, materiais fluorescentes, materiais luminescentes, materiais bioluminescentes e materiais radioactivos. Os exemplos de enzimas adequadas incluem peroxidase de rábano, fosfatase alcalina, β -galactosidase ou acetilcolinesterase; os exemplos de complexos de grupos prostéticos adequados incluem estreptavidina/biotina e avidina/biotina; os exemplos de materiais fluorescentes adequados incluem umbeliferona, fluoresceína, isotiocianato de fluoresceína, rodamina, diclorotriazinilamino fluoresceína, cloreto de dansilo ou ficoeritrina; um exemplo de um material luminescente inclui luminol; os exemplos de materiais bioluminescentes incluem luciferase, luciferina e aequorina, e os exemplos de material radioactivo adequados incluem ^{125}I , ^{131}I , ^{35}S ou ^3H .

III. Vectores de Expressão e Células Hospedeiras Recombinantes

Outro aspecto da invenção refere-se a vectores, preferivelmente vectores de expressão, contendo um ácido nucleico que codifica CRSP (ou uma sua porção). Tal como aqui se utiliza, o termo "vector" refere-se a uma molécula de ácido nucleico capaz de transportar outro ácido nucleico ao qual foi ligado. Um tipo de vector é um "plasmídeo", que se refere a um ADN circular de cadeia dupla ao qual podem ser ligados segmentos de ADN adicionais. Outro tipo de vector é um vector viral, em que segmentos de ADN adicionais podem ser ligados ao genoma viral. Certos vectores são capazes de replicação autónoma numa célula hospedeira na qual são introduzidas (e.g., vectores bacterianos possuindo uma origem de replicação bacteriana e vectores epissómicos de mamífero). Outros vectores (e.g., vectores não epissómicos de mamífero) são integrados no genoma de uma célula hospedeira por introdução na célula hospedeira, e desse modo são replicados juntamente com o genoma do hospedeiro. Adicionalmente, certos vectores são capazes de dirigir a expressão de genes aos quais estão

operativamente ligados. Estes vectores são aqui referidos como "vectores de expressão". Em geral, os vectores de expressão de utilidade em técnicas de ADN recombinante estão frequentemente na forma de plasmídeos. No presente fascículo, "plasmídeo" e "vector" podem ser utilizados indiferentemente pois o plasmídeo é a forma de vector mais vulgarmente utilizada. Contudo, pretende-se incluir na invenção essas outras formas de vectores de expressão, tais como vectores virais (e.g., retrovírus defectivos na replicação, adenovírus e vírus adeno-associados), que servem funções equivalentes.

Os vectores de expressão recombinantes da invenção compreendem um ácido nucleico da invenção numa forma adequada para expressão do ácido nucleico numa célula hospedeira, o que significa que os vectores de expressão recombinantes incluem uma ou mais sequências reguladoras, seleccionadas com base nas células hospedeiras a utilizar para a expressão, que estão operativamente ligadas à sequência de ácido nucleico a expressar. Num vector de expressão recombinante, "operativamente ligado" significa que a sequência nucleotídica de interesse está ligada a sequência(s) reguladora(s) de uma maneira que permite a expressão da sequência nucleotídica (e.g., num sistema de transcrição/tradução *in vitro* ou numa célula hospedeira quando o vector é introduzido na célula hospedeira). Na expressão "sequência reguladora" pretende-se incluir promotores, potenciadores e outros elementos de controlo da expressão (e.g., sinais de poliadenilação). Estas sequências reguladoras estão descritas, por exemplo, em Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). As sequências reguladoras incluem aquelas que dirigem a expressão constitutiva de uma sequência nucleotídica em muitos tipos de células hospedeiras e aquelas que dirigem a expressão da sequência nucleotídica apenas em determinadas células hospedeiras (e.g., sequências reguladoras específicas de tecidos). Os peritos na especialidade notarão que o desenho do vector de expressão pode depender de factores como a escolha da célula hospedeira a transformar, o nível de expressão de proteína desejado, etc. Os vectores de expressão da invenção podem ser introduzidos em células hospedeiras para desse modo produzir proteínas ou péptidos, incluindo proteínas ou péptidos de fusão, codificados por ácidos nucleicos como aqui

descrito (e.g., proteínas CRSP, formas mutantes de CRSP, proteínas de fusão, etc.).

Os vectores de expressão recombinantes da invenção podem ser desenhados para expressão de CRSP em células procariotas ou eucariotas. Por exemplo, a CRSP pode ser expressa em células bacterianas tais como *E. coli*, células de insecto (utilizando vectores de expressão de baculovírus) células de levedura ou células de mamífero. As células hospedeiras adequadas estão adicionalmente discutidas em Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternativamente, o vector de expressão recombinante pode ser transcrito e traduzido *in vitro*, por exemplo utilizando sequências reguladoras do promotor de T7 e polimerase de T7.

A expressão de proteínas em procariotas é mais frequentemente realizada em *E. coli* com vectores contendo promotores constitutivos ou indutíveis que dirigem a expressão quer de proteínas de fusão quer de proteínas que não de fusão. Os vectores de fusão adicionam um número de aminoácidos a uma proteína nele codificada, usualmente no terminal amino da proteína recombinante. Estes vectores de fusão servem tipicamente três finalidades: 1) aumentar a expressão de proteína recombinante; 2) aumentar a solubilidade da proteína recombinante; e 3) auxiliar a purificação da proteína recombinante actuando como ligando em purificação de afinidade. Frequentemente, em vectores de expressão de fusão, é introduzido um local de clivagem proteolítica na junção da porção de fusão e da proteína recombinante para permitir a separação da proteína recombinante da porção de fusão subseqüentemente à purificação da proteína de fusão. Estas enzimas, e suas sequências de reconhecimento cognatas, incluem Factor Xa, trombina e enteroquinase. Os vectores de expressão de fusão típicos incluem pGEX (Pharmacia Biotech Inc; Smith, D.B. e Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) e pRIT5 (Pharmacia, Piscataway, NJ) que fundem glutathione-S-transferase (GST), proteína de ligação a maltose E, ou proteína A, respectivamente, à proteína recombinante alvo.

As proteínas de fusão purificadas podem ser utilizadas em ensaios à actividade de CRSP, em ligação ao ligando de CRSP (e.g., ensaios directos ou ensaios competitivos descritos adiante com detalhes), para gerar anticorpos específicos para proteínas CRSP, como exemplos. Numa concretização preferida, uma fusão de CRSP expressa num vector de expressão retroviral da presente invenção pode ser utilizada para infectar células de medula óssea que são subsequentemente transplantada em beneficiários irradiados. A patologia do indivíduo beneficiário é então examinada depois de passado um tempo suficiente (e.g seis (6) semanas).

Os exemplos de vectores de expressão de *E. coli*, que não de fusão, indutíveis, adequados, incluem pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) e pET11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). A expressão do gene alvo a partir do vector pTrc assenta na transcrição pela ARN-polimerase do hospedeiro a partir de um promotor de fusão trp-lac híbrido. A expressão do gene alvo a partir do vector pET11d assenta na transcrição a partir de um promotor de fusão gn10-lac de T7 mediada por uma ARN-polimerase viral co-expressa (gn1 de T7). Esta polimerase viral é fornecida por estirpes hospedeiras BL21(DE3) ou HMS174(DE3) a partir de um profago λ residente portador de um gene gn1 de T7 sob o controlo da transcrição do promotor lacUV 5.

Uma estratégia para maximizar a expressão da proteína recombinante em *E. coli* consiste em expressar a proteína numa bactéria hospedeira com uma capacidade deficitária para clivar proteoliticamente a proteína recombinante (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Outra estratégia consiste em alterar a sequência de ácido nucleico do ácido nucleico a inserir num vector de expressão de modo a que os codões individuais para cada aminoácido sejam aqueles que são preferencialmente utilizados em *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Esta alteração de sequências de ácido nucleico da invenção pode ser realizada através de técnicas *standard* de síntese de ADN.

Noutra concretização, o vector de expressão de CRSP é um vector de expressão de levedura. Os exemplos de vectores para a expressão na levedura *S. cerevisiae* incluem pYepSec1 (Baldari, et al., (1987) *Embo J* 6:229-234), pMFa (Kurjan e Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), e picZ (InVitrogen Corp, San Diego, CA).

Alternativamente, a CRSP pode ser expressa em células de insecto utilizando vectores de expressão de baculovírus. Os vectores de baculovírus disponíveis para a expressão de proteínas em células de insecto de cultura (e.g., células Sf9) incluem a série pAc (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) e a série pVL (Lucklow e Summers (1989) *Virology* 170:31-39).

Ainda noutra concretização, um ácido nucleico da invenção é expresso em células de mamífero utilizando um vector de expressão de mamífero. Os exemplos de vectores de expressão de mamífero incluem pCDM8 (Seed, B. (1987) *Nature* 329:840) e pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). Quando utilizado em células de mamífero, as funções de controlo do vector de expressão são frequentemente proporcionadas por elementos reguladores virais. Por exemplo, os promotores vulgarmente utilizados são derivados de polioma, Adenovírus 2, citomegalovírus e Vírus Símio 40. Para outros sistemas de expressão adequados tanto para células procariotas como eucariotas, vejam-se os capítulos 16 e 17 de Sambrook, J., Fritsh, E. F., e Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.*

Noutra concretização, o vector de expressão de mamífero recombinante é capaz de dirigir a expressão do ácido nucleico preferencialmente num tipo de célula particular (e.g., elementos reguladores específicos de tecidos são utilizados para expressar o ácido nucleico). Os elementos reguladores específicos de tecidos são conhecidos na especialidade. Os exemplos não limitantes de promotores específicos de tecidos adequados incluem o promotor da albumina (específico do fígado; Pinkert et al. (1987) *Genes Dev.* 1:268-277), promotores específicos de linfóides (Calame e Eaton (1988)

Adv. Immunol. 43:235-275), em particular promotores de receptores de células T (Winoto e Baltimore (1989) *EMBO J.* 8:729-733) e imunoglobulinas (Banerji et al. (1983) *Cell* 33:729-740; Queen e Baltimore (1983) *Cell* 33:741-748), promotores específicos de neurónios (e.g., o promotor do neurofilamento; Byrne e Ruddle (1989) *PNAS* 86:5473-5477), promotores específicos do pâncreas (Edlund et al. (1985) *Science* 230:912-916), e promotores específicos da glândula mamária (e.g., promotor do soro do leite; Patente U.S. 4 873 316 e Publicação de Pedido de Patente Europeia 264 166). Os promotores regulados pelo desenvolvimento estão também abrangidos, por exemplo, os promotores hox murinos (Kessel e Gruss (1990) *Science* 249:374-379) e o promotor da α -fetoproteína (Campes e Tilghman (1989) *Genes Dev.* 3:537-546).

A invenção proporciona ainda um vector de expressão recombinante compreendendo uma molécula de ADN da invenção clonada no vector de expressão numa orientação anti-sentido. Ou seja, a molécula de ADN está operativamente ligada a uma sequência reguladora de uma maneira que permite a expressão (por transcrição da molécula de ADN) de uma molécula de ARN que é anti-sentido relativamente ao ARNm de CRSP. Podem ser escolhidas sequências reguladoras operativamente ligadas a um ácido nucleico clonado na orientação anti-sentido que dirigem a expressão contínua da molécula de ARN anti-sentido numa variedade de tipos de células, por exemplo promotores virais e/ou potenciadores, ou podem ser escolhidas sequências reguladoras que dirigem a expressão constitutiva, específica do tecido ou do tipo de célula, do ARN anti-sentido. O vector de expressão anti-sentido pode estar na forma de um plasmídeo recombinante, um fagemídeo ou um vírus atenuado em que são produzidos ácidos nucleicos anti-sentido sob o controlo de uma região reguladora altamente eficiente, cuja actividade pode ser determinada pelo tipo de célula no qual o vector é introduzido. Para uma discussão da regulação da expressão génica utilizando genes anti-sentido, veja-se Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1 (1) 1986.

Outro aspecto da invenção refere-se a células hospedeiras nas quais foi introduzido um vector de expressão recombinante

da invenção. As expressões "célula hospedeira" e "célula hospedeira recombinante" são utilizadas aqui indiferentemente. Entenda-se que estas expressões se referem não apenas à célula particular em questão mas também à progénie ou progénie potencial dessa célula. Como podem ocorrer certas modificações nas gerações que se sucedem devido a mutação ou a influências ambientais, esta progénie pode, de facto, não ser idêntica à célula progenitora, mas estão ainda incluídas no âmbito da expressão como aqui utilizada.

Uma célula hospedeira pode ser qualquer célula procariota ou eucariota. Por exemplo, a proteína CRSP pode ser expressa em células bacterianas tais como *E. coli*, células de insecto, células de levedura ou células de mamífero (tais como células de ovário de *hamster* chinês (CHO) ou células COS). Outras células hospedeiras adequadas são conhecidas dos peritos na especialidade.

O ADN do vector pode ser introduzido em células procariotas ou eucariotas através de técnicas convencionais de transformação ou transfecção. Tal como aqui se utilizam, com os termos "transformação" e "transfecção" pretende-se referir uma variedade de técnicas reconhecidas na especialidade para a introdução de ácido nucleico estranho (e.g., ADN) numa célula hospedeira, incluindo a co-precipitação com fosfato de cálcio ou cloreto de cálcio, a transfecção mediada por DEAE-dextrano, a lipofecção ou a electroporação. Os métodos adequados para a transformação ou transfecção de células hospedeiras podem ser encontrados em Sambrook, et al. (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989*), e outros manuais de laboratório.

Para a transfecção estável de células de mamífero, sabe-se que, dependendo do vector de expressão e da técnica de transfecção utilizados, apenas uma pequena fracção de células pode integrar o ADN estranho no seu genoma. De modo a identificar e seleccionar esses integrantes, é geralmente introduzido um gene que codifica um marcador seleccionável (e.g., resistência a antibióticos) nas células hospedeiras, juntamente com o gene de interesse. Os marcadores seleccionáveis preferidos incluem aqueles que conferem

resistência a fármacos, tais como G418, higromicina e metotrexato. O ácido nucleico que codifica um marcador seleccionável pode ser introduzido numa célula hospedeira no mesmo vector que o que codifica a CRSP ou pode ser introduzido num vector separado. As células estavelmente transfectadas com o ácido nucleico introduzido podem ser identificadas por selecção por fármacos (e.g., células que incorporaram o gene marcador seleccionável sobreviverão, enquanto as outras células morrem).

Uma célula hospedeira da invenção, tal como uma célula hospedeira procariota ou eucariota em cultura, pode ser utilizada para produzir (*i.e.*, expressar) a proteína CRSP. Deste modo, a invenção proporciona ainda métodos para a produção de proteína CRSP utilizando as células hospedeiras da invenção. Numa concretização, o método compreende a cultura da célula hospedeira da invenção (na qual foi introduzido um vector de expressão recombinante que codifica CRSP) num meio adequado de modo a que a proteína CRSP seja produzida. Noutra concretização, o método compreende ainda o isolamento da CRSP a partir do meio ou da célula hospedeira.

As células hospedeiras da invenção podem também ser utilizadas para produzir animais transgénicos não humanos. Por exemplo, numa concretização, uma célula hospedeira da invenção é um óócito fertilizado não humano ou uma célula estaminal embrionária não humana nos quais foram introduzidas sequências que codificam CRSP. Estas células hospedeiras podem então ser utilizadas para criar animais transgénicos não humanos em que as sequências de CRSP exógenas foram introduzidas no seu genoma ou animais recombinantes homólogos em que as sequências de CRSP endógenas foram alteradas. Estes animais são úteis para estudos da função e/ou da actividade da CRSP e para a identificação e/ou avaliação de moduladores da actividade de CRSP. Tal como aqui se utiliza, um "animal transgénico" é um animal não humano, preferivelmente um mamífero, mais preferivelmente um roedor tal como um rato ou um ratinho, em que uma ou mais das células do animal incluem um transgene. Outros exemplos de animais transgénicos incluem primatas não humanos, ovelhas, cães, vacas, cabras, galinhas, anfíbios, etc. Um transgene é ADN exógeno que está integrado no genoma de uma célula a partir da qual um animal transgénico se

desenvolve e que permanece no genoma do animal maduro, dirigindo deste modo a expressão de um produto génico codificado num ou mais tipos de células ou tecidos do animal transgénico. Tal como aqui se utiliza, um "animal homólogo recombinante" é um animal não humano, preferivelmente um mamífero, mais preferivelmente um ratinho, em que um gene de CRSP endógeno foi alterado por recombinação homóloga entre o gene endógeno e uma molécula de ADN exógena introduzida numa célula do animal, e.g., uma célula embrionária do animal, antes do desenvolvimentos do animal.

Um animal transgénico não humano da invenção pode ser criado introduzindo ácido nucleico que codifica CRSP nos pronúcleos masculinos de um oócito fertilizado, e.g., por micro-injecção, infecção retroviral e permitindo que o oócito se desenvolva num animal hospedeiro fêmea pseudo-grávida. A sequência de ADNc de CRSP humana da invenção pode ser introduzida na forma de um transgene no genoma de um animal não humano. Alternativamente, um homólogo não humano de um gene de CRSP humana, tal como um gene de CRSP de ratinho, pode ser isolado com base na hibridação com o ADNc de CRSP humana (adicionalmente descrita na subsecção I anterior) e pode ser utilizado como transgene. Podem também ser incluídos no transgene sequências intrónicas e sinais de poliadenilação para aumentar a eficiência da expressão do transgene. Uma ou mais sequências reguladoras específicas de tecidos podem ser operativamente ligadas ao transgene de CRSP para dirigir a expressão da CRSP proteína para células particulares. Os métodos para criar animais transgénicos não humanos por via de manipulação de embriões e micro-injecção, particularmente animais tais como ratinhos, tornaram-se convencionais na especialidade e estão descritos, por exemplo, nas Patentes U.S. 4 736 866 e 4 870 009, ambas de Leder *et al.*, na Patente U.S. 4 873 191 de Wagner *et al.*, e em Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Utilizam-se métodos similares para a produção de outros animais transgénicos não humanos. Um animal fundador transgénico não humano pode ser identificado com base na presença do transgene de CRSP no seu genoma e/ou na expressão do ARNm de CRSP em tecidos ou células dos animais. Um animal fundador transgénico não humano pode então ser utilizado para criar mais animais portadores do transgene.

Adicionalmente, os animais transgênicos não humanos portadores de um transgene que codifica CRSP podem ainda ser procriados com outros animais transgênicos não humanos portadores de outros transgenes.

Para criar um animal não humano recombinante homólogo, prepara-se um vector que contém pelo menos uma porção de um gene de CRSP no qual foi introduzida uma deleção, adição ou substituição para assim alterar, e.g., corromper funcionalmente, o gene de CRSP. O gene de CRSP pode ser um gene humano (e.g., o ADNC de SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:9 ou SEQ ID NO:12), mas mais preferivelmente, é um homólogo não humano de um gene de CRSP humana. Por exemplo, um gene de CRSP de ratinho de SEQ ID NO:16 pode ser utilizado para construir um vector de recombinação homóloga adequado para alterar um gene de CRSP endógeno no genoma do ratinho. Numa concretização preferida, o vector é desenhado de modo a que, após recombinação homóloga, o gene de CRSP endógeno seja funcionalmente corrompido (i.e., deixe de codificar uma proteína funcional; também referido como um vector "knockout"). Alternativamente, o vector pode ser desenhado de modo a que, após recombinação homóloga, o gene de CRSP endógeno seja mutado ou de outro modo alterado mas codifique ainda a proteína funcional (e.g., a região reguladora a montante pode ser alterada para desse modo alterar a expressão da proteína CRSP endógena). No vector de recombinação homóloga, a porção alterada do gene de CRSP está flanqueada nas suas extremidades 5' e 3' por ácido nucleico adicional do gene de CRSP para permitir que ocorra recombinação homóloga entre o gene de CRSP exógeno transportado pelo vector e um gene de CRSP endógeno numa célula estaminal embrionária não humana. O ácido nucleico de CRSP flanqueador adicional tem um comprimento suficiente para uma recombinação homóloga bem sucedida com o gene endógeno. Tipicamente, são incluídas no vector várias quilobases de ADN flanqueador (em ambas as extremidades, 5' e 3') (veja-se e.g., Thomas, K.R. e Capecchi, M.R. (1987) *Cell* 51:503 para uma descrição de vectores de recombinação homóloga). O vector é introduzido numa linha de células estaminais embrionárias não humanas (e.g., por electroporação) e as células em que o gene de CRSP introduzido sofreu recombinação homóloga com o gene de CRSP endógeno são seleccionadas (veja-se e.g., Li, E. et al., (1992) *Cell*

69:915). As células seleccionadas são então injectadas num blastocisto de um animal não humano (e.g., um ratinho) para formar quimeras de agregação (veja-se e.g., Bradley, A. em *Teratocarcinomas and Embryonic Stem Cells A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152. Um embrião quimérico não humano pode então ser implantado num animal hospedeiro não humano fêmea pseudo-grávida adequada e o embrião é levado a termo. A progénie portadora do ADN homologamente recombinado nas suas células germinais pode ser utilizada para criar animais em que todas as células do animal contêm o ADN homologamente recombinado por transmissão de linha germinal do transgene. Os métodos para a construção de vectores de recombinação homóloga e animais recombinantes homólogos estão descritos adicionalmente em Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 e na Publicação Internacional PCT com os N^{os}.: WO 90/11354 de Le Mouellec et al.; WO 91/01140 de Smithies et al.; WO 92/0968 de Zijlstra et al.; e WO 93/04169 de Berns et al.

Noutra concretização, podem ser produzidos animais transgénicos não humanos que contêm sistemas seleccionados que permitem a expressão regulada do transgene. Um exemplo de um destes sistemas é o sistema de recombinase *cre/loxP* do bacteriófago P1. Para uma descrição do sistema de recombinase *cre/loxP*, veja-se, e.g., Lakso et al. (1992) *PNAS* 89:6232-6236. Outro exemplo de um sistema de recombinase é o sistema de recombinase FLP de *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. Se for utilizado um sistema de recombinase *cre/loxP* para regular a expressão do transgene, são necessários animais contendo transgenes que codificam tanto a recombinase *Cre* como uma proteína seleccionada. Estes animais podem ser proporcionados através da construção de animais transgénicos "duplos" não humanos, e.g., acasalando dois animais transgénicos não humanos, um contendo um transgene que codifica uma proteína seleccionada e o outro contendo um transgene que codifica uma recombinase.

Clones dos animais transgénicos não humanos aqui descritos podem também ser produzidos de acordo com os métodos descritos em Wilmut, I. et al. (1997) *Nature* 385:810-813. Em resumo, uma célula, e.g., uma célula somática, do animal transgénico, pode ser isolada e induzida para sair do ciclo de

crescimento e entrar em fase G_0 . A célula quiescente pode então ser fundida, *e.g.*, através da utilização de pulsos eléctricos, com um oócito enucleado não humano de um animal da mesma espécie da qual é isolada a célula quiescente. O oócito não humano reconstruído é depois cultivado de modo a desenvolver-se em mórula ou blastócito e depois é transferido para o animal hospedeiro fêmea pseudo-grávida. A descendência nascida deste animal hospedeiro fêmea será um clone do animal do qual a célula, *e.g.*, a célula somática, é isolada.

IV. Composições Farmacêuticas

As moléculas de ácido nucleico de CRSP, as proteínas CRSP e os anticorpos anti-CRSP (aqui também referidos como "compostos activos") da invenção podem ser incorporados em composições farmacêuticas adequadas para administração. Estas composições tipicamente compreendem a molécula de ácido nucleico, a proteína ou o anticorpo e um transportador farmacêuticamente aceitável. Tal como aqui se utiliza, na expressão "transportador farmacêuticamente aceitável" pretende-se incluir qualquer um e todos entre solventes, meios de dispersão, revestimentos, agentes antibacterianos e antifúngicos, agentes de retardamento da absorção e isotónicos, e semelhantes, compatíveis com a administração farmacêutica. A utilização destes meios e agentes para substâncias farmacêuticamente activas é bem conhecida na especialidade. Excepto no caso de qualquer meio ou agente convencional ser incompatível com o composto activo, a sua utilização nas composições está contemplada. Compostos activos suplementares podem também ser incorporados nas composições.

Uma composição farmacêutica da invenção é formulada para ser compatível com a via de administração a que se destina. Os exemplos de vias de administração incluem administração parentérica, *e.g.*, intravenosa, intradérmica, subcutânea, oral (*e.g.*, inalação), transdérmica (tópica), transmucosa e rectal. As soluções ou suspensões utilizadas para aplicação parentérica, intradérmica ou subcutânea podem incluir os seguintes componentes: um diluente estéril tal como água para injeção, solução salina, óleos fixados, polietilenoglicóis, glicerina, propilenoglicol ou outros solventes sintéticos; agentes antibacterianos tais como álcool benzílico ou

metilparabenos; antioxidantes tais como ácido ascórbico ou bissulfito de sódio; agentes quelantes tais como ácido etilenodiaminotetra-acético; tampões tais como acetatos, citratos ou fosfatos e agentes para o ajuste de tonicidade tais como cloreto de sódio ou dextrose. O pH pode ser ajustado com ácidos ou bases, tais como ácido clorídrico ou hidróxido de sódio. A preparação parentérica pode ser encerrada em ampolas, seringas descartáveis ou frascos de doses múltiplas feitos de vidro ou plástico.

As composições farmacêuticas adequadas para utilização injectável incluem soluções aquosas estéreis (quando solúveis em água) ou dispersões e pós estéreis para a preparação extemporânea de soluções ou dispersões injectáveis estéreis. Para administração intravenosa, os transportadores adequados incluem solução salina fisiológica, água bacteriostática, Cremophor EL[™] (BASF, Parsippany, NJ) ou solução salina tamponada com fosfato (PBS). Em todos os casos, a composição tem que ser estéril e deve ser fluida na medida em que seja fácil a sua aplicação por seringa. Tem que ser estável sob as condições de fabrico e armazenagem e tem que ser conservada contra a acção de contaminação por microorganismos tais como bactérias e fungos. O transportador pode ser um solvente ou um meio de dispersão contendo, por exemplo, água, etanol, poliol (por exemplo, glicerol, propilenoglicol e polietilenoglicol líquido, e semelhantes), e suas misturas adequadas. A fluidez correcta pode ser mantida, por exemplo, pela utilização de um revestimento tal como lecitina, pela manutenção da dimensão de partícula necessária no caso de dispersão e pela utilização de tensioactivos. A prevenção da acção de microorganismos pode ser conseguida através de vários agentes antibacterianos e antifúngicos, por exemplo, parabenos, clorobutanol, fenol, ácido ascórbico, timerosal, e semelhantes. Em muitos casos, será preferível incluir na composição agentes isotónicos, por exemplo, açúcares, poliálcoois tais como manitol, sorbitol, cloreto de sódio. A absorção prolongada das composições injectáveis pode ser conseguida incluindo na composição um agente que retarda a absorção, por exemplo, monoestearato de alumínio e gelatina.

As soluções injectáveis estéreis podem ser preparadas por incorporação do composto activo (e.g., uma proteína CRSP ou um

anticorpo anti-CRSP) na quantidade necessária, num solvente apropriado, com um ingrediente ou uma combinação de ingredientes atrás enumerados, conforme necessário, seguida por esterilização por filtração. Genericamente, as dispersões são preparadas por incorporação do composto activo num veículo estéril que contém um meio de dispersão básico e os outros ingredientes necessários dos atrás enumerados. No caso de pós estéreis para a preparação de soluções injectáveis estéreis, os métodos de preparação preferidos são a secagem sob vácuo e a criodessecagem que originam um pó do ingrediente activo mais qualquer ingrediente adicional desejado a partir de uma sua solução previamente esterilizada por filtração.

As composições orais incluem geralmente um diluente inerte ou um transportador edível. Podem ser encerradas em cápsulas de gelatina ou prensadas em comprimidos. Para fins de administração terapêutica oral, o composto activo pode ser incorporado com excipientes e utilizado na forma de comprimidos, trociscos ou cápsulas. As composições orais podem também ser preparadas utilizando um transportador fluido para utilização como colutório, em que o composto no transportador fluido é aplicado oralmente e bochechado e expectorado ou deglutido. Podem ser incluídos agentes de ligação farmacologicamente compatíveis, e/ou materiais adjuvantes, como parte da composição. Os comprimidos, pílulas, cápsulas, trociscos e semelhantes, podem conter qualquer um dos seguintes ingredientes, ou compostos de natureza similar: um aglutinante tal como celulose microcristalina, goma adragante ou gelatina; um excipiente tal como amido ou lactose, um agente desintegrante tal como ácido algínico, Primogel ou amido de milho; um lubrificante tal como estearato de magnésio ou Sterotes; um deslizador tal como dióxido de silício coloidal; um agente edulcorante tal como sacarose ou sacarina; ou um agente aromatizante tal como hortelã-pimenta, salicilato de metilo ou aroma de laranja.

Para administração por inalação, os compostos são entregues na forma de uma pulverização de aerossol a partir de um recipiente ou dispensador pressurizados que contém um propulsor adequado, e.g., um gás tal como dióxido de carbono, ou um nebulizador.

A administração sistémica pode também ser por meios transmucosos ou transdérmicos. Para administração transmucosa ou transdérmica, utilizam-se na formulação penetrantes apropriados para a barreira a permear. Estes penetrantes são geralmente conhecidos na arte, e incluem, por exemplo, para administração transmucosa, detergentes, sais biliares e derivados de ácido fusídico. A administração transmucosa pode ser conseguida através da utilização de pulverizações nasais ou supositórios. Para administração transdérmica, os compostos activos são formulados em unguentos, pomadas, géis ou cremes como é geralmente conhecido na especialidade.

Os compostos podem também ser preparados na forma de supositórios (*e.g.*, com bases convencionais para supositórios tais como manteiga de cacau e outros glicéridos) ou enemas de retenção para entrega rectal.

Numa concretização, os compostos activos são preparados com transportadores que irão proteger o composto contra a rápida eliminação pelo corpo, tal como uma formulação de libertação controlada, incluindo implantes e sistemas de entrega microencapsulados. Podem utilizar-se polímeros biocompatíveis, biodegradáveis, tais como etileno-acetato de vinilo, polianidridos, ácido poliglicólico, colagénio, poliortoésteres e ácido poliláctico. Os métodos para preparação destas formulações serão evidentes para os peritos na especialidade. Os materiais podem também ser obtidos comercialmente de Alza Corporation e Nova Pharmaceuticals, Inc. Podem também ser utilizadas suspensões lipossómicas (incluindo lipossomas direccionados para células infectadas com anticorpos monoclonais para antigénios virais) como transportadores farmacêuticamente aceitáveis. Estas podem ser preparadas de acordo com métodos conhecidos dos peritos na especialidade, por exemplo, como descrito na Patente U.S. 4 522 811.

É especialmente vantajoso formular composições orais ou parentéricas na forma de dosagem unitária para facilidade de administração e uniformidade de dosagem. Forma de dosagem unitária, como aqui se utiliza, refere-se a unidades fisicamente discretas adequadas como dosagens unitárias para o indivíduo a tratar; cada unidade contendo uma quantidade

predeterminada de composto activo calculada para produzir o efeito terapêutico desejado em associação com o transportador farmacêutico requerido. As especificações para as formas de dosagem unitária da invenção são ditadas por, e directamente dependentes, das características únicas do composto activo e do efeito terapêutico particular a conseguir, e das limitações inerentes na especialidade da preparação de composições com esse composto activo para o tratamento de indivíduos.

A toxicidade e a eficácia terapêutica destes compostos podem ser determinadas por procedimentos farmacêuticos *standard* em culturas celulares ou animais experimentais, *e.g.*, para determinação da LD50 (a dose letal para 50% da população) e a ED50 (a dose terapeuticamente eficaz em 50% da população). A razão de doses entre os efeitos tóxico e terapêutico é o índice terapêutico e pode ser expresso como a razão LD50/ED50. São preferidos os compostos que exibem índices terapêuticos elevados. Embora se possam utilizar compostos que exibem efeitos secundários tóxicos, deve ter-se o cuidado de desenhar um sistema de entrega que direcione esses compostos para o local do tecido afectado de modo a minimizar potenciais danos em células não infectadas e, desse modo, reduzir os efeitos secundários.

Os resultados obtidos a partir dos ensaios de culturas celulares e estudos em animais podem ser utilizados na formulação de uma gama de dosagem para utilização em seres humanos. A dosagem destes compostos assenta preferivelmente numa gama de concentrações circulantes que inclui a ED50 com pouca ou nenhuma toxicidade. A dosagem pode variar dentro desta gama dependendo da forma de dosagem empregue e da via de administração utilizada. Para qualquer composto utilizado no método da invenção, a dose terapeuticamente eficaz pode ser estimada inicialmente a partir de ensaios de culturas celulares. Uma dose pode ser formulada em modelos animais para se conseguir uma gama de concentrações plasmáticas circulantes que inclui a IC50 (*i.e.*, a concentração do composto de teste que atinge uma inibição semi-máxima dos sintomas) como determinado em cultura celular. Esta informação pode ser utilizada para determinar mais precisamente as doses úteis em seres humanos. Os níveis no plasma podem ser medidos, por exemplo, por cromatografia líquida de alta resolução.

As moléculas de ácido nucleico da invenção podem ser inseridas em vectores e utilizadas como vectores de terapia génica. Os vectores de terapia génica podem ser entregues a um indivíduo por, por exemplo, injeção intravenosa, administração local (veja-se a Patente U.S. 5 328 470) ou por injeção estereotáctica (veja-se e.g., Chen et al. (1994) *PNAS* 91:3054-3057). A preparação farmacêutica do vector de terapia génica pode incluir o vector de terapia génica num diluente aceitável, ou pode compreender uma matriz de libertação lenta em que o veículo de entrega dos genes está implantado. Alternativamente, quando o vector de entrega de genes completo pode ser produzido intacto a partir de células recombinantes, e.g., vectores retrovirais, a preparação farmacêutica pode incluir uma ou mais células que produzem o sistema de entrega de genes.

As composições farmacêuticas podem ser incluídas num recipiente, uma embalagem ou um dispensador, juntamente com instruções para administração.

V. Utilizações e Métodos da Invenção

As moléculas da presente invenção (e.g., moléculas de ácido nucleico, proteínas, homólogos de proteínas e anticorpos, aqui descritos) podem ser utilizadas num ou mais dos seguintes métodos: a) ensaios de rastreio; b) medicina preditiva (e.g., ensaios de diagnóstico, ensaios de prognóstico, ensaios clínicos de monitoração e farmacogenética); e c) métodos de tratamento (e.g., terapêuticos e profilácticos). Tal como aqui descrito, uma proteína CRSP da invenção possui uma ou mais das seguintes actividades: cálcio intracelular, um aumento de fosfatidilinositol ou outra molécula, e pode resultar, e.g., em fosforilação de proteínas específicas, uma modulação da transcrição génica e qualquer das outras actividades biológicas aqui estabelecidas.

Numa concretização preferida, uma actividade de CRSP é pelo menos uma ou mais das seguintes actividades: (i) interacção de uma proteína CRSP com, e/ou ligação a, uma segunda molécula, (e.g., uma proteína, tal como um receptor de CRSP (e.g., CRSP-1), uma forma solúvel de um receptor de CRSP,

um receptor para um membro da família *wnt* de proteínas de sinalização, ou uma molécula de sinalização que não uma CRSP); (ii) interação de uma proteína CRSP com uma proteína intracelular por via de um receptor de CRSP ligado à membrana; (iii) formação de um complexo entre uma proteína CRSP solúvel e um segundo parceiro de ligação de CRSP solúvel (e.g., uma molécula que não uma proteína CRSP ou uma segunda molécula de proteína CRSP); (iv) interação com outras proteínas extracelulares (e.g., regulação da adesão celular dependente de *wnt* a componentes da matriz extracelular); (v) ligação a, e eliminação de, uma molécula indesejável (e.g., uma actividade desintoxicante ou função de defesa); e/ou (vi) uma actividade enzimática, e pode assim ser utilizada em, por exemplo, (1) modulação da transdução de sinais celulares, quer *in vitro* quer *in vivo* (e.g., antagonismo da actividade de membros da família *wnt* de proteínas segregadas ou supressão da transdução de sinal dependente de *wnt*); (2) regulação da comunicação entre células (e.g., regulação de interações célula-célula dependentes de *wnt*); (3) regulação da expressão de genes cuja expressão é modulada por ligação de CRSP (e.g., CRSP-1) a um receptor; (4) regulação da transcrição génica numa célula envolvida no desenvolvimento ou diferenciação, quer *in vitro* quer *in vivo* (e.g., indução de diferenciação celular); (5) regulação da transcrição génica numa célula envolvida no desenvolvimento ou diferenciação, em que pelo menos um gene codifica uma proteína específica da diferenciação; (6) regulação da transcrição génica numa célula envolvida no desenvolvimento ou na diferenciação, em que pelo menos um gene codifica uma segunda proteína segregada; (7) regulação da transcrição génica numa célula envolvida no desenvolvimento ou na diferenciação, em que pelo menos um gene codifica uma molécula de transdução de sinal; (8) regulação da proliferação celular, quer *in vitro* quer *in vivo* (e.g., indução da proliferação celular ou inibição da proliferação como no caso da supressão da tumorigénese (e.g., supressão da formação de glioblastomas)); (9) formação e manutenção de arranjos espaciais ordenados de tecidos diferenciados em vertebrados, tanto adultos como embrionários (e.g., indução da formação da cabeça durante o desenvolvimento dos vertebrados ou manutenção de células progenitoras hematopoiéticas); (10) modulação da morte celular, tal como estimulação da sobrevivência celular;

(11) regulação da migração celular; e/ou (12) modulação imunitária.

De acordo com uma concretização, a presente invenção envolve um método de utilização (e.g., um ensaio de diagnóstico, um ensaio de prognóstico ou um método de tratamento profilático/terapêutico) em que é utilizada uma molécula da presente invenção (e.g., uma proteína CRSP, um ácido nucleico de CRSP ou um modulador de CRSP), por exemplo, para diagnosticar, prognosticar e/ou tratar uma doença e/ou condição em que está indicada qualquer das actividades atrás mencionadas (i.e., actividades (i)-(vi) e (1)-(12) no parágrafo anterior). Noutra concretização, a presente invenção envolve um método de utilização (e.g., um ensaio de diagnóstico, um ensaio de prognóstico ou o fabrico de um medicamento para um método de tratamento profilático/terapêutico) em que é utilizada uma molécula da presente invenção (e.g. uma proteína CRSP, um ácido nucleico de CRSP ou um modulador de CRSP), por exemplo, no fabrico de um medicamento para o diagnóstico, prognóstico e/ou tratamento de indivíduos, preferivelmente um indivíduo humano, em que qualquer das actividades atrás mencionadas está patologicamente perturbada. Numa concretização preferida, os métodos de utilização (e.g., ensaios de diagnóstico, ensaios de prognóstico ou fabrico de medicamentos para métodos de tratamento profiláticos/terapêuticos) envolvem a produção de um medicamento para administração a um indivíduo, preferivelmente um indivíduo humano, de uma molécula da presente invenção (e.g., uma proteína CRSP, um ácido nucleico de CRSP ou um modulador de CRSP) para o diagnóstico, prognóstico e/ou tratamento terapêutico. Noutra concretização, os métodos de utilização (e.g., ensaios de diagnóstico, ensaios de prognóstico ou fabrico de medicamentos para métodos de tratamento profiláticos/terapêuticos) envolvem a produção de um medicamento para administração a um indivíduo humano de uma molécula da presente invenção (e.g., uma proteína CRSP, um ácido nucleico de CRSP ou um modulador de CRSP).

Outras concretizações da invenção referem-se à utilização de moléculas de ácido nucleico isoladas da invenção que podem ser utilizadas, por exemplo, para expressar uma proteína CRSP (e.g., por via de um vector de expressão recombinante numa

célula hospedeira em aplicações de terapia génica), para detectar ARNm de CRSP (e.g., numa amostra biológica) ou uma alteração genética num gene de CRSP, e para modular a actividade de CRSP, como descrito adicionalmente adiante. Em adição, as proteínas CRSP podem ser utilizadas para rastrear fármacos ou compostos que modulam a actividade de CRSP assim como para tratar desordens caracterizadas por produção insuficiente ou excessiva de proteína CRSP ou produção de formas da proteína CRSP que possuem actividade diminuída ou aberrante em comparação com a proteína CRSP de tipo selvagem (e.g., desordens de desenvolvimento ou doenças proliferativas tais como cancro, assim como doenças, condições ou desordens caracterizadas por diferenciação e/ou sobrevivência celulares anómalas, uma estrutura extracelular anómala ou uma anomalia num mecanismo de defesa). Adicionalmente, os anticorpos anti-CRSP da invenção podem ser utilizados para detectar e isolar proteínas CRSP, regular a biodisponibilidade de proteínas CRSP, e modular a actividade de CRSP. A expressão "uma actividade aberrante", aplicada a uma actividade de uma proteína tal como CRSP (e.g., CRSP-1), refere-se a uma actividade que difere da actividade da proteína de tipo selvagem ou nativa ou que difere da actividade da proteína num indivíduo saudável. Uma actividade de uma proteína pode ser aberrante porque é mais forte do que a actividade da sua contrapartida nativa. Alternativamente, uma actividade pode ser aberrante por ser mais fraca ou estar ausente relativamente à actividade da sua contrapartida nativa. Uma actividade aberrante pode também ser uma alteração numa actividade. Por exemplo uma proteína aberrante pode interagir com uma proteína diferente relativamente à sua contrapartida nativa. Uma célula pode ter uma actividade aberrante de CRSP (e.g., CRSP-1) devido a super-expressão ou sub-expressão do gene que codifica a CRSP.

A. Ensaios de Rastreo:

A invenção proporciona um método (também referido aqui como um "ensaio de rastreo") para a identificação de moduladores, i.e., compostos de teste ou candidatos ou agentes (e.g., péptidos, peptidomiméticos, moléculas pequenas ou outros fármacos) que se ligam a proteínas CRSP ou possuem um efeito estimulador ou inibidor sobre, por exemplo, a expressão

de CRSP ou a actividade de CRSP. Os moduladores podem incluir, por exemplo, agonistas de CRSP e/ou antagonistas de CRSP. Com o termo "agonista", tal como aqui se utiliza, pretende-se referir um agente que imita ou supra-regula (e.g. potencia ou suplementa) a bioactividade de uma CRSP (e.g., CRSP-1). Um agonista de CRSP pode ser um composto que imita a bioactividade de uma proteína CRSP, tal como transdução de um sinal a partir de um receptor de CRSP, por, e.g., interacção com um receptor de CRSP-1. Um agonista de CRSP pode também ser um composto que supra-regula a expressão de um gene de CRSP. Um agonista de CRSP pode também ser um composto que modula a expressão ou a actividade de uma proteína que está localizada a jusante de um receptor de CRSP, desse modo imitando ou aumentando o efeito da ligação de CRSP a um receptor de CRSP.

"Antagonista", como aqui se utiliza, pretende referir um agente que inibe, diminui ou suprime a bioactividade de uma CRSP (e.g., CRSP-1). Um antagonista pode ser um composto que diminui a sinalização de uma proteína CRSP, e.g., um composto que é capaz de se ligar a CRSP-1 ou a um receptor de CRSP-1. Um antagonista de CRSP preferido inibe a interacção entre uma proteína CRSP e outra molécula, tal como um receptor de CRSP. Alternativamente, um antagonista de CRSP pode ser um composto que infra-regula a expressão a um gene de CRSP. Um antagonista de CRSP pode também ser um composto que modula a expressão ou a actividade de uma proteína que está localizada a jusante de um receptor de CRSP, desse modo antagonizando o efeito da ligação de CRSP a um receptor de CRSP.

Numa concretização, a invenção proporciona ensaios para o rastreio de compostos de teste ou candidatos que se ligam a, ou modulam a actividade de, uma proteína ou polipéptido de CRSP ou uma sua porção biologicamente activa. Noutra concretização, a invenção proporciona ensaios para o rastreio de compostos de teste ou candidatos que se ligam a, ou modulam a actividade de, um receptor de CRSP. Os compostos de teste da presente invenção podem ser obtidos utilizando qualquer uma das numerosas abordagens em métodos de bibliotecas combinatórias conhecidos na especialidade, incluindo: bibliotecas biológicas; bibliotecas em fase de solução ou em fase sólida paralela espacialmente abordáveis; métodos de bibliotecas sintéticas que requerem desenrolamento; o método

da biblioteca 'uma conta, um composto'; e métodos de bibliotecas sintéticas utilizando selecção por cromatografia de afinidade. A abordagem de bibliotecas biológicas está limitada a bibliotecas peptídicas, enquanto as outras quatro abordagens são aplicáveis a bibliotecas de compostos peptídicos, oligoméricos não peptídicos ou de molécula pequena, (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Os exemplos de métodos para a síntese de bibliotecas moleculares podem ser encontrados na especialidade, por exemplo em: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al. (1994). *J Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; e em Gallop et al. (1994) *J Med. Chem.* 37:1233.

As bibliotecas de compostos podem ser apresentadas em solução (e.g., Houghten (1992) *BioTechniques* 13:412-421), ou sobre contas (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bactérias (Ladner USP 5 223 409), esporos (Ladner USP '409), plasmídeos (Cull et al. (1992) *Proc Natl Acad Sci USA* 89:1865-1869) ou sobre fagos (Scott e Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla et al. (1990) *Proc. Natl. Acad Sci.* 87:6378-6382); (Felici (1991) *J Mol. Biol.* 222:301-310); (Ladner supra.).

Numa concretização, um ensaio é um ensaio à base de células em que uma célula que expressa um receptor de CRSP sobre a superfície celular é colocado em contacto com um composto de teste e a capacidade do composto de teste para se ligar a um receptor de CRSP é determinada. A célula, por exemplo, pode ser de origem mamífera ou uma célula de levedura. A determinação da capacidade do composto de teste para se ligar a um receptor de CRSP pode ser realizada, por exemplo, por acoplamento do composto de teste com um radioisótopo ou um marcador enzimático de modo a que a ligação do composto de teste ao receptor de CRSP possa ser determinada por detecção do composto marcado num complexo. Por exemplo, os compostos de teste podem ser marcados com ^{125}I , ^{35}S , ^{14}C ou ^3H ,

directa ou indirectamente, e o radioisótopo pode ser detectado por contagem directa da radioemissão ou por contagem de cintilações. Alternativamente, os compostos de teste podem ser marcados enzimaticamente com, por exemplo, peroxidase de rábano, fosfatase alcalina ou luciferase, e o marcador enzimático pode ser detectado por determinação da conversão de um substrato apropriado, em produto.

Está também no âmbito da presente invenção determinar a capacidade de um composto de teste para interagir com um receptor de CRSP sem a marcação de nenhum dos intervenientes. Por exemplo, pode utilizar-se um microfisiómetro para detectar a interacção de um composto de teste com um receptor de CRSP sem a marcação, nem do composto de teste nem do receptor. McConnell, H.M. *et al.* (1992) *Science* 257:1906-1912. Tal como aqui se utiliza, um "microfisiómetro" (e.g., Cytosensor™) é um instrumento analítico que mede a taxa a que uma célula acidifica o seu ambiente utilizando um sensor potenciométrico abordável por luz (LAPS). As alterações nesta taxa de acidificação podem ser utilizadas como um indicador da interacção entre o ligando e o receptor.

Numa concretização preferida, o ensaio compreende o contacto de uma célula que expressa um receptor de CRSP sobre a superfície celular com uma proteína CRSP ou uma sua porção biologicamente activa, para formar uma mistura de ensaio, o contacto da mistura de ensaio com um composto de teste, e a determinação da capacidade do composto de teste para interagir com um receptor de CRSP, em que a determinação da capacidade do composto de teste para interagir com um receptor de CRSP compreende a determinação da capacidade do composto de teste para se ligar preferencialmente ao receptor de CRSP em comparação com a capacidade da CRSP, ou de uma sua porção biologicamente activa, se ligar ao receptor.

Noutra concretização, um ensaio é um ensaio à base de células compreendendo o contacto de uma célula que expressa uma molécula alvo de CRSP com um composto de teste e a determinação da capacidade do composto de teste para modular (e.g. estimular ou inibir) a actividade da molécula alvo de CRSP. A determinação da capacidade do composto de teste para modular a actividade de uma molécula alvo de CRSP pode ser

conseguida, por exemplo, por determinação da capacidade da proteína CRSP se ligar a, ou interagir com, a molécula alvo de CRSP.

A determinação da capacidade da proteína CRSP se ligar ou interagir com uma molécula alvo de CRSP pode ser realizada através de um dos métodos atrás descritos para a determinação da ligação directa. Numa concretização preferida, a determinação da capacidade da proteína CRSP se ligar a, ou interagir com, uma molécula alvo de CRSP, pode ser realizada por determinação da actividade da molécula alvo. Por exemplo, a actividade da molécula alvo pode ser determinada por detecção da indução de um segundo mensageiro celular do alvo (*i.e.* Ca^{2+} intracelular, diacilglicerol, IP_3 , etc.), detecção da actividade catalítica/enzimática do alvo para um substrato apropriado, detecção da indução de um gene repórter (compreendendo um elemento regulador responsivo a CRSP operativamente ligado a um ácido nucleico que codifica um marcador detectável, *e.g.*, luciferase), ou detecção de uma resposta celular, por exemplo, desenvolvimento, diferenciação ou taxa de proliferação.

Ainda noutra concretização, um ensaio da presente invenção é um ensaio isento de células em que uma proteína CRSP ou uma sua porção biologicamente activa é colocada em contacto com um composto de teste e a capacidade do composto de teste para se ligar à proteína CRSP ou a uma sua porção biologicamente activa é determinada. A ligação do composto de teste à proteína CRSP pode ser determinada quer directa quer indirectamente como atrás descrito. Numa concretização preferida, o ensaio inclui o contacto da proteína CRSP ou de uma sua porção biologicamente activa com um composto conhecido que se liga a CRSP para formar uma mistura de ensaio, o contacto da mistura de ensaio com um composto de teste, e a determinação da capacidade do composto de teste para interagir com uma proteína CRSP, em que a determinação da capacidade do composto de teste para interagir com uma proteína CRSP compreende a determinação da capacidade do composto de teste para se ligar preferencialmente a CRSP ou a uma sua porção biologicamente activa em comparação com o composto conhecido.

Noutra concretização, o ensaio é um ensaio isento de células em que uma proteína CRSP ou uma sua porção biologicamente activa é colocada em contacto com um composto de teste e a capacidade do composto de teste para modular (e.g., estimular ou inibir) a actividade da proteína CRSP ou da sua porção biologicamente activa é determinada. A determinação da capacidade do composto de teste para modular a actividade de uma proteína CRSP pode ser realizada, por exemplo, por determinação da capacidade da proteína CRSP se ligar a uma molécula alvo de CRSP através de um dos métodos atrás descritos para a determinação da ligação directa. A determinação da capacidade da proteína CRSP se ligar a uma molécula alvo de CRSP pode também ser realizada utilizando uma tecnologia tal como a Análise de Interação Biomolecular (BIA) em tempo real. Sjolander, S. e Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 e Szabo et al. (1995) *Curr. Opin. Struct. Biol.* 5:699-705. Tal como aqui se utiliza, "BIA" é uma tecnologia para o estudo de interações bioespecíficas em tempo real, sem marcação de qualquer dos intervenientes (e.g., BIAcore™). Podem utilizar-se alterações na ressonância plasmónica de superfície (SPR) de fenómeno óptico como uma indicação de reacções em tempo real entre moléculas biológicas.

Numa concretização alternativa, a determinação da capacidade do composto de teste para modular a actividade de uma proteína CRSP pode ser conseguida por determinação da capacidade da proteína CRSP para adicionalmente modular a actividade de uma molécula alvo de CRSP. Por exemplo, a actividade catalítica/enzimática da molécula alvo sobre um substrato apropriado pode ser determinada como anteriormente descrito.

Ainda noutra concretização, o ensaio isento de células envolve o contacto de uma proteína CRSP ou uma sua porção biologicamente activa com um composto conhecido que se liga à proteína CRSP para formar uma mistura de ensaio, o contacto da mistura de ensaio com um composto de teste, e a determinação da capacidade do composto de teste para interagir com a proteína CRSP, em que a determinação da capacidade do composto de teste para interagir com a proteína CRSP compreende a determinação da capacidade da proteína CRSP para se ligar

preferencialmente a, ou modular a actividade de, uma molécula alvo de CRSP.

Em muitos programas de rastreio de fármacos que testam bibliotecas de compostos e extractos naturais, são desejáveis ensaios de alto rendimento para maximizar o número de compostos avaliados num determinado período de tempo. Ensaios que são realizados em sistemas isentos de células, tais como os que podem ser derivados com proteínas purificadas ou semi-purificadas, são frequentemente preferidos como rastreios "primários" na medida em que podem ser gerados para permitir um desenvolvimento rápido e uma relativamente fácil detecção de uma alteração num alvo molecular que é mediada por um composto de teste. Adicionalmente, os efeitos de toxicidade celular e/ou biodisponibilidade do composto de teste, podem ser geralmente ignorados no sistema *in vitro*, sendo em vez disso o ensaio focado principalmente no efeito do fármaco sobre o alvo molecular como pode ser manifesto numa alteração de afinidade de ligação com elementos a montante ou a jusante. Deste modo, num ensaio de rastreio exemplificativo da presente invenção, o composto de interesse é colocado em contacto com uma proteína CRSP (e.g., CRSP-1) ou um parceiro de ligação de CRSP (e.g., CRSP-1), e.g., um receptor. O receptor pode ser solúvel ou o receptor pode estar presente na superfície de uma célula. À mistura do composto e da proteína CRSP ou do parceiro de ligação de CRSP é então adicionada uma composição contendo um parceiro de ligação de CRSP ou uma proteína CRSP, respectivamente. A detecção e a quantificação de complexos de proteínas CRSP e parceiros de ligação de CRSP proporcionam um meio para determinação da eficácia de um composto na inibição (ou potenciação) da formação de complexos entre CRSP e um parceiro de ligação. A eficácia do composto pode ser avaliada gerando curvas de resposta à dose a partir de resultados obtidos utilizando várias concentrações do composto de teste. Adicionalmente, pode também ser realizado um ensaio de controlo para proporcionar uma linha de base para comparação. No ensaio de controlo, o polipéptido de CRSP isolado e purificado ou o parceiro de ligação são adicionados a uma composição contendo o parceiro de ligação de CRSP ou o polipéptido de CRSP, e é quantificada a formação de um complexo na ausência do composto de teste.

Os ensaios isentos de células da presente invenção são susceptíveis de utilização de ambas as formas, solúvel e/ou ligada a membranas, das proteínas isoladas (e.g. proteínas CRSP ou suas porções biologicamente activas ou moléculas alvo de CRSP). No caso de ensaios isentos de células em que se utiliza uma forma ligada à membrana de uma proteína isolada (e.g., uma molécula alvo de CRSP ou um receptor) pode ser desejável utilizar um agente solubilizante de modo a que a forma ligada à membrana da proteína isolada se mantenha em solução. Os exemplos destes agentes solubilizantes incluem detergentes não iónicos tais como n-octilglucósido, n-dodecilglucósido, n-dodecilmaltósido, octanoíl-N-metilglucamida, decanoíl-N-metilglucamida, Triton[®] X-100, Triton[®] X-114, Thesit[®], Isotridecilpoli(éter de etilenoglicol)_n, 3-[(3-colamidopropil)dimetilamínio]-1-propanossulfonato (CHAPS), 3-[(3-colamidopropil)dimetilamínio]-2-hidroxi-1-propanossulfonato (CHAPSO) ou N-dodecil=N,N-dimetil-3-amónio-1-propanossulfonato.

Em mais do que uma concretização dos métodos de ensaio da presente invenção anteriores, pode ser desejável imobilizar a CRSP ou a sua molécula alvo para facilitar a separação das formas complexadas das não complexadas de uma ou ambas as proteínas, assim como para acomodar a automatização do ensaio. A ligação de um composto de teste a uma proteína CRSP, ou a interacção de uma proteína CRSP com uma molécula alvo na presença e ausência de um composto candidato, podem ser realizadas em qualquer vaso adequado para conter os reagentes. Os exemplos destes vasos incluem placas de microtítulo, tubos de ensaio e tubos de microcentrífuga. Numa concretização, pode ser proporcionada uma proteína de fusão que adiciona um domínio que permite que uma ou ambas as proteínas se liguem à matriz. Por exemplo, proteínas de fusão glutaciona-S-transferase/CRSP ou proteínas de fusão glutaciona-S-transferase/alvo podem ser adsorvidas em contas de glutaciona-Sepharose (Sigma Chemical, St. Louis, MO) ou placas de microtítulo derivatizadas com glutaciona, que são depois combinadas com o composto de teste ou o composto de teste e a proteína CRSP ou a proteína alvo não adsorvida, e a mistura é incubada sob condições que conduzem à formação do complexo (e.g., em condições fisiológicas relativamente ao sal e ao

pH). Após a incubação, as contas ou os poços da placa de microtítulo são lavados para remover quaisquer componentes não ligados, a matriz é imobilizada no caso de contas, o complexo é determinado directa ou indirectamente, por exemplo, como descrito atrás. Alternativamente, os complexos podem ser dissociados da matriz, e o nível de ligação de CRSP ou actividade determinados utilizando técnicas *standard*.

Outras técnicas para a imobilização de proteínas sobre matrizes podem também ser utilizadas nos ensaios de rastreio da invenção. Por exemplo, quer uma proteína CRSP quer uma molécula alvo de CRSP podem ser imobilizadas utilizando conjugação de biotina e estreptavidina. A proteína CRSP ou as moléculas alvo biotiniladas podem ser preparadas a partir de biotina-NHS (N-hidroxí-succinimida) utilizando técnicas bem conhecidas na especialidade (e.g., kit de biotinilação, Pierce Chemicals, Rockford, IL), e imobilizadas nos poços de placas de 96 poços revestidos com estreptavidina (Pierce Chemical). Alternativamente, anticorpos reactivos com proteína CRSP ou moléculas alvo mas que não interferem com a ligação da proteína CRSP à sua molécula alvo podem ser derivatizadas nos poços da placa, e o alvo ou proteína CRSP não ligados aprisionados nos poços por conjugação com o anticorpo. Os métodos para detectar estes complexos, em adição aos atrás descritos para os complexos imobilizados em GST, incluem imunodeteccção de complexos utilizando anticorpos reactivos com a proteína CRSP ou a molécula alvo, assim como ensaios com enzima ligada que assentam na detecção de uma actividade enzimática associada à proteína CRSP ou à molécula alvo.

Noutra concretização, são identificados moduladores da expressão de CRSP num método em que uma célula é colocada em contacto com um composto candidato e é determinada a expressão de ARNm ou proteína de CRSP na célula. O nível de expressão do ARNm ou proteína de CRSP na presença do composto candidato é comparado com o nível de expressão de ARNm ou proteína de CRSP na ausência do composto candidato. O composto candidato pode então ser identificado como modulador da expressão de CRSP com base nesta comparação. Por exemplo, quando a expressão do ARNm ou proteína de CRSP é maior (maior de maneira estatisticamente significativa) na presença do composto candidato do que na sua ausência, o composto candidato é identificado como estimulador

da expressão de ARNm ou proteína CRSP. Alternativamente, quando a expressão de ARNm ou proteína de CRSP é inferior (inferior de maneira estatisticamente significativa) na presença do composto candidato do que na sua ausência, o composto candidato é identificado como inibidor da expressão de ARNm ou proteína de CRSP. O nível de expressão de ARNm ou proteína de CRSP nas células pode ser determinado por métodos aqui descritos para detecção de ARNm ou proteína CRSP.

Ainda noutro aspecto da invenção, as proteínas CRSP podem ser utilizadas como "proteínas isco" num ensaio de dois híbridos ou ensaio três híbridos (veja-se, e.g., Patente U.S. 5 283 317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *BioTechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; e Brent WO94/10300), para identificar outras proteínas, que se ligam a, ou interagem com, CRSP ("proteínas de ligação a CRSP" ou "CRSP-pb") e modulam a actividade de CRSP. Estas proteínas de ligação a CRSP estão também provavelmente envolvidas na propagação de sinais pelas proteínas CRSP como, por exemplo, elementos a jusante de uma via de sinalização mediada por CRSP. Alternativamente, estas proteínas de ligação a CRSP são provavelmente moléculas de superfície celular associadas a células que não expressam CRSP, em que estas proteínas de ligação a CRSP estão envolvidas na transdução de sinal.

O sistema de dois híbridos baseia-se na natureza modular da maioria dos factores de transcrição, que consistem em domínios separáveis de ligação a ADN e de activação. Resumidamente, o ensaio utiliza duas construções diferentes de ADN. Numa construção, o gene que codifica para uma proteína CRSP é fundido com um gene que codifica o domínio de ligação a ADN de um factor de transcrição conhecido (e.g., GAL-4). Na outra construção, uma sequência de ADN, a partir de uma biblioteca de sequências de ADN, que codifica uma proteína não identificada ("preza" ou "amostra") é fundida com um gene que codifica para o domínio de activação do factor de transcrição conhecido. Se as proteínas "isco" e "preza" forem capazes de interagir, *in vivo*, formando um complexo dependente de CRSP, os domínios de ligação a ADN e de activação do factor de transcrição são colocados em estreita proximidade. Esta

proximidade permite a transcrição de um gene repórter (e.g., LacZ) que está operativamente ligado a um local regulador da transcrição responsivo ao factor de transcrição. A expressão do gene repórter pode ser detectada e as colónias de células contendo o factor de transcrição funcional podem ser isoladas e utilizadas para obter o gene clonado que codifica a proteína que interage com a proteína CRSP.

A presente invenção refere-se ainda a novos agentes identificados pelos ensaios de rastreio atrás descritos e a processos para a produção desses agentes por utilização desses ensaios. Deste modo, numa concretização, a presente invenção inclui um composto ou agente obteníveis por um método compreendendo os passos de qualquer um dos supramencionados ensaios de rastreio (e.g., ensaio à base de células ou ensaios isentos de células). Por exemplo, numa concretização, a invenção inclui um composto ou agente obteníveis por um método compreendendo o contacto de uma célula que expressa uma molécula alvo de CRSP com um composto de teste e a determinação da capacidade do composto de teste para se ligar a, ou modular a actividade da molécula alvo de CRSP. Noutra concretização, a invenção inclui um composto ou agente obteníveis por método compreendendo o contacto de uma célula que expressa uma molécula alvo de CRSP com uma proteína CRSP ou uma sua porção biologicamente activa, para formar uma mistura de ensaio, o contacto da mistura de ensaio com um composto de teste, e a determinação da capacidade do composto de teste para interagir com, ou modular a actividade da molécula alvo de CRSP. Noutra concretização, a invenção inclui um composto ou agente obteníveis por um método compreendendo o contacto de uma proteína CRSP ou uma sua porção biologicamente activa com um composto de teste e a determinação da capacidade do composto de teste para se ligar a, ou modular (e.g., estimular ou inibir) a actividade da proteína CRSP ou da sua porção biologicamente activa. Em ainda outra concretização, a presente invenção inclui um composto ou agente obteníveis por um método compreendendo o contacto de uma proteína CRSP ou uma sua porção biologicamente activa com um composto conhecido que liga a proteína CRSP para formar uma mistura de ensaio, o contacto da mistura de ensaio com um composto de teste, e a determinação da capacidade do composto de teste para interagir com, ou modular a actividade da proteína CRSP.

Deste modo, está dentro do âmbito da presente invenção utilizar ainda um agente identificado como aqui descrito num modelo animal apropriado. Por exemplo, um agente identificado como aqui descrito (e.g., um agente modulador de CRSP, uma molécula de ácido nucleico de CRSP anti-sentido, um anticorpo específico de CRSP ou um parceiro de ligação de CRSP) pode ser utilizado num modelo animal para determinar a eficácia, a toxicidade ou efeitos secundários do tratamento com esse agente. Alternativamente, um agente identificado como aqui descrito pode ser utilizado num modelo animal para determinar o mecanismo de acção de um tal agente.

O presente invento também se refere a utilizações de novos agentes identificados pelos ensaios de rastreio atrás descritos para diagnósticos, prognósticos e tratamentos como aqui descrito. Desse modo, está dentro do âmbito da presente invenção a utilização destes agentes no desenho, formulação, síntese, fabrico e/ou produção de um fármaco ou de uma composição farmacêutica para utilização em diagnóstico, prognóstico ou tratamento, tal como aqui descrito. Por exemplo, numa concretização, a presente invenção inclui um método de síntese ou produção de um fármaco ou de uma composição farmacêutica por referência à estrutura e/ou propriedades de um composto obténível por um dos ensaios de rastreio atrás descritos. Por exemplo, um fármaco ou uma composição farmacêutica podem ser sintetizados com base na estrutura e/ou nas propriedades de um composto obtido por um método em que uma célula que expressa uma molécula alvo de CRSP está em contacto com um composto de teste e é determinada a capacidade do composto de teste para se ligar a, ou modular a actividade da molécula alvo de CRSP. Noutra concretização exemplificativa, a presente invenção inclui um método de síntese ou produção de um fármaco ou de uma composição farmacêutica com base na estrutura e/ou nas propriedades de um composto obténível por um método em que uma proteína CRSP ou uma sua porção biologicamente activa está em contacto com um composto de teste e é determinada a capacidade do composto de teste para se ligar a, ou modular (e.g., estimular ou inibir) a actividade da proteína CRSP ou da sua porção biologicamente activa.

B. Ensaio de Detecção

Porções ou fragmentos das sequências de ADNc aqui identificadas (e as correspondentes sequências de genes completos) podem ser utilizadas de numerosas maneiras como reagentes polinucleotídicos. Por exemplo, estas sequências podem ser utilizadas para: (i) mapear os seus respectivos genes num cromossoma; e, assim, localizar regiões de genes associadas a doença genética; (ii) identificar um indivíduo a partir de uma amostra biológica diminuta (tipificação de tecidos); e (iii) auxiliar na identificação forense de uma amostra biológica. Estas aplicações estão descritas nas subsecções que se seguem.

1. Mapeamento de Cromossomas

Uma vez isolada a sequência (ou uma porção da sequência) de um gene, esta sequência pode ser utilizada para mapear a localização do gene num cromossoma. Este processo é denominado mapeamento de cromossomas. Deste modo, porções ou fragmentos das sequências nucleotídicas de CRSP, aqui descritas, podem ser utilizadas para mapear a localização dos genes de CRSP num cromossoma. O mapeamento das sequências de CRSP em cromossomas é um importante primeiro passo na correlação destas sequências com genes associados a doença.

Resumidamente, os genes de CRSP podem ser mapeados em cromossomas preparando iniciadores de PCR (preferivelmente com 15-25 pb de comprimento) a partir das sequências nucleotídicas de CRSP. Pode utilizar-se análise em computador das sequências de CRSP para prever iniciadores que não se estendam a mais do que um exão no ADN genómico complicando assim o processo de amplificação. Estes iniciadores podem então ser utilizados para o rastreio por PCR de híbridos de células somáticas contendo cromossomas humanos individuais. Apenas os híbridos contendo o gene humano correspondente às sequências de CRSP originará um fragmento amplificado.

Os híbridos de células somáticas são preparados fundindo células somáticas de diferentes mamíferos (e.g., células humanas e de ratinho). Como os híbridos de células humanas e de ratinho crescem e dividem-se, então perdem gradualmente os

cromossomas humanos por ordem aleatória, mas retêm os cromossomas de ratinho. Utilizando meios em que as células de ratinho não podem crescer, porque lhes falta uma enzima particular, mas as células humanas podem, o cromossoma humano que contém o gene que codifica a enzima necessária será retido. Utilizando vários meios, podem ser estabelecidos painéis de linhas de células híbridas. Cada linha celular num painel contém ou um único cromossoma humano ou um pequeno número de cromossomas humanos, e um conjunto completo de cromossomas de ratinho, permitindo o fácil mapeamento de genes individuais nos cromossomas humanos específicos. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Os híbridos de células somáticas contendo apenas fragmentos de cromossomas humanos podem também ser produzidos utilizando cromossomas humanos com translocações e deleções.

O mapeamento por PCR de híbridos de células somáticas é um procedimento rápido para atribuir uma determinada sequência a um determinado cromossoma. Podem ser atribuídas três ou mais sequências por dia utilizando um único ciclizador térmico. Utilizando as sequências nucleotídicas de CRSP para desenhar iniciadores oligonucleotídicos, a sub-localização pode ser conseguida com painéis de fragmentos de cromossomas específicos. Outras estratégias de mapeamento que podem ser similarmente utilizadas para mapear uma sequência 90, 1p, ou 1v no seu cromossoma incluem hibridação *in situ* (descrita em Fan, Y. *et al.* (1990) *PNAS*, 87:6223-27), pré-rastrear com cromossomas marcados separados por fluxo e pré-selecção por hibridação a bibliotecas de ADNc específicos de cromossomas.

A hibridação *in situ* com fluorescência (FISH) de uma sequência de ADN a uma extensão cromossómica em metafase pode ainda ser utilizada para proporcionar uma localização cromossómica precisa num só passo. As extensões cromossómicas podem ser efectuadas utilizando células cuja divisão foi bloqueada em metafase por um produto químico tal como colcemida que corrompe o fuso mitótico. Os cromossomas podem ser tratados sucintamente com tripsina, e depois corados com Giemsa. Desenvolve-se um padrão bandas claras e escuras em cada cromossoma, de modo que os cromossomas podem ser identificados individualmente. A técnica FISH pode ser utilizada com uma sequência de ADN tão curta quanto 500 ou

600 bases. Contudo, clones maiores que 1000 bases têm uma maior probabilidade de se ligarem a uma única localização cromossômica com intensidade de sinal suficiente para uma detecção simples. Preferivelmente 1000 bases, e mais preferivelmente 2000 bases serão suficientes para se obter bons resultados num período de tempo razoável. Para uma revisão desta técnica, veja-se Verma et al., *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

Reagentes para o mapeamento de cromossomas podem ser utilizados individualmente para marcar um único cromossoma ou um único local nesse cromossoma, ou podem ser utilizados painéis de reagentes para marcar múltiplos locais e/ou múltiplos cromossomas. Os reagentes correspondentes a regiões não codificantes dos genes realmente envolvidos são preferidos para fins de mapeamento. As sequências de codificação têm maior probabilidade de serem conservadas nas famílias de genes, aumentando assim a probabilidade de hibridações cruzadas durante o mapeamento cromossômico.

Uma vez mapeada uma sequência numa localização cromossômica precisa, a posição física da sequência no cromossoma pode ser correlacionada com dados do mapa genético. (Estes dados são encontrados, por exemplo, em V. McKusick, *Mendelian Inheritance in Man*, disponível on-line através da Biblioteca Johns Hopkins University Welch Medical Library). A relação entre um gene e uma doença, mapeado na mesma região cromossômica, pode ser então identificada através de análise de ligação (co-herança de genes fisicamente adjacentes), descritos, por exemplo, em Egeland, J. et al. (1987) *Nature*, 325:783-787.

Adicionalmente, podem ser determinadas diferenças nas sequências de ADN entre indivíduos afectados e não afectados com uma doença associada com o gene de CRSP. Se for observada uma mutação em algum ou todos dos indivíduos afectados mas não o for em nenhum indivíduo não afectado, então a mutação é provavelmente o agente causador da doença particular. A comparação de indivíduos afectados e não afectados envolve geralmente observar primeiro alterações estruturais nos cromossomas, tais como deleções ou translocações que são

visíveis a partir de extensões cromossômicas ("*Chromosome Spreads*") ou detectáveis utilizando PCR baseada nessa sequência de ADN. Finalmente, pode ser realizada a sequenciação completa de genes de vários indivíduos para confirmar a presença de uma mutação e para distinguir mutações de polimorfismos.

2. Tipificação de Tecidos

As sequências de CRSP da presente invenção podem também ser utilizadas para identificar indivíduos a partir de amostras biológicas diminutas. O exército dos Estados Unidos, por exemplo, está a considerar a utilização de polimorfismo de comprimento de fragmentos de restrição (RFLP) para a identificação do seu pessoal. Nesta técnica, o ADN genómico de um indivíduo é digerido com uma ou mais enzimas de restrição, e sondado num *Southern blot* para se obterem bandas únicas para identificação. Este método não sofre das correntes limitações das chapas de identificação ("*Dog Tags*") que podem ser perdidas, trocadas ou furtadas, tornando difícil a identificação positiva. As sequências da presente invenção são úteis como marcadores adicionais de ADN para RFLP (descrito na Patente U.S. 5 272 057).

Adicionalmente, as sequências da presente invenção podem ser utilizadas para proporcionar uma técnica alternativa que determina a verdadeira sequência de ADN base-por-base de porções seleccionadas do genoma de um indivíduo. Assim, as sequências nucleotídicas de CRSP aqui descritas podem ser utilizadas para preparar dois iniciadores de PCR a partir das extremidades 5' e 3' das sequências. Estes iniciadores podem então ser utilizados para amplificar o ADN de um indivíduo e subsequentemente sequenciá-lo.

Painéis de sequências de ADN correspondentes de indivíduos, preparados desta maneira, podem proporcionar identificações únicas de indivíduos, pois cada indivíduo terá um conjunto único dessas sequências de ADN devido a diferenças alélicas. As sequências da presente invenção podem ser utilizadas para obter estas sequências de identificação a partir de indivíduos e de tecidos. As sequências nucleotídicas de CRSP da invenção representam unicamente porções do genoma

humano. Ocorre variação alélica em algum grau nas regiões codificantes destas sequências, e num maior grau nas regiões não codificantes. Estima-se que a variação alélica entre indivíduos humanos ocorre com uma frequência de cerca de uma vez por cada 500 bases. Cada uma das sequências aqui descritas pode, em algum grau, ser utilizada como um padrão contra o qual o ADN de um indivíduo pode ser comparado para fins de identificação. Como ocorrem maiores números de polimorfismos nas regiões não codificantes, são necessárias menos sequências para diferenciar indivíduos. As sequências não codificantes de SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7 ou SEQ ID NO:10, podem proporcionar confortavelmente a identificação positiva de indivíduos com um painel de talvez 10 a 1000 iniciadores que originam, cada um, uma sequência não codificante amplificada de 100 bases. Se forem utilizadas sequências codificantes previstas, tais como em SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:9, ou SEQ ID NO:12, um número mais apropriado de iniciadores para a identificação positiva de indivíduos será de 500-2000.

Se for utilizado um painel de reagentes a partir de sequências nucleotídicas de CRSP aqui descritas para gerar uma base de dados de identificação única para um indivíduo, esses mesmos reagentes podem mais tarde ser utilizados para identificar tecidos desse indivíduo. Utilizando a base de dados de identificação única, a identificação positiva do indivíduo, morto ou vivo, pode ser realizada a partir de amostras de tecidos extremamente pequenas.

3. Utilização de Sequências Parciais de CRSP em Biologia Forense

As técnicas de identificação à base de ADN também podem ser utilizadas em biologia forense. A biologia forense é um campo científico que emprega a tipificação genética de vestígios biológicos encontrados num local de crime como um meio para a identificação positiva, por exemplo, de um perpetrador de um crime. Para fazer uma tal identificação, pode ser utilizada tecnologia de PCR para amplificar sequências de ADN tomadas de amostras biológicas muito pequenas tais como tecidos, e.g., cabelo ou pele, ou fluidos corporais, e.g., sangue, saliva ou sêmen encontrados num local de crime. A sequência amplificada pode então ser comparada com

um padrão, permitindo desse modo a identificação da origem da amostra biológica.

As sequências da presente invenção podem ser utilizadas para proporcionar reagentes polinucleotídicos, *e.g.*, iniciadores de PCR, direccionados para *loci* específicos no genoma humano, que podem aumentar a fiabilidade de identificações forenses à base de ADN, por exemplo, proporcionando outro "marcador de identificação" (*i.e.* outra sequência de ADN que é única para um indivíduo particular). Tal como atrás mencionado, a actual informação de sequências base pode ser utilizada para identificação como uma alternativa exacta a padrões formados por fragmentos gerados através de enzimas de restrição. As sequências direccionadas para regiões não codificantes das sequências nucleotídicas da invenção são particularmente apropriadas para esta utilização pois ocorrem maiores números de polimorfismos nas regiões não codificantes, tornando mais fácil a diferenciação de indivíduos utilizando esta técnica. Os exemplos de reagentes polinucleotídicos incluem as sequências nucleotídicas de CRSP ou suas porções *e.g.*, fragmentos derivados de regiões não codificantes de sequências nucleotídicas da invenção, possuindo um comprimento de pelo menos 20 bases, preferivelmente pelo menos 30 bases.

As sequências nucleotídicas de CRSP aqui descritas podem ainda ser utilizadas para proporcionar reagentes polinucleotídicos, *e.g.*, sondas marcadas ou marcáveis que podem ser utilizadas, por exemplo, numa técnica de hibridação *in situ*, para identificar um tecido específico, *e.g.*, tecido cerebral. Isto pode ser muito útil em casos onde um patologista forense é confrontado com um tecido de origem desconhecida. Painéis destas sondas de CRSP podem ser utilizados para identificar tecidos por espécies e/ou por tipo de órgão.

De uma maneira semelhante, estes reagentes, *e.g.*, iniciadores ou sondas de CRSP, podem ser utilizados para rastrear culturas de tecidos quanto a contaminação (*i.e.* rastrear a presença de uma mistura de diferentes tipos de células numa cultura).

C. Medicina Preditiva:

A presente invenção também se refere ao campo da medicina preditiva na qual são utilizados ensaios de diagnóstico, ensaios de prognóstico e ensaios clínicos de monitoração para fins de prognóstico (preditivos) para desse modo tratar um profilacticamente um indivíduo. Deste modo, um aspecto da presente invenção refere-se a ensaios de diagnóstico para a determinação da proteína e/ou expressão de ácido nucleico de CRSP, assim como da actividade de CRSP, no contexto de uma amostra biológica (e.g., sangue, soro, células, tecido) para desse modo determinar se um indivíduo está afectado com uma doença ou desordem, ou está em risco de desenvolver uma desordem, associadas com a expressão ou actividade aberrantes de CRSP, tais como proliferação, diferenciação e/ou sobrevivência celulares aberrantes, resultando por exemplo numa doença neurodegenerativa ou cancro. A invenção também proporciona ensaios de prognóstico (ou preditivos) para determinar se um indivíduo está em risco de desenvolver uma desordem associada à proteína, expressão de ácido nucleico ou actividade de CRSP. Por exemplo, mutações num gene de CRSP podem ser ensaiadas numa amostra biológica. Estes ensaios podem ser utilizados para fins de prognóstico ou preditivos para desse modo tratar profilacticamente um indivíduo antes do início de uma desordem caracterizada por, ou associada a, proteína, expressão ácido nucleico ou actividade de CRSP.

Outra utilização da invenção refere-se à monitoração da influência de agentes (e.g., fármacos, compostos) sobre a expressão ou a actividade de CRSP em ensaios clínicos.

Estes e outros agentes são descritos com mais detalhes nas secções que se seguem.

1. Ensaios de Diagnóstico

Um método exemplificativo para a detecção da presença ou ausência de proteína ou ácido nucleico de CRSP numa amostra biológica envolve a obtenção de uma amostra biológica a partir de um indivíduo de teste e o contacto da amostra biológica com um composto ou um agente capazes de detectar a proteína ou o ácido nucleico (e.g., ARNm, ADN genómico) de CRSP que

codificam proteína CRSP de modo a que a presença de proteína ou ácido nucleico de CRSP seja detectada na amostra biológica. Um agente preferido para a detecção de ARNm ou ADN genómico de CRSP é uma sonda de ácido nucleico marcada capaz de hibridar com ARNm ou ADN genómico de CRSP. A sonda de ácido nucleico pode ser, por exemplo, um ácido nucleico de CRSP de comprimento completo, tal como um ácido nucleico da invenção, ou uma sua porção, tal como um oligonucleótido com pelo menos 15, 30, 50, 100, 250 ou 500 nucleótidos de comprimento e suficiente para hibridar especificamente sob condições rigorosas com ARNm ou ADN genómico de CRSP. Outras sondas adequadas para utilização nos ensaios de diagnóstico da invenção estão aqui descritos.

Um agente preferido para a detecção da proteína CRSP é um anticorpo capaz de se ligar à proteína CRSP, preferivelmente um anticorpo com um marcador detectável. Os anticorpos podem ser policlonais, ou mais preferivelmente, monoclonais. Podem utilizar-se um anticorpo intacto, ou um seu fragmento (e.g., Fab ou F(ab')₂). No termo "marcado", em relação à sonda ou ao anticorpo, pretende-se abranger a marcação directa da sonda ou do anticorpo por acoplamento (i.e., ligação física) de uma substância detectável à sonda ou ao anticorpo, assim como a marcação indirecta da sonda ou do anticorpo por reactividade com outro reagente que é marcado directamente. Os exemplos de marcação indirecta incluem a detecção de um anticorpo primário utilizando um anticorpo secundário marcado com fluorescência e marcando por fim uma sonda de ADN com biotina de modo a que possa ser detectada com estreptavidina marcada com fluorescência. No termo "amostra biológica" pretendem-se incluir tecidos, células e fluidos biológicos isolados de um indivíduo, assim como tecidos, células e fluidos presentes num indivíduo. Ou seja, o método de detecção da invenção pode ser utilizado para detectar ARNm, proteína ou ADN genómico de CRSP numa amostra biológica *in vitro* assim como *in vivo*. Por exemplo, as técnicas *in vitro* para a detecção de ARNm de CRSP incluem hibridações de *Northern* e hibridações *in situ*. As técnicas *in vitro* para a detecção de proteína CRSP incluem ensaios imunossorventes com enzimas ligadas (ELISA), *Western blots*, imunoprecipitações e imunofluorescência. As técnicas *in vitro* para a detecção de ADN genómico de CRSP incluem hibridações de *Southern*. Adicionalmente, as técnicas *in vivo*

para a detecção de proteína CRSP incluem a introdução num indivíduo de um anticorpo anti-CRSP marcado. Por exemplo, o anticorpo pode ser marcado com um marcador radioactivo cuja presença e localização num indivíduo pode ser detectada por técnicas *standard* de imagética.

Numa concretização, a amostra biológica contém moléculas de proteína do indivíduo de teste. Alternativamente, a amostra biológica pode conter moléculas de ARNm do indivíduo de teste ou moléculas de ADN genómico do indivíduo de teste. Uma amostra biológica preferida é uma amostra de soro isolada de um indivíduo por meios convencionais.

Noutra concretização, os métodos envolvem adicionalmente a obtenção de uma amostra biológica de controlo de um indivíduo de controlo, o contacto da amostra de controlo com um composto ou um agente capazes de detectar proteína, ARNm, ou ADN genómico de CRSP, de modo a que a presença de proteína, ARNm ou ADN genómico de CRSP seja detectada na amostra biológica, e a comparação da presença de proteína, ARNm ou ADN genómico de CRSP na amostra de controlo com a presença de proteína, ARNm ou ADN genómico de CRSP na amostra de teste.

A invenção também abrange *kits* para a detecção da presença de CRSP numa amostra biológica. Por exemplo, o *kit* pode compreender um composto ou agente marcados capazes de detectar proteína ou ARNm de CRSP numa amostra biológica; meios para a determinação da quantidade de CRSP na amostra; e meios para a comparação da quantidade de CRSP na amostra com um padrão. O composto ou agente podem ser embalados num receptáculo adequado. O *kit* pode ainda compreender instruções para a utilização do *kit* para detectar proteína ou ácido nucleico de CRSP.

2. Ensaios de Prognóstico

Os métodos de diagnóstico aqui descritos podem adicionalmente ser utilizados para identificar indivíduos possuindo, ou em risco de desenvolver, uma doença ou desordem associadas com a expressão ou actividade aberrantes de CRSP. Por exemplo, os ensaios aqui descritos, tais como os ensaios de diagnóstico anteriores ou os ensaios seguintes, podem ser

utilizados para identificar um indivíduo possuindo, ou em risco de desenvolver, uma desordem associada à proteína, expressão de ácido nucleico ou actividade de CRSP tal como uma desordem proliferativa, uma desordem da diferenciação ou do desenvolvimento, uma desordem hematopoiética, assim como doenças, condições ou desordens caracterizadas por sobrevivência celular anómala, estrutura extracelular anómala ou uma anomalia num mecanismo de defesa. Alternativamente, os ensaios de prognóstico podem ser utilizados para identificar um indivíduo possuindo, ou em risco de desenvolver, uma doença de diferenciação ou proliferativa (e.g., cancro). Assim, a presente invenção proporciona um método para a identificação de uma doença ou desordem associadas com expressão ou actividade de CRSP aberrantes em que uma amostra de teste é obtida a partir de um indivíduo e a proteína ou o ácido nucleico de CRSP (e.g., ARNm, ADN genómico) são detectados, em que a presença de proteína ou ácido nucleico de CRSP constitui um diagnóstico para um indivíduo possuindo, ou em risco de desenvolver, uma doença ou desordem associadas com expressão ou actividade de CRSP aberrantes. Tal como aqui se utiliza, uma "amostra de teste" refere-se a uma amostra biológica obtida de um indivíduo de interesse. Por exemplo, uma amostra de teste pode ser um fluido biológico (e.g., soro), uma amostra celular ou um tecido.

Adicionalmente, os ensaios de prognóstico aqui descritos podem ser utilizados para determinar se um indivíduo pode receber a administração de um agente (e.g., um agonista, antagonista, peptidomimético, proteína, péptido, ácido nucleico, molécula pequena ou outro fármaco candidato) para tratar uma doença ou uma desordem associadas com expressão ou actividade aberrantes de CRSP. Por exemplo, estes métodos podem ser utilizados para determinar se um indivíduo pode ser eficazmente tratado com um agente para uma desordem, tal como uma desordem proliferativa, uma desordem da diferenciação ou do desenvolvimento, uma desordem hematopoiética, assim como desordens caracterizadas por sobrevivência celular anómala, uma estrutura extracelular anómala, ou uma anomalia num mecanismo de defesa. Alternativamente, estes métodos podem ser utilizados para determinar se um indivíduo pode ser eficazmente tratado com um agente para uma doença de diferenciação ou proliferativa (e.g., cancro). Assim, a

presente invenção proporciona métodos para determinar se um indivíduo pode ser eficazmente tratado com um agente para uma desordem associada com expressão ou actividade aberrantes de CRSP em que é obtida uma amostra de teste e a proteína ou a expressão de ácido nucleico ou a actividade de CRSP são detectadas (e.g., em que a abundância da proteína ou da expressão de ácido nucleico ou da actividade de CRSP constitui um diagnóstico para um indivíduo que pode receber a administração do agente para tratar uma desordem associada com expressão ou actividade aberrantes de CRSP.)

Os métodos da invenção podem também ser utilizados para detectar alterações genéticas num gene de CRSP, desse modo determinando se um indivíduo com o gene alterado está em risco de uma desordem caracterizada por desenvolvimento aberrante, diferenciação celular aberrante, proliferação celular aberrante ou uma resposta hematopoiética aberrante. Em concretizações preferidas, os métodos incluem a detecção, numa amostra de células do indivíduo, da presença ou ausência de uma alteração genética caracterizada por pelo menos uma alteração que afecta a integridade de um gene que codifica uma proteína CRSP, ou a expressão errada do gene de CRSP. Por exemplo, estas alterações genéticas podem ser detectadas avaliando a existência de pelo menos uma de 1) uma deleção de um ou mais nucleótidos de um gene de CRSP; 2) uma adição de um ou mais nucleótidos a um gene de CRSP; 3) uma substituição de um ou mais nucleótidos de um gene de CRSP, 4) um arranjo cromossómico de um gene de CRSP; 5) uma alteração no nível de um transcrito de ARN mensageiro de um gene de CRSP, 6) modificação aberrante de um gene de CRSP, tal como do padrão de metilação do ADN genómico, 7) a presença de padrão de *splicing* que não o de tipo selvagem de um transcrito de ARN mensageiro de um gene de CRSP, 8) um nível que não o de tipo selvagem de uma proteína CRSP, 9) perda alélica de um gene de CRSP, e 10) modificação pós-tradução inapropriada de uma proteína CRSP. Como aqui descrito, existe um grande número de técnicas de ensaio conhecidas na especialidade que podem ser utilizadas para detectar alterações num gene de CRSP. Uma amostra biológica preferida é uma amostra de tecido ou soro isolada de um indivíduo por meios convencionais.

Em certas concretizações, a detecção da alteração envolve a utilização de uma sonda/iniciador numa reacção em cadeia com polimerase (PCR) (veja-se, e.g., Patentes U.S. 4 683 195 e 4 683 202), tal como PCR de ancoragem ou PCR-RACE, ou, alternativamente, numa reacção de ligação em cadeia (LCR) (veja-se, e.g., Landegran et al. (1988) *Science* 241:1077-1080; e Nakazawa et al. (1994) *PNAS* 91:360-364), podendo esta última ser particularmente útil para a detecção de mutações pontuais no gene de CRSP (veja-se Abravaya et al. (1995) *Nucleic Acids Res.*, 23:675-682). Este método pode incluir os passos de recolha de uma amostra de células de um paciente, isolamento de ácido nucleico (e.g., genómico, ARNm ou ambos) das células da amostra, contacto da amostra de ácido nucleico com um ou mais iniciadores que hibridam especificamente com um gene de CRSP sob condições tais que ocorra a hibridação e a amplificação do gene de CRSP (se presente), e detecção da presença ou ausência de um produto de amplificação, ou detecção do tamanho do produto de amplificação e comparação do comprimento com uma amostra de controlo. Prevê-se que a PCR e/ou a LCR possam ser desejáveis para utilização como passo de amplificação preliminar em conjunto com quaisquer das técnicas utilizadas para a detecção de mutações aqui descritas.

Os métodos de amplificação alternativos incluem: replicação auto-sustentada de sequências (Guatelli, J.C. et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:1874-1878), sistema de amplificação transcricional (Kwoh, D.Y. et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. et al., 1988, *Bio/Technology* 6:1197), ou qualquer outro método de amplificação de ácido nucleico, seguidos da detecção das moléculas amplificadas utilizando técnicas bem conhecidas dos peritos na especialidade. Estes esquemas de detecção são especialmente úteis para a detecção de moléculas de ácido nucleico se estas moléculas estiverem presentes em números muito pequenos.

Numa concretização alternativa, as mutações num gene de CRSP de uma amostra de células podem ser identificadas através de alterações em padrões de clivagem por enzimas de restrição. Por exemplo, o ADN da amostra e de controlo é isolado, amplificado (opcionalmente), digerido com uma ou mais endonucleases de restrição, e as dimensões do comprimento dos

fragmentos são determinadas por electroforese em gel e comparadas. As diferenças nas dimensões do comprimento dos fragmentos entre o ADN da amostra e de controlo indicam mutações no ADN da amostra. Adicionalmente, a utilização de ribozimas específicas da sequência (veja-se, por exemplo, Patente U.S. 5 498 531) pode ser utilizada para classificar a presença de mutações específicas por desenvolvimento ou perda de um local de clivagem para ribozima.

Noutras concretizações, as mutações genéticas na CRSP podem ser identificadas por hibridação de ácidos nucleicos da amostra e de controlo, e.g., ADN ou ARN, com conjuntos de alta densidade contendo centenas ou milhares de sondas oligonucleotídicas (Cronin, M.T. et al. (1996) *Human Mutation* 7: 244-255; Kozal, M.J. et al. (1996) *Nature Medicine* 2: 753-759). Por exemplo, as mutações genéticas na CRSP podem ser identificadas em dois conjuntos dimensionais contendo sondas de ADN gerados por luz como descrito em Cronin, M.T. et al. *supra*. Resumidamente, um primeiro conjunto de hibridação de sondas pode ser utilizado para varrer longos trechos de ADN numa amostra e num controlo para identificar alterações de bases entre as sequências fazendo conjuntos lineares de sondas de sobreposição sequenciais. Este passo permite a identificação de mutações pontuais. Este passo é seguido por um segundo conjunto de hibridação que permite a caracterização de mutações específicas utilizando conjuntos mais pequenos e especializados de sondas complementares a todas as variantes ou mutações detectadas. Cada conjunto de mutação é composto por conjuntos de sondas paralelas, um complementar com o gene de tipo selvagem e o outro complementar com o gene mutante.

Em ainda outra concretização, qualquer uma de uma variedade de reacções de sequenciação conhecidas na especialidade pode ser utilizada para sequenciar directamente o gene de CRSP e detectar mutações por comparação da sequência da CRSP da amostra com a correspondente sequência de tipo selvagem (controlo). Os exemplos de reacções de sequenciação incluem as que são baseadas em técnicas desenvolvidas por Maxim e Gilbert ((1977) *PNAS* 74:560) ou Sanger ((1977) *PNAS* 74:5463). Está igualmente contemplado que possa ser utilizado qualquer um de uma variedade de procedimentos de sequenciação automática aquando da realização dos ensaios de diagnóstico

((1995) *BioTechniques* 19:448), incluindo sequenciação por espectrometria de massa (veja-se, e.g., Publicação Internacional PCT WO 94/16101; Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; e Griffin et al. (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Outros métodos para a detecção de mutações no gene da CRSP incluem métodos em que a protecção contra agentes de clivagem é utilizada para detectar bases mal emparelhadas em ARN/ARN ou heterodúpliques ARN/ADN (Myers et al. (1985) *Science* 230:1242). Em geral, a técnica de "clivagem de erros de emparelhamento" inicia-se proporcionando heterodúpliques formadas por hibridação de ARN ou ADN (marcado) contendo a sequência de CRSP de tipo selvagem com ARN ou ADN potencialmente mutante obtido de uma amostra de tecido. Os dúpliques de cadeia dupla são tratados com um agente que cliva regiões de cadeia simples do dúplex tais como as que irão existir devido a erros de emparelhamento de bases entre as cadeias de controlo e da amostra. Por exemplo, os dúpliques de ARN/ADN podem ser tratados com ARNase e os híbridos ADN/ADN tratados com nuclease S1 para digerir enzimaticamente as regiões com erros de emparelhamento. Noutras concretizações, ambos os dúpliques de ADN/ADN ou de ARN/ADN podem ser tratados com hidroxilamina ou tetróxido de ósmio e com piperidina para digerir regiões com erros de emparelhamento. Após a digestão das regiões com erros de emparelhamento, o material resultante é então separado por tamanhos sobre géis de poliacrilamida desnaturantes para determinar o local da mutação. Veja-se, por exemplo, Cotton et al. (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba et al. (1992) *Methods Enzymol.* 217:286-295. Numa concretização preferida, o ADN ou ARN de controlo podem ser marcados para detecção.

Ainda noutra concretização, a reacção de clivagem de erros de emparelhamento emprega uma ou mais proteínas que reconhecem pares de bases mal emparelhadas em ADN de cadeia dupla (as denominadas enzimas de "reparação de erros de emparelhamento no ADN") em sistemas definidos para a detecção e mapeamento de mutações pontuais em ADNc de CRSP obtidos de amostras de células. Por exemplo, a enzima mutY de *E. coli* cliva A em erros de emparelhamento G/A e a timidina-ADN-glicosilase de células HeLa cliva T em erros de emparelhamento

G/T (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662). De acordo com uma concretização exemplificativa, uma sonda baseada numa sequência de CRSP, e.g., uma sequência de CRSP de tipo selvagem, é hibridada com um ADNc ou outro produto de ADN de uma célula ou células de teste. O dúplex é tratado com uma enzima de reparação de erros de emparelhamento no ADN, e os produtos de clivagem, se existirem, podem ser detectados a partir de protocolos de electroforese ou semelhantes. Veja-se, por exemplo, Patente U.S. 5 459 039.

Noutras concretizações, utilizar-se-ão alterações na mobilidade electroforética para identificar mutações em genes de CRSP. Por exemplo, pode utilizar-se o polimorfismo de conformação de cadeias simples (SSCP) para detectar diferenças na mobilidade electroforética entre ácidos nucleicos mutantes e de tipo selvagem (Orita et al. (1989) *Proc Natl. Acad. Sci USA*: 86:2766, veja-se também Cotton (1993) *Mutat Res* 285:125-144; e Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Fragmentos de ADN em cadeia simples de ácidos nucleicos de CRSP da amostra e de controlo serão desnaturados e deixados renaturar. A estrutura secundária de ácidos nucleicos de cadeia simples varia de acordo com a sequência, a alteração resultante na mobilidade electroforética permite a detecção mesmo de uma única alteração de base. Os fragmentos de ADN podem ser marcados ou detectados com sondas marcadas. A sensibilidade do ensaio pode ser aumentada utilizando ARN (em vez de ADN), em que a estrutura secundária é mais sensível a uma alteração na sequência. Numa concretização preferida, o método em questão utiliza a análise de heterodúplexes para separar moléculas de heterodúplexes de cadeia dupla com base nas alterações na mobilidade electroforética (Keen et al. (1991) *Trends Genet* 7:5).

Em ainda outra concretização o movimento de fragmentos mutantes ou de tipo selvagem em géis de poliacrilamida contendo um gradiente de desnaturante, é ensaiado utilizando electroforese em gel de gradiente desnaturante (DGGE) (Myers et al. (1985) *Nature* 313:495). Quando se utiliza DGGE como método de análise, o ADN será modificado para assegurar que não desnatura completamente, por exemplo por adição de um grampo GC de aproximadamente 40 pb de ADN rico em GC de elevado ponto de fusão, por PCR. Noutra concretização,

utiliza-se um gradiente de temperaturas em vez de um gradiente desnaturante para identificar diferenças na mobilidade de ADN de controlo e da amostra (Rosenbaum e Reissner (1987) *Biophys Chem* 265:12753).

Os exemplos de outras técnicas para detecção de mutações pontuais incluem, mas não se lhes limitam, hibridação selectiva de oligonucleótidos, amplificação selectiva ou extensão selectiva de iniciadores. Por exemplo, podem preparar-se iniciadores oligonucleotídicos em que a mutação conhecida é colocada centralmente e depois hibridam-se com ADN alvo sob condições que permitem a hibridação apenas se for encontrada uma correspondência perfeita (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Estes oligonucleótidos específicos de alelos são hibridados com ADN alvo amplificado por PCR ou um número de diferentes mutações quando os oligonucleótidos são ligados à membrana de hibridação e hibridados com ADN alvo marcado.

Alternativamente, a tecnologia de amplificação específica de alelos, que depende da amplificação selectiva por PCR, pode ser utilizada em conjunto com a presente invenção. Os oligonucleótidos utilizados como iniciadores para a amplificação específica podem possuir a mutação de interesse no centro da molécula (de modo a que a amplificação dependa da hibridação diferencial) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) ou no extremo da extremidade 3' de um iniciador onde, sob condições apropriadas, os erros de emparelhamento possam evitar, ou reduzir, o alongamento com polimerase (Prossner (1993) *Tibtech* 11:238). Em adição, pode ser desejável introduzir um novo local de restrição na região da mutação para criar uma detecção à base de clivagem (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). Antecipa-se que em certas concretizações a amplificação possa também ser realizada utilizando ligase Taq para a amplificação (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). Nestes casos, a ligação ocorrerá apenas se houver uma correspondência perfeita na extremidade 3' da sequência a 5' que torna possível detectar a presença de uma mutação conhecida num local específico procurando a presença ou ausência de amplificação.

Os métodos aqui descritos podem ser realizados, por exemplo, utilizando kits de diagnóstico pré-embalados compreendendo pelo menos uma sonda de ácido nucleico ou um anticorpo reagente aqui descritos, que podem ser convenientemente utilizados, e.g., em estabelecimentos clínicos para diagnosticar pacientes que apresentam sintomas ou história familiar de uma doença ou enfermidade que envolva um gene de CRSP.

Adicionalmente, qualquer tipo de célula ou tecido em que a CRSP é expressa podem ser utilizados nos ensaios de prognóstico aqui descritos.

3. Monitoração de Efeitos Durante Ensaio Clínicos

A monitoração da influência de agentes (e.g., fármacos, compostos) sobre a expressão ou actividade de CRSP (e.g., modulação da transdução de sinais celulares, regulação da transcrição de genes numa célula envolvida no desenvolvimento ou diferenciação, regulação da proliferação celular) pode ser aplicada não apenas em rastreio básico de fármacos, mas também em ensaios clínicos. Por exemplo, a eficácia de um agente determinada por um ensaio de pesquisa como aqui descrito para aumentar a expressão de um gene de CRSP, os níveis de proteína ou supra-regular a actividade de CRSP, pode ser monitorada em ensaios clínicos de indivíduos que apresentam uma expressão diminuída de genes de CRSP, de níveis de proteína, ou actividade de CRSP supra-regulada. Alternativamente, a eficácia de um agente determinada por um ensaio de pesquisa para diminuir a expressão de um gene de CRSP, os níveis de proteína, ou infra-regular a actividade de CRSP, pode ser monitorada em ensaios clínicos de indivíduos que apresentam uma expressão aumentada de genes de CRSP, de níveis de proteína, ou actividade de CRSP supra-regulada. Nestes ensaios clínicos, a expressão ou actividade de CRSP e, preferivelmente, outros genes que foram implicados, por exemplo, numa desordem proliferativa, podem ser utilizadas como "read out" ou marcadores do fenótipo de uma célula particular.

Por exemplo, e não como limitação, genes, incluindo os de CRSP, que são modulados em células por tratamento com um

agente (e.g., composto, fármaco ou molécula pequena) que modula a actividade de CRSP (e.g., identificados num ensaio de pesquisa como aqui descrito) podem ser identificados. Assim, para estudar o efeito de agentes sobre desordens proliferativas, desordens do desenvolvimento ou de diferenciação, desordens hematopoiéticas assim como desordens caracterizadas por diferenciação e/ou sobrevivência celulares anómalas, uma estrutura extracelular anómala ou uma anomalia num mecanismo de defesa, por exemplo, num ensaio clínico, podem ser isoladas células e o ARN preparado e analisado quanto aos níveis de expressão de CRSP e outros genes implicados na desordem proliferativa, desordem do desenvolvimento ou de diferenciação, desordem hematopoiética assim como desordens caracterizadas por diferenciação e/ou sobrevivência celulares anómalas, uma estrutura extracelular anómala, ou uma anomalia num mecanismo de defesa, respectivamente. Os níveis de expressão dos genes (i.e., o padrão de expressão génica) podem ser quantificados por análise *Northern blot* ou RT-PCR, como aqui descrito, ou alternativamente medindo a quantidade de proteína produzida, por um dos métodos aqui descritos, ou medindo os níveis de actividade de CRSP ou outros genes. Desta maneira, o padrão de expressão génica pode servir como marcador, indicador da resposta fisiológica das células ao agente. Deste modo, este estado de resposta pode ser determinado antes, e em vários pontos durante o tratamento do indivíduo com o agente.

A presente invenção pode ser utilizada como base para um método para monitoração da eficácia de tratamento de um indivíduo com um agente (e.g., agonista, antagonista, peptidomimético, proteína, péptido, ácido nucleico, molécula pequena, ou outro fármaco candidato identificado pelos ensaios de rastreio aqui descritos) compreendendo os passos de (i) obtenção de uma amostra pré-administração de um indivíduo antes da administração do agente, (ii) detecção do nível de expressão de uma proteína, um ARNm ou um ADN genómico de CRSP na amostra pré-administração; (iii) obtenção de uma ou mais amostras pós-administração do indivíduo; (iv) detecção do nível de expressão ou da actividade da proteína, ARNm ou ADN genómico de CRSP nas amostras pós-administração; (v) comparação do nível de expressão ou da actividade da proteína, ARNm ou ADN genómico de CRSP na amostra pré-administração com

a proteína, ARNm ou ADN genómico de CRSP na amostra ou amostras pós-administração; e (vi) alteração da administração do agente ao indivíduo em concordância. Por exemplo, pode ser desejável uma administração aumentada do agente para aumentar a expressão ou a actividade de CRSP para níveis superiores aos detectados, *i.e.*, para aumentar a eficácia do agente. Alternativamente, pode ser desejável uma administração diminuída do agente para diminuir a expressão ou a actividade de CRSP para níveis inferiores aos detectados, *i.e.* para diminuir a eficácia do agente. De acordo com esta concretização, a expressão ou a actividade de CRSP podem ser utilizadas como um indicador da eficácia de um agente, mesmo na ausência de uma resposta fenotípica observável.

C. Métodos de Tratamento:

A presente invenção pode ser utilizada como base para métodos de tratamento, tanto profilácticos como terapêuticos, de um indivíduo em risco de (ou susceptível a) uma desordem ou possuindo uma desordem associada a expressão ou actividade aberrantes de CRSP. Em relação a métodos de tratamento, tanto profilácticos como terapêuticos, estes tratamentos podem ser especificamente desenhados ou modificados, com base em conhecimentos obtidos no campo da farmacogenómica. "Farmacogenómica", como aqui se utiliza, refere-se à aplicação de tecnologias genómicas tais como sequenciação de genes, genética estatística e análise da expressão de genes para fármacos em desenvolvimento clínico e no mercado. Mais especificamente, o termo refere-se ao estudo de como os genes de um paciente determinam a sua resposta a um fármaco (*e.g.*, "fenótipo de resposta a um fármaco" ou "genótipo de resposta a um fármaco" de um paciente.) Assim, a invenção proporciona a base de métodos para o desenho de um tratamento profiláctico ou terapêutico de um indivíduo com as moléculas de CRSP da presente invenção ou com moduladores de CRSP de acordo com o genótipo de resposta ao fármaco do indivíduo. A farmacogenómica permite que um clínico ou um médico direcione os tratamentos profilácticos ou terapêuticos para os pacientes que irão mais beneficiar do tratamento e evite o tratamento de pacientes que sofrerão efeitos secundários tóxicos relacionados com o fármaco.

1. Métodos Profiláticos

Num aspecto, a invenção proporciona o fabrico de um medicamento para a prevenção num indivíduo, de uma doença ou condição associadas a uma expressão ou actividade aberrantes de CRSP. O medicamento inclui um agente que modula a expressão de CRSP ou pelo menos uma actividade de CRSP. Os indivíduos em risco de uma doença que é causada ou para que contribuem a expressão ou actividade aberrantes de CRSP, podem ser identificados, por exemplo, através de qualquer um ou uma combinação de ensaios de diagnóstico ou prognóstico aqui descritos. A administração de um agente profilático pode ocorrer antes da manifestação de sintomas característicos da aberrância da CRSP, de modo que uma doença ou uma desordem são evitadas ou, alternativamente, retardadas na sua progressão. Dependendo do tipo de aberrância da CRSP, por exemplo, pode ser utilizado um agente agonista de CRSP ou antagonista de CRSP para o tratamento do indivíduo. O agente apropriado pode ser determinado com base em ensaios de rastreio aqui descritos. Os métodos profiláticos são adicionalmente discutidos nas subsecções que se seguem.

2. Métodos Terapêuticos

Outro aspecto da invenção proporciona o fabrico de um medicamento para a modulação da expressão ou actividade de CRSP para fins terapêuticos. O método modulador da invenção envolve o contacto de uma célula com um agente que modula uma ou mais das actividades da proteína CRSP, actividade associada à célula. Um agente que modula a actividade da proteína CRSP pode ser um agente como aqui descrito, tal como um ácido nucleico ou uma proteína, uma molécula de ocorrência natural alvo de uma proteína CRSP, um péptido, um peptidomimético de CRSP ou outra molécula pequena. Numa concretização, o agente estimula uma ou mais actividades da proteína CRSP. Os exemplos destes agentes estimuladores incluem proteína CRSP activa e uma molécula de ácido nucleico que codifica CRSP que foi introduzida na célula. Noutra concretização, o agente inibe uma ou mais actividades da proteína CRSP. Os exemplos destes agentes inibidores incluem moléculas de ácido nucleico de CRSP anti-sentido e anticorpos anti-CRSP. Estes métodos moduladores podem ser realizados *in vitro* (e.g., por cultura da célula com

o agente) ou, alternativamente, *in vivo*, (e.g., por administração do agente a um indivíduo). Como tal, a presente invenção proporciona o fabrico de medicamentos para o tratamento de um indivíduo afectado com uma doença ou desordem caracterizadas por uma expressão ou actividade aberrantes de uma proteína ou de uma molécula de ácido nucleico de CRSP. Numa concretização, o medicamento inclui um agente (e.g., um agente identificado por um ensaio de rastreio aqui descrito), ou uma combinação de agentes que modulam (e.g., supra-regulam ou infra-regulam) a expressão ou a actividade de CRSP. Uma proteína ou uma molécula de ácido nucleico de CRSP podem ser administradas como terapia para compensar uma expressão ou actividade reduzidas ou aberrantes de CRSP.

A estimulação da actividade de CRSP é desejável em situações em que a CRSP está anormalmente infra-regulada e/ou em que uma actividade de CRSP aumentada tem probabilidade de ter um efeito benéfico. Do mesmo modo, a inibição da actividade de CRSP é desejável em situações em que a CRSP está anormalmente supra-regulada e/ou em que uma actividade de CRSP diminuída tem probabilidade de ter um efeito benéfico. Um exemplo de uma tal situação é quando um indivíduo tem uma desordem caracterizada por desenvolvimento ou diferenciação celular aberrantes. Outro exemplo de uma tal situação é quando o indivíduo tem uma doença proliferativa (e.g., cancro) ou uma desordem caracterizada por uma resposta hematopoiética aberrante. Ainda outro exemplo de uma tal situação é quando é desejável conseguir regeneração de tecidos num indivíduo (e.g., quando um indivíduo sofreu danos cerebrais ou na medula espinal e é desejável regenerar tecido neuronal de uma maneira regulada.)

Deste modo, numa concretização, a doença é uma doença caracterizada por uma proliferação, diferenciação e/ou sobrevivência celulares anómalas. Por exemplo, a doença pode ser uma doença hiper- ou hipoproliferativa. A invenção também proporciona o fabrico de um medicamento para o tratamento de doenças caracterizadas por uma proliferação, diferenciação e/ou sobrevivência celulares anómalas num indivíduo, que não são caracterizadas por uma actividade de CRSP anómala (e.g., actividade de CRSP-1). De facto, como a CRSP é provavelmente capaz de modular o estado proliferativo de uma célula (*i.e.*, o

estado de proliferação, diferenciação e ou sobrevivência de uma célula), a CRSP pode regular a doença em que o estado proliferativo anómalo de uma célula resulta de um defeito outro que não uma actividade de CRSP anómala.

As doenças hiperproliferativas que podem ser tratadas com terapêuticos de CRSP (e.g., CRSP-1) incluem doenças neoplásicas e hiperplásicas, tais como várias formas de cancro e leucemias, e desordens fibroproliferativas. Outras doenças hiperproliferativas que podem ser tratadas ou prevenidas com os terapêuticos de CRSP em questão (e.g. terapêuticos de CRSP-1) incluem condições malignas, condições pré-malignas e condições benignas. A condição a tratar ou prevenir pode ser um tumor sólido, tal como um tumor que surge num tecido epitelial. Assim, o tratamento de um tal cancro pode compreender a administração ao indivíduo de um terapêutico de CRSP que diminui a interacção da CRSP com um receptor de CRSP. Outros cancros que podem ser tratados ou prevenidos com uma proteína CRSP incluem sarcomas e carcinomas, e.g., cancro do pulmão, cancro do cólon, da próstata, da mama, dos ovários, do esófago, cancro do pulmão, melanoma, seminoma e adenocarcinoma escamoso. Tumores sólidos adicionais no âmbito da invenção incluem aqueles que podem ser encontrados num livro de texto de medicina.

A condição a tratar ou prevenir pode também ser um tumor solúvel, tal como leucemia, crónica ou aguda, incluindo leucemia mielogenosa crónica ou aguda, leucemia linfocítica crónica ou aguda, leucemia promielocítica, leucemia monocítica, leucemia mielomonocítica e eritroleucemia. Ainda outras desordens proliferativas que podem ser tratadas com um terapêutico de CRSP da invenção incluem doença de cadeia pesada, mieloma múltiplo, linfoma, e.g., linfoma de Hodgkin e linfoma não de Hodgkin, e macroglobulemia de Waldenstroem.

As doenças ou condições caracterizadas por um tumor sólido ou solúvel podem ser tratadas por administração de um terapêutico de CRSP, local ou sistemicamente, de modo a que a proliferação celular aberrante seja inibida ou diminuída. Os métodos para administração dos compostos da invenção são adicionalmente descritos adiante.

A invenção também proporciona o fabrico de um medicamento para a prevenção da formação e/ou do desenvolvimento de tumores. Por exemplo, o desenvolvimento de um tumor pode ser precedido pela presença de uma lesão específica, tal como uma lesão pré-neoplásica, *e.g.*, hiperplasia, metaplasia e displasia, que podem ser detectadas, *e.g.*, por métodos citológicos. Estas lesões podem ser encontradas, *e.g.*, em tecido epitelial. Assim, a invenção proporciona o fabrico de um medicamento para inibição da progressão de uma tal lesão para uma lesão neoplásica. O indivíduo possuindo uma lesão pré-neoplásica pode receber a administração de uma quantidade de um terapêutico de CRSP-1 suficiente para inibir a progressão da lesão pré-neoplásica para uma lesão neoplásica.

A invenção também proporciona o fabrico de um medicamento para o tratamento ou prevenção de doenças ou condições em que se deseja a proliferação de células. Por exemplo, podem ser utilizados terapêuticos de CRSP para estimular a reparação de tecidos ou a cura de feridas, tal como após cirurgia ou para estimular a cura de tecidos de queimaduras. Outras doenças em que a proliferação de células é desejada são as doenças hipoproliferativas, *i.e.*, doenças caracterizadas por uma proliferação anormalmente baixa de certas células.

Ainda noutra concretização, a invenção proporciona o fabrico de um medicamento para o tratamento ou prevenção de doenças ou condições caracterizadas por diferenciação celular aberrante. Deste modo, métodos para estimulação da diferenciação celular em condições caracterizadas por uma inibição da diferenciação celular normal que pode ou não ser acompanhada por proliferação excessiva. Alternativamente, os terapêuticos de CRSP podem ser utilizados para inibir a diferenciação de células específicas.

Num método preferido, a célula com proliferação e/ou diferenciação aberrantes é uma célula presente no sistema nervoso. Um papel da CRSP no sistema nervoso é sugerido, pelo menos em parte, do facto de a CRSP-1 humana ser expressa em cérebro humano fetal. Deste modo, a invenção proporciona o fabrico de um medicamento para o tratamento de doenças ou condições associadas ao sistema nervoso central ou periférico. Por exemplo, a invenção proporciona o fabrico de um

medicamento para o tratamento de lesões do sistema nervoso associadas a uma proliferação, diferenciação ou sobrevivência aberrantes de quaisquer das seguintes células: neurónios, células de Schwann, células gliais e outros tipos de células neurais. As desordens do sistema nervoso incluem, mas não se lhes limitam: lesões na medula espinal, lesões cerebrais, lesões associadas a cirurgia, lesões isquémicas, lesões malignas, lesões infecciosas, lesões degenerativas (doença de Parkinson, doença de Alzheimer, coreia de Huntington, esclerose lateral amiotrófica), doenças desmielinizantes (esclerose múltipla, mielopatia associada a imunodeficiência humana, mielopatia transversa, leucoencefalopatia multifocal progressiva, mielinolise pontina), lesões de neurónios motores, atrofia muscular espinal progressiva, palsy bulbar progressiva, esclerose lateral primária, atrofia muscular infantil e juvenil, paralisia bulbar progressiva da infância (síndrome de Fazio-Londe), poliomielite e neuropatia motossensorial hereditária (doença de Charcot-Marie-Tooth).

Noutra concretização, a invenção proporciona o fabrico de um medicamento para aumentar a sobrevivência e/ou estimular a proliferação e/ou a diferenciação de células e tecidos *in vitro*. Numa concretização preferida, os terapêuticos de CRSP são utilizados para promover a regeneração e/ou reparação de tecidos (e.g., para tratar lesões nervosas). Por exemplo, tecidos de um indivíduo podem ser obtidos e crescidos *in vitro* na presença de um terapêutico de CRSP, de modo a que as células do tecido sejam estimuladas a proliferar e/ou diferenciar-se. O tecido pode então ser readministrado ao indivíduo.

Entre as abordagens que podem ser utilizadas para melhorar sintomas de doenças que envolvem uma actividade de CRSP aberrante e/ou uma proliferação, diferenciação e/ou sobrevivência celulares anómalas, estão, por exemplo, as moléculas anti-sentido, de ribozima e de hélice tripla aqui descritas. Os exemplos de compostos adequados incluem os antagonistas, agonistas ou homólogos descritos atrás em detalhe.

Ainda outro terapêutico de CRSP consiste num primeiro péptido compreendendo um péptido de CRSP capaz de se ligar a

um receptor de CRSP, e um segundo péptido que é citotóxico. Estes terapêuticos podem ser utilizados para especificamente atingir e lisar células que expressam ou super-expressam um receptor para a CRSP.

3. Farmacogenómica

As moléculas de CRSP da presente invenção, assim como agentes ou moduladores que têm um efeito estimulador ou inibidor sobre a actividade de CRSP (e.g., expressão de genes de CRSP) como identificados por um ensaio de rastreio aqui descrito, podem ser administrados a indivíduos para tratar (profiláctica ou terapêuticamente) desordens (e.g., desordens proliferativas ou de desenvolvimento) associadas com uma actividade de CRSP aberrante. Em conjunto com este tratamento, pode ser considerada a farmacogenómica (i.e., o estudo da relação entre o genótipo de um indivíduo e a resposta desse indivíduo a um composto ou fármaco estranhos). As diferenças no metabolismo de terapêuticos pode levar a uma toxicidade grave ou a uma falha terapêutica por alteração da relação entre a dose e a concentração sanguínea do fármaco farmacologicamente activo. Assim, um médico ou um clínico podem considerar a aplicação do conhecimento obtido em estudos de farmacogenómica relevantes para determinar se devem administrar uma molécula de CRSP ou um modulador de CRSP, assim como no desenho do regime de dosagem e/ou terapêutico de tratamento com uma molécula de CRSP ou um modulador de CRSP.

A farmacogenómica lida com variações hereditárias clinicamente significativas na resposta a fármacos devido a uma disposição alterada do fármaco e acção anómala em pessoas afectadas. Veja-se e.g., Eichelbaum, M., *Clin Exp Pharmacol Physiol*, 1996, 23(10-11): 983-985 e Linder, M.W., *Clin Chem*, 1997, 43(2):254-266. Em geral, podem-se diferenciar dois tipos de condições farmacogenéticas. As condições genéticas transmitidas na forma de um único factor que altera o modo como os fármacos actuam no corpo (acção alterada do fármaco) ou condições genéticas transmitidas na forma de factores únicos que alteram o modo como o corpo actua sobre os fármacos (metabolismo alterado do fármaco). Estas condições farmacogenéticas podem ocorrer quer na forma de defeitos genéticos raros quer na forma de polimorfismos de ocorrência

natural. Por exemplo, uma deficiência em glucose-6-fosfato-desidrogenase (G6PD) é uma enzimapatia hereditária comum em que a principal complicação clínica é a hemólise após a ingestão de fármacos oxidantes (antimaláricos, sulfonamidas, analgésicos, nitrofuranos) e o consumo de favas.

Uma abordagem farmacogenómica para identificar genes que prevêm a resposta ao fármaco, conhecida como "uma associação ampla ao genoma", assenta principalmente num mapa de alta resolução do genoma humano consistindo nos marcadores relacionados com o gene já conhecidos (e.g., um mapa de marcadores génicos "bialélicos" que consiste em 60 000-100 000 locais polimórficos ou variáveis no genoma humano, cada um com duas variantes.) Este mapa genético de alta resolução pode ser comparado com um mapa do genoma de cada um de um número estatisticamente significativo de pacientes que fazem parte de um ensaio de Fase II/III do fármaco para identificar marcadores associados com uma resposta ou efeito secundário particulares observados do fármaco. Alternativamente, este mapa de alta resolução pode ser gerado a partir de uma combinação de alguns dez milhões de polimorfismos nucleotídicos simples (SNP) conhecidos no genoma humano. Tal como aqui se utiliza, um "SNP" é uma alteração comum que ocorre numa única base nucleotídica num trecho de ADN. Por exemplo, um SNP pode ocorrer uma vez em cada 1000 bases de ADN. Um SNP pode estar envolvido num processo de doença, contudo, a vasta maioria pode não estar associada a doenças. Dado um mapa genético baseado na ocorrência destes SNP, os indivíduos podem ser agrupados em categorias genéticas dependendo de um padrão particular de SNP no seu genoma individual. Desta maneira, os regimes de tratamento podem ser desenhados para grupos de indivíduos geneticamente similares, tendo em conta traços que podem ser comuns entre esses indivíduos geneticamente similares.

Alternativamente, um método denominado a "abordagem do gene candidato", pode ser utilizado para identificar genes que prevêm a resposta a fármacos. De acordo com este método, se um gene que codifica um alvo do fármaco é conhecido (e.g., uma proteína CRSP ou um receptor de CRSP da presente invenção), todas as variantes comuns desse gene podem ser muito facilmente identificadas na população e pode-se determinar se

possuir uma versão do gene *versus* outra está associado com uma resposta particular a um fármaco.

Como concretização ilustrativa, a actividade de enzimas metabolizantes do fármaco é um determinante principal tanto da intensidade como da duração de acção do fármaco. A descoberta de polimorfismos genéticos de enzimas metabolizantes de fármacos (e.g., N-acetiltransferase 2 (NAT 2) e citocromo P450 enzimas CYP2D6 e CYP2C19) proporcionou uma explicação para o porquê de alguns pacientes não obterem os efeitos esperados do fármaco ou apresentarem uma resposta exagerada ao fármaco e uma grave toxicidade após tomarem a dose padrão e segura de um fármaco. Estes polimorfismos são expressos em dois fenótipos na população, o metabolizador extensivo (EM) e o metabolizador fraco (PM). A prevalência de PM é diferente entre diferentes populações. Por exemplo, o gene que codifica para CYP2D6 é altamente polimórfico e foram identificadas várias mutações em PM, todas conduzindo à ausência de CYP2D6 funcional. Os metabolizadores fracos de CYP2D6 e CYP2C19 sofrem muito frequentemente uma resposta exagerada ao fármaco e efeitos secundários quando recebem as doses padrão. Se um metabolito for a porção terapêuticamente activa, o PM não apresenta qualquer resposta terapêutica, como demonstrado para o efeito analgésico de codeína mediado pelo seu metabolito formado por CYP2D6, morfina. No outro extremo estão os denominados metabolizadores ultra-rápidos que não respondem a doses padrão. Recentemente, a base molecular do metabolismo ultra-rápido foi identificada como sendo devida a amplificação do gene de CYP2D6.

Alternativamente, um método denominado o "traçado do perfil de expressão génica" ("*gene expression profiling*"), pode ser utilizado para identificar genes que prevêm a resposta a fármacos. Por exemplo, a expressão génica de um animal que toma um fármaco (e.g., uma molécula de CRSP ou um modulador de CRSP da presente invenção) pode dar uma indicação de se as vias génicas relacionadas com toxicidade foram "ligadas".

A informação gerada a partir de mais do que uma das abordagens farmacogenómicas anteriores pode ser utilizada para determinar os regimes de dosagem e tratamento apropriados para

o tratamento profilático ou terapêutico de um indivíduo. Este conhecimento, quando aplicado ao doseamento ou selecção do fármaco, pode evitar reacções adversas ou a falha da terapêutica e assim aumentar a eficiência terapêutica ou profiláctica aquando do tratamento de um indivíduo com uma molécula de CRSP ou um modulador de CRSP, tal como um modulador identificado por um dos ensaios de rastreio exemplificativos aqui descritos.

A presente invenção é ainda ilustrada pelos exemplos que se seguem que não deverão ser entendidos como limitantes.

EXEMPLOS

Exemplo 1: Isolamento E Caracterização de ADNc de CRSP-1 Humana

Neste exemplo, é descrito o isolamento e a caracterização do gene que codifica CRSP-1 humana (também referida como "CRISPY-1" ou "TANGO 59").

Isolamento do ADNc de CRSP-1 humana

A invenção baseia-se, pelo menos em parte, na identificação de um gene humano que codifica uma proteína segregada, referida aqui como Proteína-1 Segregada Rica em cisteínas (CRSP-1). Isolou-se um ADNc parcial utilizando um método da sequência de sinal Trap. Esta metodologia tira vantagem do facto de moléculas como a CRSP possuírem uma sequência de sinal amino terminal que dirige certas proteínas segregadas e ligadas a membranas através do aparelho de secreção celular.

Resumidamente, preparou-se uma biblioteca de ADNc iniciada aleatoriamente utilizando ARNm preparado a partir de tecido cerebral fetal humano (Clontech, Palo Alto CA) utilizando o *kit* Stratagene-ZAP-ADNc Síntese™, (n.º de catálogo 20041). Ligou-se o ADNc ao vector de expressão de mamífero pTrap adjacente a um ADNc que codifica fosfatase alcalina placentária a que falta um sinal de secreção. Os plasmídeos foram transformados em *E. coli* e o ADN foi preparado utilizando o *kit* de purificação de ADN Wizard™ (Promega). O ADN foi transfectado em células COS-7 com Lipofectamine™ (Gibco-BRL). Após 48 horas de incubação,

ensaiaram-se os sobrenadantes das células COS quanto a fosfatase alcalina num contador de cintilação Wallac Micro-Beta utilizando o *kit* Phospha-Light™ (Tropix Inc., n.º de Catálogo BP300). Os ADN plasmídicos individuais que foram positivos no ensaio de secreção de Fosfatase alcalina das células COS foram adicionalmente analisados por sequenciação do ADN utilizando procedimentos *standard*.

Utilizando um ADNc parcial isolado através do método atrás descrito, clonou-se um ADNc de comprimento completo que codifica a CRSP-1 humana. A sequência nucleotídica que codifica a proteína CRSP-1 humana de comprimento completo está apresentada na Figura 1 e definida em SEQ ID NO:1. A proteína de comprimento completo codificada por este ácido nucleico é constituída por cerca de 350 aminoácidos e possui a sequência de aminoácidos apresentada na Figura 1 e definida em SEQ ID NO:2. A porção de codificação (quadro de leitura aberto) de SEQ ID NO:1 está definida em SEQ ID NO:3.

Análise de CRSP-1 Humana

A determinação do perfil de hidrofobicidade da CRSP-1 humana possuindo a sequência de aminoácidos definida em SEQ ID NO:2 indicou a presença de uma região hidrófoba desde cerca do aminoácido 1 até cerca do aminoácido 23 de SEQ ID NO:2. A análise adicional da sequência de aminoácidos de SEQ ID NO:2 utilizando programas de previsão de péptidos de sinal previu a presença de um péptido de sinal desde cerca do aminoácido 1 até cerca do aminoácido 19, 21 ou 23 de SEQ ID NO:2. Deste modo, a proteína CRSP-1 madura inclui cerca de 327, 329 ou 331 aminoácidos que se estendem desde cerca do aminoácido 20, 22 ou 24 até cerca do aminoácido 350 de SEQ ID NO:2. A presença da sequência de sinal, em adição ao facto de a CRSP-1 ter sido identificada utilizando um sistema de sequências de sinal Trap, indica que a CRSP-1 é uma proteína segregada. Adicionalmente, a previsão de um tal péptido de sinal e local de clivagem do péptido de sinal pode ser feita, por exemplo, utilizando o algoritmo de computador SIGNALP (Henrik, *et al.* (1997) *Protein Engineering* 10:1-6).

Adicionalmente, quando o ADNc que codifica a CRSP-1 humana foi expresso em células bacterianas com um marcador

FLAG C-terminal, verificou-se que a proteína CRSP-1 produto marcada com FLAG é segregada para o meio celular.

O exame da sequência de ADNc representada na Figura 1 mostra que a CRSP-1 humana é particularmente rica em resíduos de cisteína. Como mostrado na Figura 1, a CRSP-1 contém 20 resíduos de cisteína localizados entre o aminoácido 147 e o aminoácido 284 de SEQ ID NO:2. Estes resíduos de cisteína podem possivelmente formar 10 pontes de dissulfureto.

Uma pesquisa BLAST (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403) das sequências de nucleótidos e de aminoácidos de CRSP-1 revelou que a CRSP-1 é significativamente similar a um ADNc de galinha que codifica uma proteína de função desconhecida possuindo o N.º de Acesso no GenBank D26311. Este ADNc foi isolado de uma biblioteca de ADNc de cristalino de galinha e mostrou-se que era expresso em fibras de cristalino e epitélio de cristalino, mas não na retina neural nem em células hepáticas. (Sawada *et al.* (1996) *Int. J. Dev. Biol.* 40:531). A CRSP-1 e a proteína de galinha possuem 56% de identidade de sequência de aminoácidos e 72% de similaridade de sequência de aminoácidos. A similaridade de sequência de aminoácidos entre a proteína de galinha e a CRSP-1 humana é particularmente elevada no domínio rico em cisteínas de CRSP-1 que está localizado entre os aminoácidos 147 e 284 de SEQ ID NO:2. Em particular, os 20 resíduos de cisteínas de CRSP-1 localizados nesta região estão presentes na proteína de galinha (veja-se a Figura 3).

Dois genes recentemente identificados num rastreo de supressores da formação de glioblastoma (Ligon *et al.*, 1997 *Oncogene* 14 1075-1081) também mostrou homologia significativa com hCRSP-1. Estes genes, RIG ("Regulados em Glioblastoma") e RIG-like 7-1 (N.ºs de Acesso no GenBank U32331 e AF034208, respectivamente), foram identificados num rastreo diferencial de ARNm regulado pela introdução de uma cópia normal do cromossoma 10 numa linha celular de glioblastoma portadora de uma deleção no cromossoma 10 que promove a tumorigênese. Um diagrama esquemático que resume a relação entre as sequências da CRSP-1 e os genes RIG é apresentado na Figura 9. A região de identidade indicada entre CRSP-1, RIG e RIG-like 7-1 compreende uma região da UTR a 3' do ARNm da CRSP-1 humana.

Este ADNc foi inicialmente isolado no rastreamento diferencial descrito e subsequentemente foi utilizado para isolar mensagens de RIG e RIG-like 7-1 de comprimento completo, como representado. RIG-like 7-1 é altamente homólogo (96,8% similar como determinado utilizando o programa de alinhamento de proteínas de Lipman-Pearson, *K-tuple*: 2; Penalidade por Hiato: 4, Penalidade por Comprimento de Hiato: 12; e 64,7% homólogo utilizando o programa de alinhamento de ADN de Wilbur Lipman, *K-tuple*: 3; Penalidade por Hiato: 3; Janela: 20) a CRSP-1 embora à proteína codificada falte a sequência de sinal N-terminal e não seja portanto previsível que seja uma proteína segregada. O ARNm de RIG de comprimento completo apenas apresenta homologia com CRSP-1 na região do ADNc inicial. Estes resultados associam a CRSP-1 com glioblastoma humano e sugerem que a CRSP-1 pode ser importante na supressão do fenótipo tumorigénico. Um papel no glioblastoma é também consistente com o elevado nível de expressão de ARNm de CRSP-1 observado em tecido cerebral humano. Em adição, a co-localização dos genes de CRSP-1, RIG e RIG-like numa região do cromossoma 11 (11p15.1) implicada no desenvolvimento de astrocitoma maligno humano (Ligon *et al.* 1997 *Oncogene* 14 1075-1081) indica adicionalmente um papel para esses genes na tumorigénese.

A proteína CRSP-1 humana possui também alguma similaridade de sequência de aminoácidos com metalotioneína, particularmente no domínio rico em cisteínas.

Distribuição nos Tecidos de ARNm de CRSP-1

Este Exemplo descreve a distribuição nos tecidos de ARNm de CRSP-1, como determinado por hibridação *Northern blot*.

As hibridações *Northern blot* com as várias amostras de ARN foram realizadas sob condições *standard* e lavadas sob condições rigorosas, *i.e.*, SSC 0,2x a 65°C. Em cada amostra, a sonda hibridou com um único ARN de cerca de 2,5 kb. Os resultados da hibridação da sonda com várias amostras de ARNm estão descritos adiante.

A hibridação de um *Northern blot* de múltiplos tecido fetais (MTN - *Multiple Tissue Northern*) da Clontech (Clontech,

LaJolla, CA) contendo ARN de cérebro, pulmão, fígado e rim fetais indicou a presença de elevados níveis de ARNm de CRSP-1 em cérebro, pulmão fetais e níveis ligeiramente menores de ARNm de GRSP-1 em rim fetal. Contudo, não se encontrou qualquer nível significativo de ARNm de CRSP-1 em fígado fetal.

A hibridação de um *Northern blot* de múltiplos tecido fetais humanos (MTN) da Clontech (Clontech, LaJolla, CA) contendo ARN de coração, cérebro, placenta, pulmão, fígado, músculo esquelético, rim e pâncreas adultos com uma sonda de CRSP-1 humana indicou a presença de níveis elevados de ARNm de CRSP-1 em coração, níveis ligeiramente menores em cérebro, e níveis muito menores em placenta e pulmão. Foi também encontrado algum ARNm CRSP-1 em músculo esquelético adulto. Contudo, não se observaram níveis significativos de ARNm de CRSP-1 em fígado, rim ou pâncreas adultos. De forma interessante, o gene de galinha que é homólogo a CRSP-1 também não foi expresso em níveis detectáveis no fígado (Sawada et al. (1996) *Int. J. Dev. Biol.* 40:531).

Outra hibridação de um *Northern blot* de múltiplos tecido fetais humanos (MTN) da Clontech (Clontech, LaJolla, CA) incluindo ARN de medula espinal adulta revelou níveis elevados de expressão de CRSP-1 em ARNm isolado de medula espinal adulta.

A análise *in situ* revelou ainda os seguintes padrões de expressão quando secções de tecidos foram hibridadas com sondas de CRSP-1. O ARNm de CRSP-1 foi expresso em coração, pulmão, aorta, olho (retina e cristalino) e cérebro (córtex) em embriões de ratinho de 14,5 dias. O ARNm de CRSP-1 foi expresso em coração, pulmão, olho (retina e cristalino), cérebro (cortical) e cartilagem (pé, traqueia, laringe, cabeça, esterno e coluna espinal) em tecidos de ratinho com 1,5 dias pós-natal. O ARNm de CRSP-1 foi expresso em cérebro (córtex, hipocampo, tronco cerebral), coração (apenas aurículas), olho (retina e camada externa do cristalino) em tecidos de ratinho adulto.

Assim, a CRSP-1 é expressa de uma maneira específica do tecido, com a expressão mais forte observada em cérebro, coração e medula espinal.

Resultados de Expressão de Proteína CRSP-1

Construções

Prepararam-se construções de expressão para duas formas de CRSP-1 utilizando o vector de expressão de mamífero pMET-stop. A forma-1 incluía um ADNc que incorpora a sequência de codificação da proteína CRSP-1 completa de 350aa (CRSP-1 flag.long) e a forma-2 incluía a sequência de codificação da proteína CRSP-1 inteira excepto os 18 aminoácidos finais (CRSP-1 flag.short). Uma sequência C-terminal que codifica o epítipo FLAG (DYKDDDDK) foi adicionada a ambas as formas de CRSP-1 para facilidade de detecção e purificação. Os ADNc de CRSP-1flag foram gerados por PCR a partir de um molde de ADNc de CRSP-1 de comprimento completo e foram ligados a pMET-stop utilizando os locais de restrição EcoR1 e Sall1.

Ensaio de Transfecção - Expressão em Pequena Escala

As construções de expressão para CRSP-1flag.long e CRSP-1flag.short foram transfectadas em células 293T utilizando 10 µl de Lipofectamine (GIBCO/BRL) e 2 µg de ADN por poço de uma placa de 6 poços de células que eram 70-80% confluentes. Após 5 horas a 37°C, as células foram alimentadas com 1 ml de FCS a 20%/DMEM. Após incubação durante a noite a 37°C, as células foram condicionadas em 1 ml de OptiMEM durante 48 horas a 37°C. Amostras de sobrenadante e peletes celulares foram solubilizadas em tampão de gel de SDS-PAGE ebuliente, corridas num gel de SDS-PAGE a 4-20%, transferidas para uma membrana de nylon e sondadas com o anticorpo monoclonal anti-FLAG M2. Tanto as amostras do sobrenadante como as amostras da pelete apresentaram imunoreactividade significativa dentro de uma gama de pesos moleculares de 40-65 kDa num filme autorradiográfico utilizando um anticorpo secundário conjugado com HRP e reagentes de detecção de ECL. Assim, ambas as formas de CRSP-1 testadas são segregadas por células 293T, confirmando assim experimentalmente que a CRSP-1 é uma proteína segregada. Deverá notar-se que os pesos moleculares de ambas as formas de CRSP-1 testadas são maiores do que o previsto a partir da sequência de aminoácidos, sugerindo que

as proteínas CRSP-1 segregadas por células 293T podem estar glicosiladas. Isto é consistente com a presença de quatro potenciais locais para glicosilação ligada a N na proteína CRSP-1.

Produção de Proteína CRSP-1 em Grande Escala

Para o aumento de escala da expressão da proteína CRSP-1flag.long, placas (30 x 150mm) de células 293T a 70-80% da confluência foram transfectadas com 27 µg de ADN, 100 µl de Lipofectamine em 18 ml de OptiMEM durante 5 horas a 37°C. Adicionaram-se 18 ml de FCS a 10%/DMEM a cada placa e incubou-se durante a noite a 37°C. 24 horas após o início da transfecção, aspirou-se o sobrenadante da transfecção e adicionaram-se 35 ml de OptiMEM a cada placa e incubaram-se as placas a 37°C durante 72 horas. Colheu-se o meio condicionado, centrifugou-se a 4000 rpm durante 30 min. a 4°C, e filtrou-se através de uma unidade de filtração de 0,45 microns. Passaram-se 1100 ml através de uma coluna de afinidade anti-FLAG M2 de 1,6 x 10 cm pré-equilibrada em tampão PBS de pH 7,4 a um caudal de 2,0 ml por minuto. Após lavagem com 200 ml de tampão PBS de pH 7,4, eluiu-se o material ligado através de um passo de tampão de Glicina 200 mM de pH 3,0 e colheram-se fracções de 0,5 ml. Após eluição, foi detectado um pico significativo de proteína por absorvância a 280 nm. As amostras correspondentes ao meio condicionado, o fluido de passagem e as fracções eluídas, foram analisados por SDS-PAGE com coloração com azul de Coomassie e com prata e por análise *Western blot* como descrito atrás. Foi detectada imunoreactividade significativa dentro de uma gama de pesos moleculares de 40-65 kDa em meio condicionado e nas fracções eluídas mas não na amostra de fluido de passagem, indicando que a proteína CRSP-1 flag.long segregada se ligava especificamente à coluna de afinidade e era eluída eficientemente pelas condições descritas. A coloração com azul de Coomassie dos géis de SDS-PAGE sugeriu que a proteína imunoreactiva predominante constituía >90% da proteína presente no pico de proteína ligada e eluída. As fracções de pico de proteína eluída foram reunidas e dialisadas contra Solução Salina Tamponada com Fosfato resultando um volume de 4 ml de proteína CRSP-1 flag.long recombinante numa concentração de aproximadamente 1 mg/ml.

Exemplo 2: Isolamento E Caracterização de ADNc de CRSP-2, CRSP-3 e CRSP-4 Humanas

Para identificar novas proteínas relacionadas com a CRSP-1, utilizou-se a sequência de aminoácidos da CRSP-1 humana para rastrear bases de dados privadas e a dbEST utilizando TBLASTN (WashUversion, 2.0, matriz de pesquisa BLOSUM62). Deste rastreio, identificaram-se quatro sequências parciais distintas da proteína que apresentavam homologia significativa com a CRSP-1 humana. Estas sequências foram designadas como sequências parciais para as novas hCRSP; hCRSP-2, h-CRSP-3, h-CRSP-4 e h-CRSP-like-n.

A sequência de proteína parcial hCRSP-2 era derivada de um único clone de dbEST com o Número de Acesso AA565546. Este clone foi subsequentemente obtido a partir da colecção IMAGE e totalmente sequenciado para definir a totalidade da sequência de hCRSP-2 representada na Figura 2.

A sequência de proteína parcial hCRSP-3 era derivada de um clone jthKb075a10 com o código de identificação de sequência jthKb075a10. Este clone foi ainda sequenciado e esta informação de sequência foi combinada com sequências adicionais de dbEST utilizando o programa de montagem Phrap (P. Green, U. Washington) para definir a totalidade da sequência hCRSP-3 representada na Figura 3. O ADN para o clone jthKb075a10 foi depositado na ATCC com o N.º de Acesso 98633. A análise *Northern blot* de vários tecidos utilizando uma sonda específica para hCRSP-3 revelou que a expressão do ARNm de CRSP-3 estava restringida à placenta. A expressão do ARNm de CRSP-3 não foi ainda demonstrada em outros tecidos humanos adultos normais.

Foi recentemente descrito um homólogo de CRSP-3 de murino e de *Xenopus* denominado dkk-1 ("*dickkopf*", em alemão) que se propõe codificar uma molécula de sinalização segregada (*Nature*, 391, 357-368). A dkk-1 é descrita como um poderoso indutor da formação da cabeça durante o início do desenvolvimento de *Xenopus*, em que o seu mecanismo de acção envolve aparentemente a inibição da família *wnt* de factores segregados.

Dada a homologia entre CRSP-3 e *dkk-1*, é sugerido que a hCRSP-3 pode apresentar uma profunda actividade indutora da cabeça durante o início do desenvolvimento embrionário humano. Adicionalmente, pensa-se que os mecanismos de acção de *dkk-1* envolvem antagonismo de membros da família *wnt* de proteínas segregadas. A super-expressão de proteínas *wnt* foi associada a certas malignidades. Por exemplo, a activação da expressão de *wnt-1* por inserção proviral de MMTV causa cancro da mama no ratinho (veja-se Cadigan e Nusse, *Genes and Dev.*, 1997 11 3286-3305). Portanto, pelo menos a CRSP-3, e potencialmente outros membros da família CRSP, podem também desempenhar um papel na supressão de sinalização de *wnt* no cancro.

A sequência de proteína parcial h-CRSP-4 era derivada de um único clone de dbEST com o Número de Acesso W55979. Este clone foi subsequentemente obtido a partir da colecção IMAGE e totalmente sequenciado para definir a sequência de hCRSP-4 representada na Figura 4. A análise *Northern blot* de vários tecidos utilizando uma sonda específica para hCRSP-4 revelou que o ARNm de CRSP-4 era expresso em vários tecidos humanos adultos. A expressão do ARNm de CRSP-4 foi maior no coração, cérebro, placenta, pulmão e músculo esquelético.

A sequência de proteína parcial h-CRSP-like-n era derivada de um único clone de dbEST com o Número de Acesso AA397836. Este clone foi subsequentemente obtido a partir da colecção IMAGE e totalmente sequenciado para definir a totalidade da sequência de hCRSP-like-n representada na Figura 5.

Estrutura das Proteínas da Família CRSP

Um alinhamento das sequências de aminoácidos da CRSP-1 humana, CRSP-2 humana, CRSP-3 humana e CRSP-4 humana, é apresentado na Figura 6. Os resíduos de aminoácido que são conservados entre os membros da família de CRSP humanas estão em caixas. Os 20 resíduos de cisteína conservados estão indicados por um asterisco. Os domínios ricos em cisteínas estão indicados como CRD-1 e CRD-2. A estrutura de domínios das proteínas CRSP de comprimento completo da presente invenção está representada na Figura 7. A homologia de aminoácidos e de nucleótidos entre os membros da família CRSP é como se segue nas Tabelas 1 e 2.

Tabela 1:

	CRSP-1	CRSP-2	CRSP-3	CRSP-4	mdkk-1	xdkk-1	CLFEST
CRSP-1	100	16,0	18,6	15,1	18,5	16,5	53,0
CRSP-2		100	33,7	35,2	32,6	33,7	16,2
CRSP-3			100	33,1	80,2	53,5	17,4
CRSP-4				100	30,5	33,7	12,5

As percentagens de homologias de aminoácidos foram determinadas utilizando o programa ALIGN, (versão 2.0) (pacote de *software* GCG) utilizando uma tabela de resíduos ponderal PAM120, uma penalidade por comprimento de hiatos de 12, e uma penalidade por hiatos de 4.

Tabela 2:

	CRSP-1	CRSP-2	CRSP-3	CRSP-4	mdkk-1	xdkk-1	CLFEST
CRSP-1	100	30,0	37,2	34,7	31,5	45,4	58,8
CRSP-2		100	43,0	35,9	38,8	38,4	36,7
CRSP-3			100	59,3	66,4	53,7	32,1
CRSP-4				100	38,8	38,4	36,7

As percentagens de homologias de nucleótidos foram determinadas utilizando o programa de alinhamento de ADN Wilbur Lipman, *K-tuple*: 3; Penalidade por hiatos: 3; Janela: 20.

LISTAGEM DE SEQUÊNCIAS

(1) INFORMAÇÃO GERAL:

(i) REQUERENTE:

- (A) NOME: MILLENNIUM BIOTHERAPEUTICS, INC.
- (B) RUA: 620 MEMORIAL DRIVE
- (C) CIDADE: CAMBRIDGE
- (D) ESTADO: MASSACHUSETTS
- (E) PAÍS: EUA
- (F) CÓDIGO POSTAL: 02319
- (G) TELEFONE:
- (H) TELEFAX:

(ii) TÍTULO DA INVENÇÃO: NOVAS MOLÉCULAS DE PROTEÍNA CRSP E DE ÁCIDO NUCLEICO DE CRSP E SUAS UTILIZAÇÕES

- (iii) NÚMERO DE SEQUÊNCIAS: 18
 - (iv) ENDEREÇO PARA CORRESPONDÊNCIA:
 - (A) DESTINATÁRIO: LAHIVE & COCKFIELD, LLP
 - (B) RUA: 28 STATE STREET
 - (C) CIDADE: BOSTON
 - (D) ESTADO: MASSACHUSETTS
 - (E) PAÍS: EUA
 - (F) CÓDIGO POSTAL: 02109
 - (v) FORMATO LEGÍVEL EM COMPUTADOR:
 - (A) TIPO DE MEIO: Disquete
 - (B) COMPUTADOR: PC compatível com IBM
 - (C) SISTEMA OPERATIVO: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Versão #1.25
 - (vi) DADOS DO PRESENTE PEDIDO:
 - (A) NÚMERO DE PEDIDO: PCT/US98/
 - (B) DATA DE APRESENTAÇÃO: 16 ABRIL 1998
 - (C) CLASSIFICAÇÃO:
 - (vii) DADOS DE PEDIDOS ANTERIORES:
 - (A) NÚMERO DE PEDIDO: US 08/843 704
 - (B) DATA DE APRESENTAÇÃO: 16 ABRIL 1997
 - (C) NÚMERO DE PEDIDO: US 08/842 898
 - (D) DATA DE APRESENTAÇÃO: 17 ABRIL 1997
 - (E) NÚMERO DE PEDIDO: 60/071 589
 - (F) DATA DE APRESENTAÇÃO: 15 JANEIRO 1998
 - (G) NÚMERO DE PEDIDO: 09/009,802
 - (H) DATA DE APRESENTAÇÃO: 20 JANEIRO 1998
 - (viii) INFORMAÇÃO SOBRE REPRESENTANTE/AGENTE:
 - (A) NOME: MANDRAGOURAS, AMY E.
 - (B) NÚMERO DE REGISTRO: 36,207
 - (C) NÚMERO DE REFERÊNCIA/PROCESSO: MEI-008CPPC
 - (ix) INFORMAÇÃO PARA TELECOMUNICAÇÕES:
 - (A) TELEFONE: (617)227-7400
 - (B) TELEFAX: (617)742-4214
- (2) INFORMAÇÃO PARA SEQ ID NO:1:
- (i) CARACTERÍSTICAS DA SEQUÊNCIA:
 - (A) COMPRIMENTO: 2479 pares de bases
 - (B) TIPO: ácido nucleico
 - (C) TIPO DE CADEIA: simples
 - (D) TOPOLOGIA: linear
 - (ii) TIPO DE MOLÉCULA: ADNc
 - (ix) CARACTERÍSTICA:
 - (A) NOME/CHAVE: CDS
 - (B) LOCALIZAÇÃO: 38..1087

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:1:

GGCACGAGGG GCGGGCGGCT GCGGGCGCAG AGCGGAG ATG CAG CGG CTT GGG GCC	55
Met Gln Arg Leu Gly Ala	
1 5	
ACC CTG CTG TGC CTG CTG CTG GCG GCG GCG GTC CCC ACG GCC CCC GCG	103
Thr Leu Leu Cys Leu Leu Leu Ala Ala Ala Val Pro Thr Ala Pro Ala	
10 15 20	
CCC GCT CCG ACG GCG ACC TCG GCT CCA GTC AAG CCC GGC CCG GCT CTC	151
Pro Ala Pro Thr Ala Thr Ser Ala Pro Val Lys Pro Gly Pro Ala Leu	
25 30 35	
AGC TAC CCG CAG GAG GAG GCC ACC CTC AAT GAG ATG TTC CGC GAG GTT	199
Ser Tyr Pro Gln Glu Glu Ala Thr Leu Asn Glu Met Phe Arg Glu Val	
40 45 50	
GAG GAA CTG ATG GAG GAC ACG CAG CAC AAA TTG CGC AGC GCG GTG GAA	247
Glu Glu Leu Met Glu Asp Thr Gln His Lys Leu Arg Ser Ala Val Glu	
55 60 65 70	
GAG ATG GAG GCA GAA GAA GCT GCT GCT AAA GCA TCA TCA GAA GTG AAC	295
Glu Met Glu Ala Glu Glu Ala Ala Ala Lys Ala Ser Ser Glu Val Asn	
75 80 85	
CTG GCA AAC TTA CCT CCC AGC TAT CAC AAT GAG ACC AAC ACA GAC ACG	343
Leu Ala Asn Leu Pro Pro Ser Tyr His Asn Glu Thr Asn Thr Asp Thr	
90 95 100	
AAC GTT GGA AAT AAT ACC ATC CAT GTG CAC CGA GAA ATT CAC AAG ATA	391
Asn Val Gly Asn Asn Thr Ile His Val His Arg Glu Ile His Lys Ile	
105 110 115	

ACC AAC AAC CAG ACT GGA CAA ATG GTC TTT TCA GAG ACA GTT ATC ACA	439
Thr Asn Asn Gln Thr Gly Gln Met Val Phe Ser Glu Thr Val Ile Thr	
120 125 130	
TCT GTG GGA GAC GAA GAA GGC AGA AGG AGC CAC GAG TGC ATC ATC GAC	487
Ser Val Gly Asp Glu Glu Gly Arg Arg Ser His Glu Cys Ile Ile Asp	
135 140 145 150	
GAG GAC TGT GGG CCC AGC ATG TAC TGC CAG TTT GCC AGC TTC CAG TAC	535
Glu Asp Cys Gly Pro Ser Met Tyr Cys Gln Phe Ala Ser Phe Gln Tyr	
155 160 165	
ACC TGC CAG CCA TGC CGG GGC CAG AGG ATG CTC TGC ACC CGG GAC AGT	583
Thr Cys Gln Pro Cys Arg Gly Gln Arg Met Leu Cys Thr Arg Asp Ser	
170 175 180	
GAG TGC TGT GGA GAC CAG CTG TGT GTC TGG GGT CAC TGC ACC AAA ATG	631
Glu Cys Cys Gly Asp Gln Leu Cys Val Trp Gly His Cys Thr Lys Met	
185 190 195	
GCC ACC AGG GGC AGC AAT GGG ACC ATC TGT GAC AAC CAG AGG GAC TGC	679
Ala Thr Arg Gly Ser Asn Gly Thr Ile Cys Asp Asn Gln Arg Asp Cys	
200 205 210	
CAG CCG GGG CTG TGC TGT GCC TTC CAG AGA GGC CTG CTG TTC CCT GTG	727
Gln Pro Gly Leu Cys Cys Ala Phe Gln Arg Gly Leu Leu Phe Pro Val	
215 220 225 230	
TGC ACA CCC CTG CCC GTG GAG GGC GAG CTT TGC CAT GAC CCC GCC AGC	775
Cys Thr Pro Leu Pro Val Glu Gly Glu Leu Cys His Asp Pro Ala Ser	
235 240 245	
CGG CTT CTG GAC CTC ATC ACC TGG GAG CTA GAG CCT GAT GGA GCC TTG	823
Arg Leu Leu Asp Leu Ile Thr Trp Glu Leu Glu Pro Asp Gly Ala Leu	
250 255 260	
GAC CGA TGC CCT TGT GCC AGT GGC CTC CTC TGC CAG CCC CAC AGC CAC	871
Asp Arg Cys Pro Cys Ala Ser Gly Leu Leu Cys Gln Pro His Ser His	
265 270 275	
AGC CTG GTG TAT GTG TGC AAG CCG ACC TTC GTG GGG AGC CGT GAC CAA	919
Ser Leu Val Tyr Val Cys Lys Pro Thr Phe Val Gly Ser Arg Asp Gln	
280 285 290	
GAT GGG GAG ATC CTG CTG CCC AGA GAG GTC CCC GAT GAG TAT GAA GTT	967
Asp Gly Glu Ile Leu Leu Pro Arg Glu Val Pro Asp Glu Tyr Glu Val	
295 300 305 310	
GGC AGC TTC ATG GAG GAG GTG CGC CAG GAG CTG GAG GAC CTG GAG AGG	1015
Gly Ser Phe Met Glu Glu Val Arg Gln Glu Leu Glu Asp Leu Glu Arg	
315 320 325	
AGC CTG ACT GAA GAG ATG GCG CTG AGG GAG CCT GCG GCT GCC GCC GCT	1063
Ser Leu Thr Glu Glu Met Ala Leu Arg Glu Pro Ala Ala Ala Ala Ala	
330 335 340	

GCA CTG CTG GGA AGG GAA GAG ATT TAGATCTGGA CCAGGCTGTG GGTAGATGTG 1117
 Ala Leu Leu Gly Arg Glu Glu Ile
 345 350
 CAATAGAAAT AGCTAATTTA TTTCCCCANG TGTGTGCTTT AAGCGTGGGC TGACCAGGCT 1177
 TCTTCCTACA TCTTCTTCCC AGTAAGTTTC CCCTCTGGCT TGACAGCATG AGGTGTTGTG 1237
 CATTTGTTCA GCTCCCCCAG GCTGTTCTCC AGGCTTCACA GTCTGGTGCT TGGGAGAGTC 1297
 AGGCAGGGTT AAACGCAGG AGCAGTTTGC CACCCCTGTC CAGATTATTG GCTGCTTTGC 1357
 CTCTACCAGT TGGCAGACAG CCGTTTGTTC TACATGGCTT TGATAATTGT TTGAGGGGAG 1417
 GAGATGGAAA CAATGTGGAG TCTCCCTCTG ATTGGTTTGG GGGAAATGTG GAGAAGAGTG 1477
 CCCTGCTTTG CAAACATCAA CCTGGCAAAA ATGCAACAAA TGAATTTTCC ACGCAGTTCT 1537
 TTCCATGGGC ATAGGTAAGC TGTGCCTTCA GCTGTTGCAG ATGAAATGTT CTGTTACACC 1597
 TGCATTACAT GTGTTTATTTC ATCCAGCAGT GTTGCTCAGC TCCTACCTCT GTGCCAGGGC 1657
 AGCATTTTCA TATCCAAGAT CAATCCCTC TCTCAGCACA GCCTGGGGAG GGGGTCATTG 1717
 TTCTCCTCGT CCATCAGGGA TTTCAGAGGC TCAGAGACTG CAAGCTGCTT GCCCAAAGTCA 1777
 CACAGCTAGT GAAGACCAGA GCAGTTTCAT CTGGTTGTGA CTCTAAGCTC AGTGCTCTCT 1837
 CCACTACCCC ACACCAGCCT TGGTGCCACC AAAAGTGCTC CCCAAAAGGA AGGAGAATGG 1897
 GATTTTTTCTT TTGAGGCATG CACATCTGGA ATTAAGGTCA AACTAATTCT CACATCCCTC 1957
 TAAAAGTAAA CTA CTACTGTTAG GAACAGCAGT GTTCTCACAG TGTGGGGCAG CCGTCCTTCT 2017
 AATGAAGACA ATGATATTGA CACTGTCCCT CTTTGGCAGT TGCATTAGTA ACTTTGAAAG 2077
 GTATATGACT GAGCGTAGCA TACAGGTTAA CCTGCAGAAA CAGTACTTAG GTAATTGTAG 2137
 GCGGAGGATT ATAAATGAAA TTTGCAAAAT CACTTAGCAG CAACTGAAGA CAATTATCAA 2197
 CCACGTGGAG AAAATCAAAC CGAGCAGGGC TGTGTGAAAC ATGGTTGTAA TATGCGACTG 2257
 CGAACACTGA ACTCTACGCC ACTCCACAAA TGATGTTTTC AGGTGTCATG GACTGTTGCC 2317
 ACCATGTATT CATCCAGAGT TCTTAAAGTT TAAAGTTGCA CATGATTGTA TAAGCATGCT 2377
 TTCTTTGAGT TTAAATTAT GTATAAACAT AAGTTGCATT TAGAAATCAA GCATAAATCA 2437
 CTFCAACTGC TAAAAAATAA AAAAAAATAA AAAAAAATAA AA 2479

(2) INFORMAÇÃO PARA SEQ ID NO:2:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 350 aminoácidos
- (B) TIPO: aminoácido
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:2:

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

TTG CGC AGC GCG GTG GAA GAG ATG GAG GCA GAA GAA GCT GCT GCT AAA	240
Leu Arg Ser Ala Val Glu Glu Met Glu Ala Glu Glu Ala Ala Ala Lys	
65 70 75 80	
GCA TCA TCA GAA GTG AAC CTG GCA AAC TTA CCT CCC AGC TAT CAC AAT	288
Ala Ser Ser Glu Val Asn Leu Ala Asn Leu Pro Pro Ser Tyr His Asn	
85 90 95	
GAG ACC AAC ACA GAC ACG AAC GTT GGA AAT AAT ACC ATC CAT GTG CAC	336
Glu Thr Asn Thr Asp Thr Asn Val Gly Asn Asn Thr Ile His Val His	
100 105 110	
CGA GAA ATT CAC AAG ATA ACC AAC AAC CAG ACT GGA CAA ATG GTC TTT	384
Arg Glu Ile His Lys Ile Thr Asn Asn Gln Thr Gly Gln Met Val Phe	
115 120 125	
TCA GAG ACA GTT ATC ACA TCT GTG GGA GAC GAA GAA GGC AGA AGG AGC	432
Ser Glu Thr Val Ile Thr Ser Val Gly Asp Glu Glu Gly Arg Arg Ser	
130 135 140	
CAC GAG TGC ATC ATC GAC GAG GAC TGT GGG CCC AGC ATG TAC TGC CAG	480
His Glu Cys Ile Ile Asp Glu Asp Cys Gly Pro Ser Met Tyr Cys Gln	
145 150 155 160	
TTT GCC AGC TTC CAG TAC ACC TGC CAG CCA TGC CGG GGC CAG AGG ATG	528
Phe Ala Ser Phe Gln Tyr Thr Cys Gln Pro Cys Arg Gly Gln Arg Met	
165 170 175	
CTC TGC ACC CGG GAC AGT GAG TGC TGT GGA GAC CAG CTG TGT GTC TGG	576
Leu Cys Thr Arg Asp Ser Glu Cys Cys Gly Asp Gln Leu Cys Val Trp	
180 185 190	
GGT CAC TGC ACC AAA ATG GCC ACC AGG GGC AGC AAT GGG ACC ATC TGT	624
Gly His Cys Thr Lys Met Ala Thr Arg Gly Ser Asn Gly Thr Ile Cys	
195 200 205	
GAC AAC CAG AGG GAC TGC CAG CCG GGG CTG TGC TGT GCC TTC CAG AGA	672
Asp Asn Gln Arg Asp Cys Gln Pro Gly Leu Cys Cys Ala Phe Gln Arg	
210 215 220	
GGC CTG CTG TTC CCT GTG TGC ACA CCC CTG CCC GTG GAG GGC GAG CTT	720
Gly Leu Leu Phe Pro Val Cys Thr Pro Leu Pro Val Glu Gly Glu Leu	
225 230 235 240	
TGC CAT GAC CCC GCC AGC CGG CTT CTG GAC CTC ATC ACC TGG GAG CTA	768
Cys His Asp Pro Ala Ser Arg Leu Leu Asp Leu Ile Thr Trp Glu Leu	
245 250 255	
GAG CCT GAT GGA GCC TTG GAC CGA TGC CCT TGT GCC AGT GGC CTC CTC	816
Glu Pro Asp Gly Ala Leu Asp Arg Cys Pro Cys Ala Ser Gly Leu Leu	
260 265 270	
TGC CAG CCC CAC AGC CAC AGC CTG GTG TAT GTG TGC AAG CCG ACC TTC	864
Cys Gln Pro His Ser His Ser Leu Val Tyr Val Cys Lys Pro Thr Phe	
275 280 285	

GTG GGG AGC CGT GAC CAA GAT GGG GAG ATC CTG CTG CCC AGA GAG GTC	912
Val Gly Ser Arg Asp Gln Asp Gly Glu Ile Leu Leu Pro Arg Glu Val	
290 295 300	
CCC GAT GAG TAT GAA GTT GGC AGC TTC ATG GAG GAG GTG CGC CAG GAG	960
Pro Asp Glu Tyr Glu Val Gly Ser Phe Met Glu Glu Val Arg Gln Glu	
305 310 315 320	
CTG GAG GAC CTG GAG AGG AGC CTG ACT GAA GAG ATG GCG CTG AGG GAG	1008
Leu Glu Asp Leu Glu Arg Ser Leu Thr Glu Glu Met Ala Leu Arg Glu	
325 330 335	
CCT GCG GCT GCC GCC GCT GCA CTG CTG GGA AGG GAA GAG ATT	1050
Pro Ala Ala Ala Ala Ala Ala Leu Leu Gly Arg Glu Glu Ile	
340 345 350	

(2) INFORMAÇÃO PARA SEQ ID NO:4:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 848 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADnc

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 125..796

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:4:

GAATTCGGCA CGAGAGACGA CGTGCTGAGC TGCCAGCTTA GTGGAAGCTC TGCTCTGGGT	60
GGAGAGCAGC CTCGCTTTGG TGACGCACAG TGCTGGGACC CTCCAGGAGC CCCGGGATTG	120
AAGG ATG GTG GCG GCC GTC CTG CTG GGG CTG AGC TGG CTC TGC TCT CCC	169
Met Val Ala Ala Val Leu Leu Gly Leu Ser Trp Leu Cys Ser Pro	
1 5 10 15	
CTG GGA GCT CTG GTC CTG GAC TTC AAC AAC ATC AGG AGC TCT GCT GAC	217
Leu Gly Ala Leu Val Leu Asp Phe Asn Asn Ile Arg Ser Ser Ala Asp	
20 25 30	
CTG CAT GGG GCC CGG AAG GGC TCA CAG TGC CTG TCT GAC ACG GAC TGC	265
Leu His Gly Ala Arg Lys Gly Ser Gln Cys Leu Ser Asp Thr Asp Cys	
35 40 45	
AAT ACC AGA AAG TTC TGC CTC CAG CCC CGC GAT GAG AAG CCG TTC TGT	313
Asn Thr Arg Lys Phe Cys Leu Gln Pro Arg Asp Glu Lys Pro Phe Cys	
50 55 60	

GCT ACA TGT CGT GGG TTG CGG AGG AGG TGC CAG CGA GAT GCC ATG TGC	361
Ala Thr Cys Arg Gly Leu Arg Arg Arg Cys Gln Arg Asp Ala Met Cys	
65 70 75	
TGC CCT GGG ACA CTC TGT GTG AAC GAT GTT TGT ACT ACG ATG GAA GAT	409
Cys Pro Gly Thr Leu Cys Val Asn Asp Val Cys Thr Thr Met Glu Asp	
80 85 90 95	
GCA ACC CCA ATA TTA GAA AGG CAG CTT GAT GAG CAA GAT GGC ACA CAT	457
Ala Thr Pro Ile Leu Glu Arg Gln Leu Asp Glu Gln Asp Gly Thr His	
100 105 110	
GCA GAA GGA ACA ACT GGG CAC CCA GTC CAG GAA AAC CAA CCC AAA AGG	505
Ala Glu Gly Thr Thr Gly His Pro Val Gln Glu Asn Gln Pro Lys Arg	
115 120 125	
AAG CCA AGT ATT AAG AAA TCA CAA GGC AGG AAG GGA CAA GAG GGA GAA	553
Lys Pro Ser Ile Lys Lys Ser Gln Gly Arg Lys Gly Gln Glu Gly Glu	
130 135 140	
AGT TGT CTG AGA ACT TTT GAC TGT GGC CCT GGA CTT TGC TGT GCT CGT	601
Ser Cys Leu Arg Thr Phe Asp Cys Gly Pro Gly Leu Cys Cys Ala Arg	
145 150 155	
CAT TTT TGG ACG AAA ATT TGT AAG CCA GTC CTT TTG GAG GGA CAG GTC	649
His Phe Trp Thr Lys Ile Cys Lys Pro Val Leu Leu Glu Gly Gln Val	
160 165 170 175	
TGC TCC AGA AGA GGG CAT AAA GAC ACT GCT CAA GCT CCA GAA ATC TTC	697
Cys Ser Arg Arg Gly His Lys Asp Thr Ala Gln Ala Pro Glu Ile Phe	
180 185 190	
CAG CGT TGC GAC TGT GGC CCT GGA CTA CTG TGT CGA AGC CAA TTG ACC	745
Gln Arg Cys Asp Cys Gly Pro Gly Leu Leu Cys Arg Ser Gln Leu Thr	
195 200 205	
AGC AAT CGG CAG CAT GCT CGA TTA AGA GTA TGC CAA AAA ATA GAA AAG	793
Ser Asn Arg Gln His Ala Arg Leu Arg Val Cys Gln Lys Ile Glu Lys	
210 215 220	
CTA TAAATATTTT AAAATAAAGA AGAATCCACA TTGCAAAAAA AAAAAAAAAA AA	848
Leu	

(2) INFORMAÇÃO PARA SEQ ID NO:5:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 224 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:5:

Met Val Ala Ala Val Leu Leu Gly Leu Ser Trp Leu Cys Ser Pro Leu
 1 5 10 15

Gly Ala Leu Val Leu Asp Phe Asn Asn Ile Arg Ser Ser Ala Asp Leu
 20 25 30

His Gly Ala Arg Lys Gly Ser Gln Cys Leu Ser Asp Thr Asp Cys Asn
 35 40 45

Thr Arg Lys Phe Cys Leu Gln Pro Arg Asp Glu Lys Pro Phe Cys Ala
 50 55 60

Thr Cys Arg Gly Leu Arg Arg Arg Cys Gln Arg Asp Ala Met Cys Cys
 65 70 75 80

Pro Gly Thr Leu Cys Val Asn Asp Val Cys Thr Thr Met Glu Asp Ala
 85 90 95

Thr Pro Ile Leu Glu Arg Gln Leu Asp Glu Gln Asp Gly Thr His Ala
 100 105 110

Glu Gly Thr Thr Gly His Pro Val Gln Glu Asn Gln Pro Lys Arg Lys
 115 120 125

Pro Ser Ile Lys Lys Ser Gln Gly Arg Lys Gly Gln Glu Gly Glu Ser
 130 135 140

Cys Leu Arg Thr Phe Asp Cys Gly Pro Gly Leu Cys Cys Ala Arg His
 145 150 155 160

Phe Trp Thr Lys Ile Cys Lys Pro Val Leu Leu Glu Gly Gln Val Cys
 165 170 175

Ser Arg Arg Gly His Lys Asp Thr Ala Gln Ala Pro Glu Ile Phe Gln
 180 185 190

Arg Cys Asp Cys Gly Pro Gly Leu Leu Cys Arg Ser Gln Leu Thr Ser
 195 200 205

Asn Arg Gln His Ala Arg Leu Arg Val Cys Gln Lys Ile Glu Lys Leu
 210 215 220

(2) INFORMAÇÃO PARA SEQ ID NO:6:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 672 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNC

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..672

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:6:

ATG GTG GCG GCC GTC CTG CTG GGG CTG AGC TGG CTC TGC TCT CCC CTG	48
Met Val Ala Ala Val Leu Leu Gly Leu Ser Trp Leu Cys Ser Pro Leu	
1 5 10 15	
GGA GCT CTG GTC CTG GAC TTC AAC AAC ATC AGG AGC TCT GCT GAC CTG	96
Gly Ala Leu Val Leu Asp Phe Asn Asn Ile Arg Ser Ser Ala Asp Leu	
20 25 30	
CAT GGG GCC CGG AAG GGC TCA CAG TGC CTG TCT GAC ACG GAC TGC AAT	144
His Gly Ala Arg Lys Gly Ser Gln Cys Leu Ser Asp Thr Asp Cys Asn	
35 40 45	
ACC AGA AAG TTC TGC CTC CAG CCC CGC GAT GAG AAG CCG TTC TGT GCT	192
Thr Arg Lys Phe Cys Leu Gln Pro Arg Asp Glu Lys Pro Phe Cys Ala	
50 55 60	
ACA TGT CGT GGG TTG CGG AGG AGG TGC CAG CGA GAT GCC ATG TGC TGC	240
Thr Cys Arg Gly Leu Arg Arg Cys Gln Arg Asp Ala Met Cys Cys	
65 70 75 80	
CCT GGG ACA CTC TGT GTG AAC GAT GTT TGT ACT ACG ATG GAA GAT GCA	288
Pro Gly Thr Leu Cys Val Asn Asp Val Cys Thr Thr Met Glu Asp Ala	
85 90 95	
ACC CCA ATA TTA GAA AGG CAG CTT GAT GAG CAA GAT GGC ACA CAT GCA	336
Thr Pro Ile Leu Glu Arg Gln Leu Asp Glu Gln Asp Gly Thr His Ala	
100 105 110	
GAA GGA ACA ACT GGG CAC CCA GTC CAG GAA AAC CAA CCC AAA AGG AAG	384
Glu Gly Thr Thr Gly His Pro Val Gln Glu Asn Gln Pro Lys Arg Lys	
115 120 125	
CCA AGT ATT AAG AAA TCA CAA GGC AGG AAG GGA CAA GAG GGA GAA AGT	432
Pro Ser Ile Lys Lys Ser Gln Gly Arg Lys Gly Gln Glu Gly Glu Ser	
130 135 140	
TGT CTG AGA ACT TTT GAC TGT GGC CCT GGA CTT TGC TGT GCT CGT CAT	480
Cys Leu Arg Thr Phe Asp Cys Gly Pro Gly Leu Cys Cys Ala Arg His	
145 150 155 160	
TTT TGG ACG AAA ATT TGT AAG CCA GTC CTT TTG GAG GGA CAG GTC TGC	528
Phe Trp Thr Lys Ile Cys Lys Pro Val Leu Leu Glu Gly Gln Val Cys	
165 170 175	
TCC AGA AGA GGG CAT AAA GAC ACT GCT CAA GCT CCA GAA ATC TTC CAG	576
Ser Arg Arg Gly His Lys Asp Thr Ala Gln Ala Pro Glu Ile Phe Gln	
180 185 190	
CGT TGC GAC TGT GGC CCT GGA CTA CTG TGT CGA AGC CAA TTG ACC AGC	624
Arg Cys Asp Cys Gly Pro Gly Leu Leu Cys Arg Ser Gln Leu Thr Ser	
195 200 205	
AAT CGG CAG CAT GCT CGA TTA AGA GTA TGC CAA AAA ATA GAA AAG CTA	672
Asn Arg Gln His Ala Arg Leu Arg Val Cys Gln Lys Ile Glu Lys Leu	
210 215 220	

(2) INFORMAÇÃO PARA SEQ ID NO:7:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1529 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNC

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 93..890

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:7:

```

CCGGACGCGT GGGCGGCACG GTTTCGTGGG GACCCAGGCT TGCAAAGTGA CGTTCATTTT    60
CTCTTTCTTT CTCCTCTTG AGTCCTTCTG AG  ATG ATG GCT CTG GGC GCA GCG    113
                                     Met Met Ala Leu Gly Ala Ala
                                       1           5

GGA GCT ACC CGG GTC TTT GTC GCG ATG GTA GCG GCG GCT CTC GGC GGC    161
Gly Ala Thr Arg Val Phe Val Ala Met Val Ala Ala Ala Leu Gly Gly
      10           15           20

CAC CCT CTG CTG GGA GTG AGC GCC ACC TTG AAC TCG GTT CTC AAT TCC    209
His Pro Leu Leu Gly Val Ser Ala Thr Leu Asn Ser Val Leu Asn Ser
      25           30           35

AAC GCT ATC AAG AAC CTG CCC CCA CCG CTG GGC GGC GCT GCG GGG CAC    257
Asn Ala Ile Lys Asn Leu Pro Pro Pro Leu Gly Gly Ala Ala Gly His
      40           45           50           55

CCA GGC TCT GCA GTC AGC GCC GCG CCG GGA ATC CTG TAC CCG GGC GGG    305
Pro Gly Ser Ala Val Ser Ala Ala Pro Gly Ile Leu Tyr Pro Gly Gly
      60           65           70

AAT AAG TAC CAG ACC ATT GAC AAC TAC CAG CCG TAC CCG TGC GCA GAG    353
Asn Lys Tyr Gln Thr Ile Asp Asn Tyr Gln Pro Tyr Pro Cys Ala Glu
      75           80           85

GAC GAG GAG TGC GGC ACT GAT GAG TAC TGC GCT AGT CCC ACC CGC GGA    401
Asp Glu Glu Cys Gly Thr Asp Glu Tyr Cys Ala Ser Pro Thr Arg Gly
      90           95           100

```

GGG GAC GCA GGC GTG CAA ATC TGT CTC GCC TGC AGG AAG CGC CGA AAA 449
 Gly Asp Ala Gly Val Gln Ile Cys Leu Ala Cys Arg Lys Arg Arg Lys
 105 110 115

CGC TGC ATG CGT CAC GCT ATG TGC TGC CCC GGG AAT TAC TGC AAA AAT 497
 Arg Cys Met Arg His Ala Met Cys Cys Pro Gly Asn Tyr Cys Lys Asn
 120 125 130 135

GGA ATA TGC GTG TCT TCT GAT CAA AAT CAT TTC CGA GGA GAA ATT GAG 545
 Gly Ile Cys Val Ser Ser Asp Gln Asn His Phe Arg Gly Glu Ile Glu
 140 145 150

GAA ACC ATC ACT GAA AGC TTT GGT AAT GAT CAT AGC ACC TTG GAT GGG 593
 Glu Thr Ile Thr Glu Ser Phe Gly Asn Asp His Ser Thr Leu Asp Gly
 155 160 165

TAT TCC AGA AGA ACC ACC TTG TCT TCA AAA ATG TAT CAC ACC AAA GGA 641
 Tyr Ser Arg Arg Thr Thr Leu Ser Ser Lys Met Tyr His Thr Lys Gly
 170 175 180

CAA GAA GGT TCT GTT TGT CTC CGG TCA TCA GAC TGT GCC TCA GGA TTG 689
 Gln Glu Gly Ser Val Cys Leu Arg Ser Ser Asp Cys Ala Ser Gly Leu
 185 190 195

TGT TGT GCT AGA CAC TTC TGG TCC AAG ATC TGT AAA CCT GTC CTG AAA 737
 Cys Cys Ala Arg His Phe Trp Ser Lys Ile Cys Lys Pro Val Leu Lys
 200 205 210 215

GAA GGT CAA GTG TGT ACC AAG CAT AGG AGA AAA GGC TCT CAT GGA CTA 785
 Glu Gly Gln Val Cys Thr Lys His Arg Arg Lys Gly Ser His Gly Leu
 220 225 230

GAA ATA TTC CAG CGT TGT TAC TGT GGA GAA GGT CTG TCT TGC CGG ATA 833
 Glu Ile Phe Gln Arg Cys Tyr Cys Gly Glu Gly Leu Ser Cys Arg Ile
 235 240 245

CAG AAA GAT CAC CAT CAA GCC AGT AAT TCT TCT AGG CTT CAC ACT TGT 881
 Gln Lys Asp His His Gln Ala Ser Asn Ser Ser Arg Leu His Thr Cys
 250 255 260

CAG AGA CAC TAAACCAGCT ATCCAAAATG CAGTGAAGCTC CTTTTATATA 930
 Gln Arg His
 265

ATAGATGCTA TGAAAACCTT TTATGACCTT CATCAACTCA ATCCTAAGGA TATACAAGTT 990

CTGTGGTTTC AGTTAAGCAT TCCAATAACA CCTTCCAAAA ACCTGGAGTG TAAGAGCTTT 1050

GTTTCTTTAT GGAACCTCCC TGTGATTGCA GTAAATTACT GTATTGTAAA TTCTCAGTGT 1110

GGCACTTACC TGTAATGCA ATGAACTTT TAATTATTTT TCTAAAGGTG CTGCACTGCC 1170

TATTTTCTCT CTGTATATGT AAATTTTGT ACACATTGAT TGTATCTTIG ACTGACAAAT 1230

ATTCTATATT GAACTGAAGT AAATCATTTT AGCTTATAGT TCTTAAAAGC ATAACCCTTT 1290

ACCCCATTTN ATTCTAGAGT CNAGAACGCA AGGATCTCTT GGAATGACAA ATGATAGGTA 1350

CCTAAAATGT AACATGAAAA TACTAGCTTA TTTTCTGAAA TGTACTATCT TAATGCTTAA 1410

ATTATATTTT CCTTTAGGCT GTGATAGTTT TTGAAATAAA ATTTAACATT TAATATCATG 1470

AAATGKTATA AGTAGACATA AAAAAAAAAA AAAAAAAAAA AGGGCGGCCG CTAGACTAG 1529

(2) INFORMAÇÃO PARA SEQ ID NO:8:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 266 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:8:

Met Met Ala Leu Gly Ala Ala Gly Ala Thr Arg Val Phe Val Ala Met
 1 5 10 15

Val Ala Ala Ala Leu Gly Gly His Pro Leu Leu Gly Val Ser Ala Thr
 20 25 30

Leu Asn Ser Val Leu Asn Ser Asn Ala Ile Lys Asn Leu Pro Pro Pro
 35 40 45

Leu Gly Gly Ala Ala Gly His Pro Gly Ser Ala Val Ser Ala Ala Pro
 50 55 60

Gly Ile Leu Tyr Pro Gly Gly Asn Lys Tyr Gln Thr Ile Asp Asn Tyr
 65 70 75 80

Gln Pro Tyr Pro Cys Ala Glu Asp Glu Glu Cys Gly Thr Asp Glu Tyr
 85 90 95

Cys Ala Ser Pro Thr Arg Gly Gly Asp Ala Gly Val Gln Ile Cys Leu
 100 105 110

Ala Cys Arg Lys Arg Arg Lys Arg Cys Met Arg His Ala Met Cys Cys
 115 120 125

Pro Gly Asn Tyr Cys Lys Asn Gly Ile Cys Val Ser Ser Asp Gln Asn
 130 135 140

His Phe Arg Gly Glu Ile Glu Glu Thr Ile Thr Glu Ser Phe Gly Asn
 145 150 155 160

Asp His Ser Thr Leu Asp Gly Tyr Ser Arg Arg Thr Thr Leu Ser Ser
 165 170 175

Lys Met Tyr His Thr Lys Gly Gln Glu Gly Ser Val Cys Leu Arg Ser
 180 185 190

Ser Asp Cys Ala Ser Gly Leu Cys Cys Ala Arg His Phe Trp Ser Lys
 195 200 205

Ile Cys Lys Pro Val Leu Lys Glu Gly Gln Val Cys Thr Lys His Arg
 210 215 220

Arg Lys Gly Ser His Gly Leu Glu Ile Phe Gln Arg Cys Tyr Cys Gly
 225 230 235 240

Glu Gly Leu Ser Cys Arg Ile Gln Lys Asp His His Gln Ala Ser Asn
 245 250 255

Ser Ser Arg Leu His Thr Cys Gln Arg His
 260 265

(2) INFORMAÇÃO PARA SEQ ID NO:9:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 798 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADnc

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..798

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:9:

ATG ATG GCT CTG GGC GCA GCG GGA GCT ACC CGG GTC TTT GTC GCG ATG	48
Met Met Ala Leu Gly Ala Ala Gly Ala Thr Arg Val Phe Val Ala Met	
1 5 10 15	
GTA GCG GCG GCT CTC GGC GGC CAC CCT CTG CTG GGA GTG AGC GCC ACC	96
Val Ala Ala Ala Leu Gly Gly His Pro Leu Leu Gly Val Ser Ala Thr	
20 25 30	
TTG AAC TCG GTT CTC AAT TCC AAC GCT ATC AAG AAC CTG CCC CCA CCG	144
Leu Asn Ser Val Leu Asn Ser Asn Ala Ile Lys Asn Leu Pro Pro Pro	
35 40 45	
CTG GGC GGC GCT GCG GGG CAC CCA GGC TCT GCA GTC AGC GCC GCG CCG	192
Leu Gly Gly Ala Ala Gly His Pro Gly Ser Ala Val Ser Ala Ala Pro	
50 55 60	
GGA ATC CTG TAC CCG GGC GGG AAT AAG TAC CAG ACC ATT GAC AAC TAC	240
Gly Ile Leu Tyr Pro Gly Gly Asn Lys Tyr Gln Thr Ile Asp Asn Tyr	
65 70 75 80	

CAG CCG TAC CCG TGC GCA GAG GAC GAG GAG TGC GGC ACT GAT GAG TAC Gln Pro Tyr Pro Cys Ala Glu Asp Glu Glu Cys Gly Thr Asp Glu Tyr 85 90 95	288
TGC GCT AGT CCC ACC CGC GGA GGG GAC GCA GGC GTG CAA ATC TGT CTC Cys Ala Ser Pro Thr Arg Gly Gly Asp Ala Gly Val Gln Ile Cys Leu 100 105 110	336
GCC TGC AGG AAG CGC CGA AAA CGC TGC ATG CGT CAC GCT ATG TGC TGC Ala Cys Arg Lys Arg Arg Lys Arg Cys Met Arg His Ala Met Cys Cys 115 120 125	384
CCC GGG AAT TAC TGC AAA AAT GGA ATA TGC GTG TCT TCT GAT CAA AAT Pro Gly Asn Tyr Cys Lys Asn Gly Ile Cys Val Ser Ser Asp Gln Asn 130 135 140	432
CAT TTC CGA GGA GAA ATT GAG GAA ACC ATC ACT GAA AGC TTT GGT AAT His Phe Arg Gly Glu Ile Glu Glu Thr Ile Thr Glu Ser Phe Gly Asn 145 150 155 160	480
GAT CAT AGC ACC TTG GAT GGG TAT TCC AGA AGA ACC ACC TTG TCT TCA Asp His Ser Thr Leu Asp Gly Tyr Ser Arg Arg Thr Thr Leu Ser Ser 165 170 175	528
AAA ATG TAT CAC ACC AAA GGA CAA GAA GGT TCT GTT TGT CTC CGG TCA Lys Met Tyr His Thr Lys Gly Gln Glu Gly Ser Val Cys Leu Arg Ser 180 185 190	576
TCA GAC TGT GCC TCA GGA TTG TGT TGT GCT AGA CAC TTC TGG TCC AAG Ser Asp Cys Ala Ser Gly Leu Cys Cys Ala Arg His Phe Trp Ser Lys 195 200 205	624
ATC TGT AAA CCT GTC CTG AAA GAA GGT CAA GTG TGT ACC AAG CAT AGG Ile Cys Lys Pro Val Leu Lys Glu Gly Gln Val Cys Thr Lys His Arg 210 215 220	672
AGA AAA GGC TCT CAT GGA CTA GAA ATA TTC CAG CGT TGT TAC TGT GGA Arg Lys Gly Ser His Gly Leu Glu Ile Phe Gln Arg Cys Tyr Cys Gly 225 230 235 240	720
GAA GGT CTG TCT TGC CGG ATA CAG AAA GAT CAC CAT CAA GCC AGT AAT Glu Gly Leu Ser Cys Arg Ile Gln Lys Asp His His Gln Ala Ser Asn 245 250 255	768
TCT TCT AGG CTT CAC ACT TGT CAG AGA CAC Ser Ser Arg Leu His Thr Cys Gln Arg His 260 265	798

(2) INFORMAÇÃO PARA SEQ ID NO:10:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:
 - (A) COMPRIMENTO: 702 pares de bases
 - (B) TIPO: ácido nucleico
 - (C) TIPO DE CADEIA: simples
 - (D) TOPOLOGIA: linear
- (ii) TIPO DE MOLÉCULA: ADNc
- (ix) CARACTERÍSTICA:
 - (A) NOME/CHAVE: CDS
 - (B) LOCALIZAÇÃO: 1..537

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:10:

GAA TTC GGC ACG AGG GTT GGG AGG TAT TGC CAC AGT CCC CAC CAA GGA	48
Glu Phe Gly Thr Arg Val Gly Arg Tyr Cys His Ser Pro His Gln Gly	
1 5 10 15	
TCA TCG GCC TGC ATG GTG TGT CGG AGA AAA AAG AAG CGC TGC CAC CGA	96
Ser Ser Ala Cys Met Val Cys Arg Arg Lys Lys Lys Arg Cys His Arg	
20 25 30	
GAT GGC ATG TGC TGC CCC AGT ACC CGC TGC AAT AAT GGC ATC TGT ATC	144
Asp Gly Met Cys Cys Pro Ser Thr Arg Cys Asn Asn Gly Ile Cys Ile	
35 40 45	
CCA GTT ACT GAA AGC ATC TTA ACC CCT CAC ATC CCG GCT CTG GAT GGT	192
Pro Val Thr Glu Ser Ile Leu Thr Pro His Ile Pro Ala Leu Asp Gly	
50 55 60	
ACT CGG CAC AGA GAT CGA AAC CAC GGT CAT TAC TCA AAC CAT GAC TTG	240
Thr Arg His Arg Asp Arg Asn His Gly His Tyr Ser Asn His Asp Leu	
65 70 75 80	
GGA TGG CAG AAT CTA GGA AGA CCA CAC ACT AAG ATG TCA CAT ATA AAA	288
Gly Trp Gln Asn Leu Gly Arg Pro His Thr Lys Met Ser His Ile Lys	
85 90 95	
GGG CAT GAA GGA GAC CCC TGC CTA CGA TCA TCA GAC TGC ATT GAA GGG	336
Gly His Glu Gly Asp Pro Cys Leu Arg Ser Ser Asp Cys Ile Glu Gly	
100 105 110	
TTT TGC TGT GCT CGT CAT TTC TGG ACC AAA ATC TGC AAA CCA GTC CTC	384
Phe Cys Cys Ala Arg His Phe Trp Thr Lys Ile Cys Lys Pro Val Leu	
115 120 125	
CAT CAG GGG GAA GTC TGT ACC AAA CAA CGC AAG AAG GGT TCT CAT GGG	432
His Gln Gly Glu Val Cys Thr Lys Gln Arg Lys Lys Gly Ser His Gly	
130 135 140	
CTG GAA ATT TTC CAG CGT TGC GAC TGT GCG AAG GGC CTG TCT TGC AAA	480
Leu Glu Ile Phe Gln Arg Cys Asp Cys Ala Lys Gly Leu Ser Cys Lys	
145 150 155 160	
GTA TGG AAA GAT GCC ACC TAC TCC TCC AAA GCC AGA CTC CAT GTG TGT	528
Val Trp Lys Asp Ala Thr Tyr Ser Ser Lys Ala Arg Leu His Val Cys	
165 170 175	
CAG AAA ATT TGATCACCAT TGAGGAACAT CATCAATTGC AGACTGTGAA	577
Gln Lys Ile	
GTTGTGTATT TAATGCATTA TAGCATGGTG GAAAATAAGG TTCAGATGCA GAAGAATGGC	637
TAAAATAAGA AACGTGATAA GAATATAGAT GATCACAAAA AAAAAAAAAA AAAAGATGCC	697
GCCGC	702

(2) INFORMAÇÃO PARA SEQ ID NO:11:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 179 aminoácidos
- (B) TIPO: aminoácido
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:11:

Glu Phe Gly Thr Arg Val Gly Arg Tyr Cys His Ser Pro His Gln Gly
 1 5 10 15

Ser Ser Ala Cys Met Val Cys Arg Arg Lys Lys Lys Arg Cys His Arg
 20 25 30

Asp Gly Met Cys Cys Pro Ser Thr Arg Cys Asn Asn Gly Ile Cys Ile
 35 40 45

Pro Val Thr Glu Ser Ile Leu Thr Pro His Ile Pro Ala Leu Asp Gly
 50 55 60

Thr Arg His Arg Asp Arg Asn His Gly His Tyr Ser Asn His Asp Leu
 65 70 75 80

Gly Trp Gln Asn Leu Gly Arg Pro His Thr Lys Met Ser His Ile Lys
 85 90 95

Gly His Glu Gly Asp Pro Cys Leu Arg Ser Ser Asp Cys Ile Glu Gly
 100 105 110

Phe Cys Cys Ala Arg His Phe Trp Thr Lys Ile Cys Lys Pro Val Leu
 115 120 125

His Gln Gly Glu Val Cys Thr Lys Gln Arg Lys Lys Gly Ser His Gly
 130 135 140

Leu Glu Ile Phe Gln Arg Cys Asp Cys Ala Lys Gly Leu Ser Cys Lys
 145 150 155 160

Val Trp Lys Asp Ala Thr Tyr Ser Ser Lys Ala Arg Leu His Val Cys
 165 170 175

Gln Lys Ile

(2) INFORMAÇÃO PARA SEQ ID NO:12:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 537 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..537

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:12:

GAA TTC GGC ACG AGG GTT GGG AGG TAT TGC CAC AGT CCC CAC CAA GGA	48
Glu Phe Gly Thr Arg Val Gly Arg Tyr Cys His Ser Pro His Gln Gly	
1 5 10 15	
TCA TCG GCC TGC ATG GTG TGT CGG AGA AAA AAG AAG CGC TGC CAC CGA	96
Ser Ser Ala Cys Met Val Cys Arg Arg Lys Lys Lys Arg Cys His Arg	
20 25 30	
GAT GGC ATG TGC TGC CCC AGT ACC CGC TGC AAT AAT GGC ATC TGT ATC	144
Asp Gly Met Cys Cys Pro Ser Thr Arg Cys Asn Asn Gly Ile Cys Ile	
35 40 45	
CCA GTT ACT GAA AGC ATC TTA ACC CCT CAC ATC CCG GCT CTG GAT GGT	192
Pro Val Thr Glu Ser Ile Leu Thr Pro His Ile Pro Ala Leu Asp Gly	
50 55 60	
ACT CGG CAC AGA GAT CGA AAC CAC GGT CAT TAC TCA AAC CAT GAC TTG	240
Thr Arg His Arg Asp Arg Asn His Gly His Tyr Ser Asn His Asp Leu	
65 70 75 80	
GGA TGG CAG AAT CTA GGA AGA CCA CAC ACT AAG ATG TCA CAT ATA AAA	288
Gly Trp Gln Asn Leu Gly Arg Pro His Thr Lys Met Ser His Ile Lys	
85 90 95	
GGG CAT GAA GGA GAC CCC TGC CTA CGA TCA TCA GAC TGC ATT GAA GGG	336
Gly His Glu Gly Asp Pro Cys Leu Arg Ser Ser Asp Cys Ile Glu Gly	
100 105 110	
TTT TGC TGT GCT CGT CAT TTC TGG ACC AAA ATC TGC AAA CCA GTG CTC	384
Phe Cys Cys Ala Arg His Phe Trp Thr Lys Ile Cys Lys Pro Val Leu	
115 120 125	
CAT CAG GGG GAA GTC TGT ACC AAA CAA CGC AAG AAG GGT TCT CAT GGG	432
His Gln Gly Glu Val Cys Thr Lys Gln Arg Lys Lys Gly Ser His Gly	
130 135 140	
CTG GAA ATT TTC CAG CGT TGC GAC TGT GCG AAG GGC CTG TCT TGC AAA	480
Leu Glu Ile Phe Gln Arg Cys Asp Cys Ala Lys Gly Leu Ser Cys Lys	
145 150 155 160	
GTA TGG AAA GAT GCC ACC TAC TCC TCC AAA GCC AGA CTC CAT GTG TGT	528
Val Trp Lys Asp Ala Thr Tyr Ser Ser Lys Ala Arg Leu His Val Cys	
165 170 175	
CAG AAA ATT	537
Gln Lys Ile	

(2) INFORMAÇÃO PARA SEQ ID NO:13:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 928 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 75..800

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:13:

CTCGAGGCCA AAATTCGGCA CGAGGCCGGG CTGTGGTCTA GCATAAAGGC GGAGCCCAGA	60
AGAAGGGGCG GGGT ATG GGA GAA GCC TCC CCA CCT GCC CCC GCA AGG CGG	110
Met Gly Glu Ala Ser Pro Pro Ala Pro Ala Arg Arg	
1 5 10	
CAT CTG CTG GTC CTG CTG CTG CTC CTC TCT ACC CTG GTG ATC CCC TCC	158
His Leu Leu Val Leu Leu Leu Leu Leu Ser Thr Leu Val Ile Pro Ser	
15 20 25	
GCT GCA GCT CCT ATC CAT GAT GCT GAC GCC CAA GAG AGC TCC TTG GGT	206
Ala Ala Ala Pro Ile His Asp Ala Asp Ala Gln Glu Ser Ser Leu Gly	
30 35 40	
CTC ACA GGC CTC CAG AGC CTA CTC CAA GGC TTC AGC CGA CTT TTC CTG	254
Leu Thr Gly Leu Gln Ser Leu Leu Gln Gly Phe Ser Arg Leu Phe Leu	
45 50 55 60	
AAA GGT AAC CTG CTT CGG GGC ATA GAC AGC TTA TTC TCT GCC CCC ATG	302
Lys Gly Asn Leu Leu Arg Gly Ile Asp Ser Leu Phe Ser Ala Pro Met	
65 70 75	
GAC TTC CGG GGC CTC CCT GGG AAC TAC CAC AAA GAG GAG AAC CAG GAG	350
Asp Phe Arg Gly Leu Pro Gly Asn Tyr His Lys Glu Glu Asn Gln Glu	
80 85 90	

CAC CAG CTG GGG AAC AAC ACC CTC TCC AGC CAC CTC CAG ATC GAC AAG	398
His Gln Leu Gly Asn Asn Thr Leu Ser Ser His Leu Gln Ile Asp Lys	
95 100 105	
ATG ACC GAC AAC AAG ACA GGA GAG GTG CTG ATC TCC GAG AAT GTG GTG	446
Met Thr Asp Asn Lys Thr Gly Glu Val Leu Ile Ser Glu Asn Val Val	
110 115 120	
GCA TCC ATT CAA CCA GCG GAG GGG AGC TTC GAG GGT GAT TTG AAG GTA	494
Ala Ser Ile Gln Pro Ala Glu Gly Ser Phe Glu Gly Asp Leu Lys Val	
125 130 135 140	
CCC AGG ATG GAG GAG AAG GAG GCC CTG GTA CCC ATC CAG AAG GCC ACG	542
Pro Arg Met Glu Glu Lys Glu Ala Leu Val Pro Ile Gln Lys Ala Thr	
145 150 155	
GAC AGC TTC CAC ACA GAA CTC CAT CCC CGG GTG GCC TTC TGG ATC ATT	590
Asp Ser Phe His Thr Glu Leu His Pro Arg Val Ala Phe Trp Ile Ile	
160 165 170	
AAG CTG CCA CGG CGG AGG TCC CAC CAG GAT GCC CTG GAG GGC GGC CAC	638
Lys Leu Pro Arg Arg Arg Ser His Gln Asp Ala Leu Glu Gly Gly His	
175 180 185	
TGG CTC AGC GAG AAG CGA CAC CGC CTG CAG GCC ATC CGG GAT GGA CTC	686
Trp Leu Ser Glu Lys Arg His Arg Leu Gln Ala Ile Arg Asp Gly Leu	
190 195 200	
CGC AAG GGG ACC CAC AAG GAC GTC CTA GAA GAG GGG ACC GAG AGC TCC	734
Arg Lys Gly Thr His Lys Asp Val Leu Glu Glu Gly Thr Glu Ser Ser	
205 210 215 220	
TCC CAC TCC AGG CTG TCC CCC CGA AAG ACC CAC TTA CTG TAC ATC CTC	782
Ser His Ser Arg Leu Ser Pro Arg Lys Thr His Leu Leu Tyr Ile Leu	
225 230 235	
AGG CCC TCT CGG CAG CTG TAGGGGTGGG GACCGGGGAG CACCTGCCTG	830
Arg Pro Ser Arg Gln Leu	
240	
TAGCCCCCAT CAGACCCTGC CCCAAGCACC ATATGGAAAT AAAGTTCTTT CTTACATCTA	890
AAAAAAAAAA AAAAAAAAAA AAAAAAATTG GCGGCCGC	928

(2) INFORMAÇÃO PARA SEQ ID NO:14:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 242 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:14:

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

(2) INFORMAÇÃO PARA SEQ ID NO:15:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 726 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNC

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..726

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:15:

ATG GGA GAA GCC TCC CCA CCT GCC CCC GCA AGG CGG CAT CTG CTG GTC	48
Met Gly Glu Ala Ser Pro Pro Ala Pro Ala Arg Arg His Leu Leu Val	
1 5 10 15	
CTG CTG CTG CTC CTC TCT ACC CTG GTG ATC CCC TCC GCT GCA GCT CCT	96
Leu Leu Leu Leu Leu Ser Thr Leu Val Ile Pro Ser Ala Ala Ala Pro	
20 25 30	
ATC CAT GAT GCT GAC GCC CAA GAG AGC TCC TTG GGT CTC ACA GGC CTC	144
Ile His Asp Ala Asp Ala Gln Glu Ser Ser Leu Gly Leu Thr Gly Leu	
35 40 45	
CAG AGC CTA CTC CAA GGC TTC AGC CGA CTT TTC CTG AAA GGT AAC CTG	192
Gln Ser Leu Leu Gln Gly Phe Ser Arg Leu Phe Leu Lys Gly Asn Leu	
50 55 60	
CTT CGG GGC ATA GAC AGC TTA TTC TCT GCC CCC ATG GAC TTC CGG GGC	240
Leu Arg Gly Ile Asp Ser Leu Phe Ser Ala Pro Met Asp Phe Arg Gly	
65 70 75 80	
CTC CCT GGG AAC TAC CAC AAA GAG GAG AAC CAG GAG CAC CAG CTG GGG	288
Leu Pro Gly Asn Tyr His Lys Glu Glu Asn Gln Glu His Gln Leu Gly	
85 90 95	
AAC AAC ACC CTC TCC AGC CAC CTC CAG ATC GAC AAG ATG ACC GAC AAC	336
Asn Asn Thr Leu Ser Ser His Leu Gln Ile Asp Lys Met Thr Asp Asn	
100 105 110	
AAG ACA GGA GAG GTG CTG ATC TCC GAG AAT GTG GTG GCA TCC ATT CAA	384
Lys Thr Gly Glu Val Leu Ile Ser Glu Asn Val Val Ala Ser Ile Gln	
115 120 125	
CCA GCG GAG GGG AGC TTC GAG GGT GAT TTG AAG GTA CCC AGG ATG GAG	432
Pro Ala Glu Gly Ser Phe Glu Gly Asp Leu Lys Val Pro Arg Met Glu	
130 135 140	
GAG AAG GAG GCC CTG GTA CCC ATC CAG AAG GCC ACG GAC AGC TTC CAC	480
Glu Lys Glu Ala Leu Val Pro Ile Gln Lys Ala Thr Asp Ser Phe His	
145 150 155 160	
ACA GAA CTC CAT CCC CGG GTG GCC TTC TGG ATC ATT AAG CTG CCA CGG	528
Thr Glu Leu His Pro Arg Val Ala Phe Trp Ile Ile Lys Leu Pro Arg	
165 170 175	
CGG AGG TCC CAC CAG GAT GCC CTG GAG GGC GGC CAC TGG CTC AGC GAG	576
Arg Arg Ser His Gln Asp Ala Leu Glu Gly Gly His Trp Leu Ser Glu	
180 185 190	
AAG CGA CAC CGC CTG CAG GCC ATC CGG GAT GGA CTC CGC AAG GGG ACC	624
Lys Arg His Arg Leu Gln Ala Ile Arg Asp Gly Leu Arg Lys Gly Thr	
195 200 205	
CAC AAG GAC GTC CTA GAA GAG GGG ACC GAG AGC TCC TCC CAC TCC AGG	672
His Lys Asp Val Leu Glu Glu Gly Thr Glu Ser Ser Ser His Ser Arg	
210 215 220	
CTG TCC CCC CGA AAG ACC CAC TTA CTG TAC ATC CTC AGG CCC TCT CGG	720
Leu Ser Pro Arg Lys Thr His Leu Leu Tyr Ile Leu Arg Pro Ser Arg	
225 230 235 240	
CAG CTG	726
Gln Leu	

(2) INFORMAÇÃO PARA SEQ ID NO:16:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 2381 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADnc

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 110..1156

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:16:

```

FGTCGACCCA CGCGTCCGCT GTGGCAGCCC AGCTACCGGT CGTGACCAGA TCCAGCTTGC      60
AGCTCAGCTT TGTTCATTTC AATTGGGCGG CGGCCAGCGC GGAACAAAC ATG CAG      115
                                     Met Gln
                                     1
CGG CTC GGG GGT ATT TTG CTG TGT ACA CTG CTG GCG GCG GCG GTC CCC      163
Arg Leu Gly Gly Ile Leu Leu Cys Thr Leu Leu Ala Ala Ala Val Pro
      5                10                15
ACT GCT CCT GCT CCT TCC CCG ACG GTC ACT TGG ACT CCG GCG GAG CCG      211
Thr Ala Pro Ala Pro Ser Pro Thr Val Thr Trp Thr Pro Ala Glu Pro
      20                25                30
GGC CCA GCT CTC AAC TAC CCT CAG GAG GAA GCT ACG CTC AAT GAG ATG      259
Gly Pro Ala Leu Asn Tyr Pro Gln Glu Glu Ala Thr Leu Asn Glu Met
      35                40                45                50

```

TTT CGA GAG GTG GAG GAG CTG ATG GAA GAC ACT CAG CAC AAA CTG CGC	307
Phe Arg Glu Val Glu Leu Met Glu Asp Thr Gln His Lys Leu Arg	
55 60 65	
AGT GCC GTG GAG GAG ATG GAG GCG GAA GAA GCA GCT GCT AAA ACG TCC	355
Ser Ala Val Glu Glu Met Glu Ala Glu Glu Ala Ala Lys Thr Ser	
70 75 80	
TCT GAG GTG AAC CTG GCA AGC TTA CCT CCC AAC TAT CAC AAT GAG ACC	403
Ser Glu Val Asn Leu Ala Ser Leu Pro Pro Asn Tyr His Asn Glu Thr	
85 90 95	
AGC ACG GAG ACC AGG GTG GGA AAT AAC ACA GTC CAT GTG CAC CAG GAA	451
Ser Thr Glu Thr Arg Val Gly Asn Asn Thr Val His Val His Gln Glu	
100 105 110	
GTT CAC AAG ATA ACC AAC AAC CAG AGT GGA CAG GTG GTC TTT TCT GAG	499
Val His Lys Ile Thr Asn Asn Gln Ser Gly Gln Val Val Phe Ser Glu	
115 120 125 130	
ACA GTC ATT ACA TCT GTA GGG GAT GAA GAA GGC AAG AGG AGC CAT GAA	547
Thr Val Ile Thr Ser Val Gly Asp Glu Glu Gly Lys Arg Ser His Glu	
135 140 145	
TGT ATC ATT GAT GAA GAC TGT GGG CCC ACC AGG TAC TGC CAG TTC TCC	595
Cys Ile Ile Asp Glu Asp Cys Gly Pro Thr Arg Tyr Cys Gln Phe Ser	
150 155 160	
AGC TTC AAG TAC ACC TGC CAG CCA TGC CGG GAC CAG CAG ATG CTA TGC	643
Ser Phe Lys Tyr Thr Cys Gln Pro Cys Arg Asp Gln Gln Met Leu Cys	
165 170 175	
ACC CGA GAC AGT GAG TGC TGT GGA GAC CAG CTG TGT GCC TGG GGT CAC	691
Thr Arg Asp Ser Glu Cys Cys Gly Asp Gln Leu Cys Ala Trp Gly His	
180 185 190	
TGC ACC CAA AAG GCC ACC AAA GGT GGC AAT GGG ACC ATC TGT GAC AAC	739
Cys Thr Gln Lys Ala Thr Lys Gly Gly Asn Gly Thr Ile Cys Asp Asn	
195 200 205 210	
CAG AGG GAT TGC CAG CCT GGC CTG TGT TGT GCC TTC CAA AGA GGC CTG	787
Gln Arg Asp Cys Gln Pro Gly Leu Cys Cys Ala Phe Gln Arg Gly Leu	
215 220 225	
CTG TTC CCC GTG TGC ACA CCC CTG CCC GTG GAG GGA GAG CTC TGC CAT	835
Leu Phe Pro Val Cys Thr Pro Leu Pro Val Glu Gly Glu Leu Cys His	
230 235 240	
GAC CCC ACC AGC CAG CTG CTG GAT CTC ATC ACC TGG GAA CTG GAG CCT	883
Asp Pro Thr Ser Gln Leu Leu Asp Leu Ile Thr Trp Glu Leu Glu Pro	
245 250 255	
GAA GGA GCT TTG GAC CGA TGC CCC TGC GCC AGT GGC CTC CTA TGC CAG	931
Glu Gly Ala Leu Asp Arg Cys Pro Cys Ala Ser Gly Leu Leu Cys Gln	
260 265 270	

CCA CAC AGC CAC AGT CTG GTG TAC ATG TGC AAG CCA GCC TTC GTG GGC 979
 Pro His Ser His Ser Leu Val Tyr Met Cys Lys Pro Ala Phe Val Gly
 275 280 285 290

AGC CAT GAC CAC AGT GAG GAG AGC CAG CTG CCC AGG GAG GCC CCG GAT 1027
 Ser His Asp His Ser Glu Glu Ser Gln Leu Pro Arg Glu Ala Pro Asp
 295 300 305

GAG TAC GAA GAT GTT GGC TTC ATA GGG GAA GTG CGC CAG GAG CTG GAA 1075
 Glu Tyr Glu Asp Val Gly Phe Ile Gly Glu Val Arg Gln Glu Leu Glu
 310 315 320

GAC CTG GAG CGG AGC CTA GCC CAG GAG ATG GCA TTT GAG GGG CCT GCC 1123
 Asp Leu Glu Arg Ser Leu Ala Gln Glu Met Ala Phe Glu Gly Pro Ala
 325 330 335

CCT GTG GAG TCA CTA GGC GGA GAG GAG GAG ATT TAGGCCCAGA CCCAGCTGAG 1176
 Pro Val Glu Ser Leu Gly Gly Glu Glu Glu Ile
 340 345

TCACTGGTAG ATGTGCAATA GAAATGGCTA ATTTATTTTC CCAGGAGTGT CCCCAAGTGT 1236

GGAATGGCCG CAGCTCCTTC CCAGTAGCTT TTCCTCTGGC TTGACAAGGT ACAGTGCAGT 1296

ACATTTCTTC CAGCCGCCCT GCTTCTCTGA CTTGGGAAAG ACAGGCATGG CGGGTAAGGG 1356

CAGCGGTGAG TCGTCCCTCG CTGTTGCTAG AAACGCTGTC TTGTCTTCA TGGATGGAAG 1416

ATTTGTTTGA AGGGAGAGGA TGGGAAGGGG TGAAGTCTGC TCATGATGGA TTTGGGGGAT 1476

ACAGGGAGGA GGATGCCTGC CTGCGAGACG TGGACTTGGC AAAATGTAAC CTTTGCTTTT 1536

GTCTTGCGCC GCTCCCATGG GCTGAGGCAG TGGCTACACA AGAGCTATGC TGCTCTGTGG 1596

CCTCCACAT ATTATCCCT GTGTTTCAGC TCCTACCTCA CTGTCAGCAC AGCCCTTCAT 1656

AGCCACGCCC CCTCTGCTC ACCACAGCCT AGGAGGGGAC CAGAGGGGAC TTCTCTCAGA 1716

GCCCCATGCT CTCTCTCTCA ACCCCATACC AGCCTCTGTG CCAGCGACAG TCCTTCCAAA 1776

TGGAGGGAGT GAAATCCTTT GGTTTAATTA TTTTCTCCTT CAAGGCACGC CTGCCACTAA 1836

GGTCAGGCTG ACTTGCATGT CCCTCTAACG TTCGTAGCAG TGTGGTGGAC ACTGTCTTCC 1896

ACCGACTGCT TCAATACCTC TGAAAGCCAG TGCTCGGAGT GCAGTTCGTG TAAATTAATT 1956

TGCAGGAAGT ATACTTGGCT AATTGTAGGG CTAGGATTGT GAATGAAATT TGCAAAGTCC 2016

CTTAGCAACA ATGGAAAGCC TTTCTCAGTC ACACCGAGAA GTCACAACCA AGCCAGGTTG 2076

TGTAGAGTAC AGCTGTGACA TACAGACAGA AGAAGGCTGG GCTGGATGTC AGGCCTCAGA 2136

TGACGGTTTC AGGTGCCAGG AACTATTACC ATTCTGTATC TATCCAGAGT TATTAATAAT 2196

GAAAGTTGCA CACATTTGTA TAAGCATGCC TTTCTCCTGA GTTTTAAATT ATATGTATAC 2256

ACAAACATGT GGCCTCAAA GATCATGCAC AAACCACTAC TCTTTGCTAA TTCTTGACT 2316

TTTCTCTTTG ATTTTCAATA AATACAAATC CCCTTCATGC AAAAAAAAAA AAAAAGGGCG 2376

GCCGC 2381

(2) INFORMAÇÃO PARA SEQ ID NO:17:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 349 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:17:

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

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

(2) INFORMAÇÃO PARA SEQ ID NO:18:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1047 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNC

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1047

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO:18:

ATG CAG CGG CTC GGG GGT ATT TTG CTG TGT ACA CTG CTG GCG GCG GCG	48
Met Gln Arg Leu Gly Gly Ile Leu Leu Cys Thr Leu Leu Ala Ala Ala	
1 5 10 15	
GTC CCC ACT GCT CCT GCT CCT TCC CCG ACG GTC ACT TGG ACT CCG GCG	96
Val Pro Thr Ala Pro Ala Pro Ser Pro Thr Val Thr Trp Thr Pro Ala	
20 25 30	

GAG CCG GGC CCA GCT CTC AAC TAC CCT CAG GAG GAA GCT ACG CTC AAT	144
Glu Pro Gly Pro Ala Leu Asn Tyr Pro Gln Glu Glu Ala Thr Leu Asn	
35 40 45	
GAG ATG TTT CGA GAG GTG GAG GAG CTG ATG GAA GAC ACT CAG CAC AAA	192
Glu Met Phe Arg Glu Val Glu Glu Leu Met Glu Asp Thr Gln His Lys	
50 55 60	
CTG CGC AGT GCC GTG GAG GAG ATG GAG GCG GAA GAA GCA GCT GCT AAA	240
Leu Arg Ser Ala Val Glu Glu Met Glu Ala Glu Glu Ala Ala Ala Lys	
65 70 75 80	
ACG TCC TCT GAG GTG AAC CTG GCA AGC TTA CCT CCC AAC TAT CAC AAT	288
Thr Ser Ser Glu Val Asn Leu Ala Ser Leu Pro Pro Asn Tyr His Asn	
85 90 95	
GAG ACC AGC ACG GAG ACC AGG GTG GGA AAT AAC ACA GTC CAT GTG CAC	336
Glu Thr Ser Thr Glu Thr Arg Val Gly Asn Asn Thr Val His Val His	
100 105 110	
CAG GAA GTT CAC AAG ATA ACC AAC AAC CAG AGT GGA CAG GTG GTC TTT	384
Gln Glu Val His Lys Ile Thr Asn Asn Gln Ser Gly Gln Val Val Phe	
115 120 125	
TCT GAG ACA GTC ATT ACA TCT GTA GGG GAT GAA GAA GGC AAG AGG AGC	432
Ser Glu Thr Val Ile Thr Ser Val Gly Asp Glu Glu Gly Lys Arg Ser	
130 135 140	
CAT GAA TGT ATC ATT GAT GAA GAC TGT GGG CCC ACC AGG TAC TGC CAG	480
His Glu Cys Ile Ile Asp Glu Asp Cys Gly Pro Thr Arg Tyr Cys Gln	
145 150 155 160	
TTC TCC AGC TTC AAG TAC ACC TGC CAG CCA TGC CGG GAC CAG CAG ATG	528
Phe Ser Ser Phe Lys Tyr Thr Cys Gln Pro Cys Arg Asp Gln Gln Met	
165 170 175	
CTA TGC ACC CGA GAC AGT GAG TGC TGT GGA GAC CAG CTG TGT GCC TGG	576
Leu Cys Thr Arg Asp Ser Glu Cys Cys Gly Asp Gln Leu Cys Ala Trp	
180 185 190	
GGT CAC TGC ACC CAA AAG GCC ACC AAA GGT GGC AAT GGG ACC ATC TGT	624
Gly His Cys Thr Gln Lys Ala Thr Lys Gly Gly Asn Gly Thr Ile Cys	
195 200 205	
GAC AAC CAG AGG GAT TGC CAG CCT GGC CTG TGT TGT GCC TTC CAA AGA	672
Asp Asn Gln Arg Asp Cys Gln Pro Gly Leu Cys Cys Ala Phe Gln Arg	
210 215 220	
GGC CTG CTG TTC CCC GTG TGC ACA CCC CTG CCC GTG GAG GGA GAG CTC	720
Gly Leu Leu Phe Pro Val Cys Thr Pro Leu Pro Val Glu Gly Glu Leu	
225 230 235 240	
TGC CAT GAC CCC ACC AGC CAG CTG CTG GAT CTC ATC ACC TGG GAA CTG	768
Cys His Asp Pro Thr Ser Gln Leu Leu Asp Leu Ile Thr Trp Glu Leu	
245 250 255	

GAG CCT GAA GGA GCT TTG GAC CGA TGC CCC TGC GCC AGT GGC CTC CTA	816
Glu Pro Glu Gly Ala Leu Asp Arg Cys Pro Cys Ala Ser Gly Leu Leu	
260 265 270	
TGC CAG CCA CAC AGC CAC AGT CTG GTG TAC ATG TGC AAG CCA GCC TTC	864
Cys Gln Pro His Ser His Ser Leu Val Tyr Met Cys Lys Pro Ala Phe	
275 280 285	
GTG GGC AGC CAT GAC CAC AGT GAG GAG AGC CAG CTG CCC AGG GAG GCC	912
Val Gly Ser His Asp His Ser Glu Glu Ser Gln Leu Pro Arg Glu Ala	
290 295 300	
CCG GAT GAG TAC GAA GAT GTT GGC TTC ATA GGG GAA GTG CGC CAG GAG	960
Pro Asp Glu Tyr Glu Asp Val Gly Phe Ile Gly Glu Val Arg Gln Glu	
305 310 315 320	
CTG GAA GAC CTG GAG CGG AGC CTA GCC CAG GAG ATG GCA TTT GAG GGG	1008
Leu Glu Asp Leu Glu Arg Ser Leu Ala Gln Glu Met Ala Phe Glu Gly	
325 330 335	
CCT GCC CCT GTG GAG TCA CTA GGC GGA GAG GAG GAG ATT	1047
Pro Ala Pro Val Glu Ser Leu Gly Gly Glu Glu Glu Ile	
340 345	

Lisboa,

REIVINDICAÇÕES

1. Molécula de ácido nucleico isolada seleccionada do grupo que consiste em:

- a) um ácido nucleico compreendendo SEQ ID NO:3 ou SEQ ID NO:9;
- b) um ácido nucleico que codifica um polipéptido compreendendo a sequência de aminoácidos dos resíduos de aminoácido 24-350 de SEQ ID NO:2, ou a sequência de aminoácidos dos resíduos de aminoácido 21-266 de SEQ ID NO:8;
- c) um ácido nucleico que codifica um polipéptido que compreende uma variante alélica de ocorrência natural da sequência de aminoácidos dos resíduos de aminoácido 20-224 de SEQ ID NO:5, SEQ ID NO:11 ou SEQ ID NO:14, em que a referida variante alélica de ocorrência natural modula a transdução de sinal celular ou a proliferação celular, e em que (i) o referido ácido nucleico que codifica uma variante alélica de ocorrência natural da sequência de aminoácidos dos resíduos de aminoácido 20-224 de SEQ ID NO:5 hibrida com o complemento de SEQ ID NO:6 sob condições de lavagem rigorosas, (ii) o referido ácido nucleico que codifica uma variante alélica de ocorrência natural da sequência de aminoácidos de SEQ ID NO:11 hibrida com o complemento de SEQ ID NO:12 sob condições de lavagem rigorosas, (iii) o referido ácido nucleico que codifica uma variante alélica de ocorrência natural da sequência de aminoácidos de SEQ ID NO:14 hibrida com o complemento de SEQ ID NO:15 sob condições de lavagem rigorosas, em que as condições de lavagem rigorosas são SSC 0,2X, SDS a 0,1%, a 50-65°C;
- d) um ácido nucleico compreendendo uma sequência nucleotídica que é o complemento de a), b) ou c); e
- e) um vector de expressão de células eucariotas recombinante compreendendo o ácido nucleico de a), b), c) ou d).

2. Molécula de ácido nucleico isolada de acordo com a reivindicação 1, em que a referida molécula de ácido nucleico isolada compreende uma sequência nucleotídica seleccionada do grupo que consiste em SEQ ID NO:1 e SEQ ID NO:3.

3. Molécula de ácido nucleico isolada de acordo com a reivindicação 1, em que a referida molécula de ácido nucleico isolada compreende uma sequência nucleotídica seleccionada do grupo que consiste em SEQ ID NO:7, SEQ ID NO:9 e a sequência nucleotídica da inserção de ADN do plasmídeo depositado na ATCC com o Número de Acesso 98633.

4. Molécula de ácido nucleico isolada de acordo com a reivindicação 1, em que a referida molécula de ácido nucleico isolada codifica um polipéptido compreendendo uma sequência de aminoácidos seleccionada do grupo que consiste em SEQ ID NO:2, nos resíduos de aminoácido 20-350 de SEQ ID NO:2, nos resíduos de aminoácido 22-350 de SEQ ID NO:2 e nos resíduos de aminoácido 24-350 de SEQ ID NO:2.

5. Molécula de ácido nucleico isolada de acordo com a reivindicação 1, em que a referida molécula de ácido nucleico isolada codifica um polipéptido compreendendo uma sequência de aminoácidos seleccionada do grupo que consiste em SEQ ID NO:8 e nos resíduos de aminoácido 21-266 de SEQ ID NO:8.

6. Molécula de ácido nucleico isolada de acordo com a reivindicação 1a), 1b) ou 1c), em que a referida molécula de ácido nucleico isolada compreende ainda:

- i) uma sequência de ácido nucleico vector; e/ou
- ii) uma sequência de ácido nucleico que codifica um polipéptido heterólogo.

7. Molécula de ácido nucleico isolada de acordo com a reivindicação 1e), em que a referida molécula de ácido nucleico isolada compreende ainda uma sequência de ácido nucleico que codifica um polipéptido heterólogo.

8. Célula hospedeira compreendendo o vector de expressão de células eucariotas recombinante de acordo com a reivindicação 1e).

9. Célula hospedeira compreendendo um vector compreendendo o ácido nucleico de acordo com qualquer uma das reivindicações 1a), 1b), 2, 3, 4 ou 5 em que o referido ácido nucleico é um ácido nucleico recombinante.

10. Polipéptido isolado seleccionado do grupo que consiste em:

- a) um polipéptido compreendendo pelo menos 15 aminoácidos contíguos de SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11 ou SEQ ID NO:14, polipéptido este que modula a transdução de sinal celular ou a proliferação celular;
- b) um polipéptido compreendendo uma variante alélica de ocorrência natural da sequência de aminoácidos dos resíduos de aminoácido 20-224 de SEQ ID NO:5, dos resíduos de aminoácido 21-266 de SEQ ID NO:8, SEQ ID NO:11 ou SEQ ID NO:14, em que a referida variante alélica de ocorrência natural modula a transdução de sinal celular ou a proliferação celular e em que (i) a referida variante alélica de ocorrência natural da sequência de aminoácidos dos resíduos de aminoácido 20-224 de SEQ ID NO:5 é codificada por uma molécula de ácido nucleico que hibrida com o complemento de SEQ ID NO:6 sob condições de lavagem rigorosas, (ii) a referida variante alélica de ocorrência natural da sequência de aminoácidos de SEQ ID NO:8 é codificada por uma molécula de ácido nucleico que hibrida com o complemento de SEQ ID NO:9 sob condições de lavagem rigorosas, (iii) a referida variante alélica de ocorrência natural da sequência de aminoácidos de SEQ ID NO:11 é codificada por uma molécula de ácido nucleico que hibrida com o complemento de SEQ ID NO:12 sob condições de lavagem rigorosas, e (iv) a referida variante alélica de ocorrência natural da sequência de aminoácidos de SEQ ID NO:14 é codificada por uma molécula de ácido nucleico que hibrida com o complemento de SEQ ID NO:15 sob condições de lavagem rigorosas, e em que as condições de lavagem rigorosas são SSC 0,2X, SDS a 0,1% a 50-65°C;
- c) um polipéptido compreendendo os resíduos de aminoácido 24-350 de SEQ ID NO:2; e
- d) um polipéptido compreendendo uma sequência de aminoácidos que possui pelo menos 80% de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID NO:5, SEQ ID NO:8 ou SEQ ID NO:11 ao longo de todo o comprimento de SEQ ID NO:5, SEQ ID NO:8 ou SEQ ID NO:11, respectivamente, e modula a transdução de sinal celular ou a proliferação celular.

11. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido compreende pelo menos 15 aminoácidos contíguos de SEQ ID NO:8 e em que o referido polipéptido modula a transdução de sinal celular ou a proliferação celular.

12. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido compreende pelo menos 20 aminoácidos contíguos de SEQ ID NO:5, SEQ ID NO:11 ou SEQ ID NO:14 e em que o referido polipéptido modula a transdução de sinal celular ou a proliferação celular.

13. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido compreende pelo menos 30 aminoácidos contíguos de SEQ ID NO:5, SEQ ID NO:11 ou SEQ ID NO:14 e em que o referido polipéptido modula a transdução de sinal celular ou a proliferação celular.

14. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido compreende pelo menos 50 aminoácidos contíguos de SEQ ID NO:5, SEQ ID NO:11 ou SEQ ID NO:14 e em que o referido polipéptido modula a transdução de sinal celular ou a proliferação celular.

15. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido compreende pelo menos 100 aminoácidos contíguos de SEQ ID NO:5, SEQ ID NO:11 ou SEQ ID NO:14 e em que o referido polipéptido modula a transdução de sinal celular ou a proliferação celular.

16. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido compreende pelo menos 20 aminoácidos contíguos de SEQ ID NO:8 e em que o referido polipéptido modula a transdução de sinal celular ou a proliferação celular.

17. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido compreende pelo menos 30 aminoácidos contíguos de SEQ ID NO:8 e em que o referido polipéptido modula a transdução de sinal celular ou a proliferação celular.

18. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido compreende pelo menos 50 aminoácidos contíguos de SEQ ID NO:8 e em que o referido polipéptido modula a transdução de sinal celular ou a proliferação celular.

19. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido compreende pelo menos 100 aminoácidos contíguos de SEQ ID NO:8 e em que o referido polipéptido modula a transdução de sinal celular ou a proliferação celular.

20. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido isolado compreende uma sequência de aminoácidos seleccionada do grupo que consiste em SEQ ID NO:8 e nos resíduos de aminoácido 21-266 de SEQ ID NO:8.

21. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido isolado compreende uma sequência de aminoácidos seleccionada do grupo que consiste em SEQ ID NO:2, nos resíduos de aminoácido 20-350 de SEQ ID NO:2, nos resíduos de aminoácido 22-350 de SEQ ID NO:2 e nos resíduos de aminoácido 24-350 de SEQ ID NO:2.

22. Polipéptido isolado de acordo com a reivindicação 10, em que o referido polipéptido isolado compreende uma sequência de aminoácidos seleccionada do grupo que consiste em SEQ ID NO:5, nos resíduos de aminoácido 20-224 de SEQ ID NO:5, SEQ ID NO:11 e SEQ ID NO:14.

23. Polipéptido isolado de acordo com qualquer uma das reivindicações 10-22 compreendendo ainda um polipéptido heterólogo.

24. Anticorpo ou sua porção imunologicamente activa que se liga especificamente a um polipéptido seleccionado do grupo que consiste em:

- a) um polipéptido que consiste em 15 aminoácidos contíguos de SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11 ou SEQ ID NO:14; e
- b) um polipéptido que consiste numa variante alélica de ocorrência natural da sequência de aminoácidos dos

resíduos de aminoácido 20-224 de SEQ ID NO:5, dos resíduos de aminoácido 21-266 de SEQ ID NO:8, SEQ ID NO:11 ou SEQ ID NO:14, em que a referida variante alélica de ocorrência natural modula a transdução de sinal celular ou a proliferação celular e em que (i) a referida variante alélica de ocorrência natural da sequência de aminoácidos dos resíduos de aminoácido 20-224 de SEQ ID NO:5 é codificada por uma molécula de ácido nucleico que hibrida com o complemento de SEQ ID NO:6 sob condições de lavagem rigorosas, (ii) a referida variante alélica de ocorrência natural da sequência de aminoácidos dos resíduos de aminoácido 21-266 de SEQ ID NO:8 é codificada por uma molécula de ácido nucleico que hibrida com o complemento de SEQ ID NO:9 sob condições de lavagem rigorosas, (iii) a referida variante alélica de ocorrência natural da sequência de aminoácidos de SEQ ID NO:11 é codificada por uma molécula de ácido nucleico que hibrida com o complemento de SEQ ID NO:12 sob condições de lavagem rigorosas, (iv) a referida variante alélica de ocorrência natural da sequência de aminoácidos de SEQ ID NO:14 é codificada por uma molécula de ácido nucleico que hibrida com o complemento de SEQ ID NO:15 sob condições de lavagem rigorosas, e em que as condições de lavagem rigorosas são SSC 0,2X, SDS a 0,1%, a 50-65°C.

25. Anticorpo ou sua porção imunologicamente activa de acordo com a reivindicação 24, em que o referido anticorpo ou a sua porção imunologicamente activa se ligam especificamente a um polipéptido que consiste em 15 aminoácidos contíguos de SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11 ou SEQ ID NO:14.

26. Anticorpo ou sua porção imunologicamente activa de acordo com a reivindicação 24, em que o referido anticorpo ou a sua porção imunologicamente activa se ligam especificamente a um polipéptido que consiste em 20 aminoácidos contíguos de SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11 ou SEQ ID NO:14.

27. Anticorpo ou sua porção imunologicamente activa de acordo com a reivindicação 24, em que o referido anticorpo ou a sua porção imunologicamente activa se ligam especificamente a um polipéptido que consiste em 30 aminoácidos contíguos de SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11 ou SEQ ID NO:14.

28. Anticorpo ou sua porção imunologicamente activa de acordo com a reivindicação 24, em que o referido anticorpo ou a sua porção imunologicamente activa se ligam especificamente a um polipéptido que consiste numa sequência de aminoácidos seleccionada do grupo que consiste em SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11 e SEQ ID NO:14.

29. Anticorpo ou sua porção imunologicamente activa de acordo com a reivindicação 28, em que o referido anticorpo ou a sua porção imunologicamente activa se ligam especificamente a um polipéptido que consiste na sequência de aminoácidos dos resíduos de aminoácido 21-266 de SEQ ID NO:8.

30. Anticorpo ou sua porção imunologicamente activa de acordo com a reivindicação 28, em que o referido anticorpo ou a sua porção imunologicamente activa se ligam especificamente a um polipéptido que consiste na sequência de aminoácidos dos resíduos de aminoácido 20-224 de SEQ ID NO:5.

31. Anticorpo ou sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-30, em que o referido anticorpo ou a sua porção imunologicamente activa são seleccionados do grupo que consiste num anticorpo humano, um anticorpo monoclonal humanizado, um anticorpo monoclonal quimérico e uma porção imunologicamente activa de qualquer um dos anteriores.

32. Anticorpo ou sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-30 em que o referido anticorpo ou a sua porção imunologicamente activa são um anticorpo monoclonal ou uma porção imunologicamente activa de um anticorpo monoclonal.

33. Anticorpo ou sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-30, em que o referido anticorpo ou a sua porção imunologicamente activa são uma porção imunologicamente activa seleccionada do grupo que consiste num fragmento Fab e um fragmento $F(ab')_2$.

34. Anticorpo ou sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-33, em que o

referido anticorpo ou a sua porção imunologicamente activa compreendem um marcador detectável.

35. Anticorpo ou sua porção imunologicamente activa de acordo com a reivindicação 34, em que o referido marcador detectável é uma enzima, um grupo prostético, um material fluorescente, um material luminescente, um material bioluminescente ou um material radioactivo.

36. Método para a produção de um polipéptido de acordo com a reivindicação 10 compreendendo a cultura da célula hospedeira de acordo com a reivindicação 8 sob condições em que o vector de expressão recombinante é expresso, e o referido polipéptido é produzido.

37. Método para a produção de um polipéptido de acordo com a reivindicação 20 ou a reivindicação 21 compreendendo a cultura da célula hospedeira de acordo com a reivindicação 9 sob condições em que o ácido nucleico recombinante é expresso, e o referido polipéptido é produzido.

38. Método para a detecção da presença de um polipéptido de acordo com qualquer uma das reivindicações 10-23 numa amostra compreendendo:

- a) o contacto da amostra com um anticorpo ou com uma sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-35; e
- b) a determinação de se o referido anticorpo se liga a um polipéptido na amostra para desse modo detectar a presença do referido polipéptido de acordo com qualquer uma das reivindicações 10-23 na amostra.

39. *Kit* compreendendo um anticorpo ou uma sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-35 e instruções de utilização.

40. Utilização de um *kit* para a detecção de um polipéptido de acordo com qualquer uma das reivindicações 10-23 numa amostra, o referido *kit* compreendendo um anticorpo ou uma sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-35 e instruções de utilização.

41. Método para a detecção da presença de uma molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1-5, numa amostra, compreendendo:

- a) o contacto da amostra com uma sonda ou um iniciador de ácido nucleico que hibridam especificamente com uma molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1-5; e
- b) a determinação de se a sonda ou o iniciador de ácido nucleico se ligam a uma molécula de ácido nucleico na amostra para desse modo detectar a presença de uma molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1-5 na amostra.

42. Método de acordo com a reivindicação 41, em que a amostra compreende moléculas de ARNm e é colocada em contacto com uma sonda de ácido nucleico.

43. Utilização de um *kit* para a detecção de uma molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1-5 numa amostra, o referido *kit* compreendendo uma sonda ou um iniciador de ácido nucleico que hibridam especificamente com uma molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1-5 e instruções de utilização.

44. Método *in vitro* para a identificação de um composto que se liga a um polipéptido de acordo com qualquer uma das reivindicações 10-23 compreendendo:

- a) o contacto de um polipéptido de acordo com qualquer uma das reivindicações 10-23, ou de uma célula que expressa o referido polipéptido, com um composto de teste; e
- b) a determinação de se o polipéptido se liga ao composto de teste.

45. Método de acordo com a reivindicação 44, em que a ligação do composto de teste ao polipéptido é detectada por um método seleccionado do grupo que consiste em:

- a) detecção de ligação por detecção directa da ligação composto de teste/polipéptido;

- b) detecção de ligação utilizando um ensaio de ligação de competição; e
- c) detecção de ligação utilizando um ensaio para detecção de transdução de sinal celular ou proliferação celular.

46. Método *in vitro* de modulação da actividade de transdução de sinal celular ou de proliferação celular de um polipéptido de acordo com qualquer uma das reivindicações 10-23 compreendendo o contacto de um polipéptido de acordo com qualquer uma das reivindicações 10-23 ou de uma célula que expressa o referido polipéptido, com um anticorpo ou uma sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-35, numa concentração suficiente para modular a actividade de transdução de sinal celular ou de proliferação celular do referido polipéptido.

47. Método para identificação de um composto que modula a actividade de transdução de sinal celular ou de proliferação celular de um polipéptido de acordo com qualquer uma das reivindicações 10-23 compreendendo:

- a) o contacto de um polipéptido de acordo com qualquer uma das reivindicações 10-23 com um composto de teste; e
- b) a determinação do efeito do composto de teste sobre a actividade de transdução de sinal celular ou de proliferação celular do referido polipéptido para desse modo identificar um composto que modula a actividade de transdução de sinal celular ou de proliferação celular do referido polipéptido.

48. Anticorpo ou sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-35 para utilização em terapia ou diagnóstico.

49. Utilização de um anticorpo ou de uma sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-35 para o fabrico de um medicamento para tratamento de uma desordem de desenvolvimento, de diferenciação ou de proliferação.

50. Polipéptido de acordo com qualquer uma das reivindicações 10-23 para utilização em terapia ou diagnóstico.

51. Utilização de um polipéptido de acordo com qualquer uma das reivindicações 10-23 para o fabrico de um medicamento para tratamento de uma desordem de desenvolvimento, de diferenciação ou de proliferação.

52. Utilização de um polipéptido de acordo com qualquer uma das reivindicações 10-23 para o fabrico de um medicamento ou diagnóstico para o tratamento, profilaxia ou diagnóstico do cancro.

53. Molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1-7 para utilização em terapia ou diagnóstico.

54. Utilização de uma molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1-7 para o fabrico de um medicamento para tratamento de uma desordem de desenvolvimento, de diferenciação ou de proliferação.

55. Utilização de um anticorpo ou de uma sua porção imunologicamente activa de acordo com qualquer uma das reivindicações 24-35 para o fabrico de um medicamento ou diagnóstico para o tratamento, profilaxia ou diagnóstico do cancro.

56. Utilização de uma molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1-7 para o fabrico de um medicamento ou diagnóstico para o tratamento, profilaxia ou diagnóstico do cancro.

57. Célula hospedeira não humana compreendendo um transgene que compreende a molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1a), 1b), 2, 3, 4 ou 5.

Lisboa,

Figura 1A

```

                                M Q R L G A T L L C   10
GGCACGAGGGGGCGGGCTGCGGGCCGAGAGCGGAG ATG CAG CGG CTT GGG GCC ACC CTG CTG TGC   67

L L L A A A V P T A P A P A P T A T S A   30
CTG CTG CTG GCG GCG GCG GTC CCC ACG GCC CCC GCG CCC GCT CCG ACG GCG ACC TCG GCT   127

P V K P G P A L S Y P Q E E A T L N E M   50
CCA GTC AAG CCC GGC CCG GCT CTC AGC TAC CCG CAG GAG GAG GCC ACC CTC AAT GAG ATG   187

F R E V E E L M E D T Q H K L R S A V E   70
TTC CGC GAG GTT GAG GAA CTG ATG GAG GAC ACG CAG CAC AAA TTG CGC AGC GCG GTG GAA   247

E M E A E E A A A K A S S E V N L A N L   90
GAG ATG GAG GCA GAA GAA GCT GCT GCT AAA GCA TCA TCA GAA GTG AAC CTG GCA AAC TTA   307

P P S Y H N E T N T D T N V G N N T I H   110
CCT CCC AGC TAT CAC AAT GAG ACC AAC ACA GAC ACG AAC GTT GGA AAT AAT ACC ATC CAT   367

V H R E I H K I T N N Q T G Q M V F S E   130
GTG CAC CGA CAA ATT CAC AAG ATA ACC AAC AAC CAG ACT GGA CAA ATG GTC TTT TCA GAG   427

T V I T S V G D E E G R R S H E C I I D   150
ACA GTT ATC ACA TCT GTG GGA GAC GAA GAA GGC AGA AGG AGC CAC GAG TGC ATC ATC GAC   487

E D C G P S M Y C Q F A S F Q Y T C Q P   170
GAG GAC TGT GGG CCC AGC ATG TAC TGC CAG TTT GCC AGC TTC CAG TAC ACC TGC CAG CCA   547

C R G Q R M L C T R D S E C C G D Q L C   190
TGC CGG GGC CAG AGG ATG CTC TGC ACC CGG GAC AGT GAG TGC TGT GGA GAC CAG CTG TGT   607

V W G H C T K M A T R G S N G T I C D N   210
GTC TGG GST CAC TGC ACC AAA ATG CCC ACC AGG GGC AGC AAT GGG ACC ATC TGT GAC AAC   667

Q R D C Q P G L C C A F Q R G L L F P V   230
CAG AGG GAC TGC CAG CCG GGG CTG TGC TGT GCC TTC CAG AGA GGC CTG CTG TTC CCT GTG   727

C T P L P V E G E L C H D P A S R L L D   250
TGC ACA CCC CTG CCC GTG GAG GGC GAG CTT TGC CAT GAC CCC GCC AGC CGG CTT CTG GAC   787

L I T W E L E P D G A L D R C P C A S G   270
CTC ATC ACC TGG GAG CTA GAG CCT GAT GGA GCC TTG GAC CGA TGC CCT TGT GCC AGT GGC   847

L L C Q P H S H S L V Y V C K P T F V G   290
CTC CTC TGC CAG CCC CAC AGC CAC AGC CTG GTG TAT GTG TGC AAG CCG ACC TTC GTG GGG   907

S R D Q D G E I L L P R E V P D E Y E V   310
AGC CGT GAC CAA GAT GGG GAG ATC CTG CTG CCC AGA GAG GTC CCC GAT GAG TAT GAA GTT   967

G S F M E E V R Q E L F D L E R S L T E   330
GGC AGC TTC ATG GAG GAG GTG CGC CAG GAG CTC CAG GAC CTG GAG AGG AGC CTG ACT GAA   1027

E M A L R E P A A A A A A L L G R E E I   350
GAG ATG GCG CTG AGG GAG CCT GCG GCT GCC GCC GCT GCA CTG CTG GGA AGG GAA GAG ATT   1087

*
TAG
                                351
                                1090

```

Figura 1B

ATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAAATTTATTTCCCCANGTCTGTGCTTTAAGCGTGGGCTG 1169
ACCAGGCTTCTTCCATACATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTTGACAGCATGAGGTGTTGTGCAITTTGTTTCAG 1248
CTCCCCCAGGCTGTTCTCCAGGCTTCACAGTCTGGTCTGGGAGACTCAGGCAGGGTTAACTGCAGGAGCAGTTTGC 1327
CACCCCTGTCCAGATTATTGGCTGCTTTGGCTCTACCCAGTTGGCAGACAGCCGTTTGTCTACATGGCTTTGATTAATTG 1406
TTTGAGGGGAGGAGTGAACAATGTGGAGTCTCCCTCTGATTTGGTATTTGGGGAATGTGGAGNACAGTCCCCCTGCTT 1485
TGCAAAATCAACCTGGCAAAAATGCAACAAAATGAATTTCCACGAGTCTTTCCATGGGCATAGGTAAGCTGTGCCT 1564
TCAGCTGTTGCAGATGAAAATGTTCTGTTCACTGTTCACTGTTTATTCATCCAGCAGTGTGCTCAGCTCCTAC 1643
CTCTGTGCCAGGGCAGCATTTTCATATCCAAAGATCAATTCCTCTCTCAGCACAGCCTGGGAGGGGTTCAITTTGTTCTC 1722
CTCGTCCATCAGGGATTTCAGAGGCTCAGAGACTGCCAAGCTGTGCCCAAGTCCACAGCTAGTGAAGACCAGAGCAG 1801
TTTCATCTGGTTGTGACTCTAAGCTCAGTGTCTCTCCACTACCCACACCAGCCTTGGTGCACCCAAAAGTGTCTCCC 1880
AAAAGGAGGAGCAATGGGATTTTCTTTTGAGGCTGCACATCTGGCAATTAAGGTCAAACTAATTTTCACATCCCTCTYA 1959
AAAGTAAACTACTGTTAGGAACAGCAGTGTCTCACAGTGTGGGGCAGCCGCTCTCTAATGAAGACAATGATTTGAC 2038
ACTGTCCCTCTTTGGCAGTTGCATTAGTAACTTTGAAAAGTATATGACTGAGGGTAGCATACAGGTTAACCTGCAGAAA 2117
CAGTACTTAGGTAATGTAGGGCAGGATTAATAATGAAATTTGCAAAATCACTTAGCAGCAACTGAAGACAATTAACA 2196
ACCAGTGGAGAAAATCAAAACGAGCAGGGCTGTGTGAAAACATGGTTGTAATATGCGACTGCGAACACTGAUCTTAGG 2275
CCACTCCACAAATGATGTTTTTCAGGTGTCTAGGACTGTTGCCACCATGTTATTCATCCAGAGTTCTTTAAAGTTTAAAGCTT 2354
GCACATGATTTGATATAGCATGCTTTCTTTGAGTTTAAATTAATGATTAACAATAGTGTGATTTAGAAAATCAAGCATPA 2433
ATCAGTTCAACTGCTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2479

Figura 2

GAATTGGCAGCAGACGACGCTGCTGAGTCCAGCTTAGTGGAGCTCTGCTGGTGGAGAGCAGCCCTGGCTTTG 79
 M V A A V L L G 8
 GTGACGCACAGTCTGGACCCCTCCAGGAGCCCGGATTAAGG ATG GTG GCG GCC GTC CTG CTG GGG 148
 L S W L C S P L G A L V L D F N M I R S 28
 CTG AGC TGG CTC TGC TCT CCC CTG GGA GCT CTG GTC GAC TTC AAC AAC ATC AGG AGC 208
 S A D L L H G A R K G S Q C L S D T D C N 48
 TCT GCT GAC CTG CAT GGG GCC CUG AAG GGC TCA CAG TGC CTG TCT GAC ACG GAC TGC AAT 268
 T R K F C L Q P R D E K P F C A T C R G 68
 ACC AGA AAG TTC TGC CTC CAG CCC CGC GAT GAG AAG CCG TTC TGT GCT ACA TGT CGT GGG 328
 L R R R C Q R D A M C C P G T L C V N D 88
 TTG CCG AGG AGG TGC CAG CGA GAT GCC ATG TGC TGC CCT CCG ACA CTC TGT GTG AAC GAT 388
 V C T T M E D A T P I L E R Q L D E Q D 108
 GTT TGT ACT ACG ATG GAA GAT GCA ACC GCA ATA TTA GAA AGG CAG CTT GAT GAG CAA GAT 448
 G T H A E G T T G H P V Q E N Q P K R K 128
 GGC ACA CAT GCA GAA GGA ACA ACT GGG CAC CCA GTC CAG GAA AAC CAA CCC AAA AGG AAG 508
 P S I K K S Q G R K G Q E G E S C L R T 148
 CCA AGT ATT AAG AAA TCA CAA GGC AGG AAG GGA CAA GAG GGA GAA AGT TGT CTG AGA ACT 568
 F D C G P G L C C A R H F W T K I C K P 168
 TTT GAC TGT GGC CCT GGA CTT TGC TGT GCT CCG CAT TTT TGG AGC AAA ATT TGT AAG CCA 628
 V L L E G Q V C S R R G H K D T A Q A P 188
 GTC CTT TTG GAG GGA CAG GTC TGC TCC AGA AGA GGG CAT AAA GAC ACT CCT CAA GCT CCA 688
 E I F Q R C D C G P G L L C R S Q L T S 208
 GAA ATC TTC CAG CBT TGC GAC TGT GGC CCT GGA CTA CTG TGT CGA AGC CAA TTG ACC AGC 748
 N R O H A R L R V C O K I E K L * 225
 AAT CCG CAG CAT GCT CGA TTA AGA GTA TGC CAA AAA ATA GAA AAG CTA TAA 799
 ATATTTCAAATATAAGAGATCCACATTTGCCAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 848

Figura 3A

CCGGACGCGTGGGCGGCACGGTTTCGTGGGGACCCAGGCTTGCAAAGTGACGGTCATTTCTCTTTCTTCTCCCTCTT 79

 M M A L G A A G A T R V F V A M 16
GAGTCTTCTGAG ATG ATG GCT CTG GGC GCA GCG GGA GCT ACC CGG GTC TTT GTC GCG ATG 140

V A A A L G G H P L L G V S A T L N S V 36
GTA GCG GCG GCT CTC GGC GGC CAC CCT CTG CTG GGA GTG AGC GCC ACC TTG AAC TCG GTT 200

L N S N A I K N L P P P L G G A A G H P 56
CTC AAT TCC AAC GCT ATC AAG AAC CTG CCC CCA CCG CTG GGC GGC GCT GCG GGG CAC CCA 260

G S A V S A A P G I L Y P G G N K Y Q T 76
GGC TCT GCA GTC AGC GCC GCG CCG GGA ATC CTG TAC CCG GGC GGG AAT AAG TAC CAG ACC 320

I D N Y Q P Y P C A E D E E C G T D E Y 96
ATT GAC AAC TAC CAG CCG TAC CCG TGC GCA GAG GAC GAG GAG TGC GGC ACT GAT GAG TAC 380

C A S P T R G G D A G V Q I C L A C R K 116
TGC GCT AGT CCC ACC CCG GGA GGG GAC GCA GGC GTG CAA ATC TGT CTC GCC TGC AGG AAG 440

R R K R C M R H A M C C P G N Y C K N G 136
CGC CGA AAA CCG TGC ATG CGT CAC GCT ATG TGC TGC CCC GGG AAT TAC TGC AAA AAT GGA 500

I C V S S D Q N H F R G E I E E T I T E 156
ATA TGC GTG TCT TCT GAT CAA AAT CAT TTC CGA GGA GAA ATT GAG GAA ACC ATC ACT GAA 560

S F G N D H S T L D G Y S R R T T L S S 176
AGC TTT GGT AAT GAT CAT AGC ACC TTG GAT GGG TAT TCC AGA AGA ACC ACC TTG TCT TCA 620

K M Y H T K G Q E G S V C L R S S D C A 196
AAA ATG TAT CAC ACC AAA GGA CAA GAA GGT TCT GTT TGT CTC CGG TCA TCA GAC TGT GCC 680

S G L C C A R H F W S K I C K P V L K E 216
TCA GGA TTG TGT TGT GCT AGA CAC TTC TGG TCC AAG ATC TGT AAA CCT GTC CTG AAA GAA 740

G Q V C T K H R R K G S H G L E I F Q R 236
GGT CAA GTG TGT ACC AAG CAT AGG AGA AAA GGC TCT CAT GGA CTA GAA ATA TTC CAG CGT 800

C Y C G E G L S C R I Q K D H H Q A S N 256
TGT TAC TGT GGA GAA GGT CTG TCT TGC CCG ATA CAG AAA GAT CAC CAT CAA GCC AGT AAT 860

S S R L H T C Q R H * 267
TCT TCT AGG CTT CAC ACT TGT CAG AGA CAC TAA 893

ACCAGCTATCCAAAATGCAGTGAACICCTTTTATATAATAGATGCTATGAAAACCTTTTATGACCTTCATCAACTCAAT 972

CCTAAGGATATACAAGTTCTGTGGTTTCAGTTAAGCATTCCAATAACACCTTCCAAAAACCTGGAGTGTAAAGAGCTTTG 1051

TTTCTTTATGGAACCTCCCTGTGATTGCAGTAAATTACTGTATGTAAATTCAGTGTGGCACTTACCTGTAATGCA 1130

ATGAAACITTTAATTTATTTTCTAAAGGTGCTGCACTGCCTATTTTCTCTTGTTCGTAATTTTGTACACATTGA 1209

TTGTTATCTTGACTGACAAATATTCTATATTGAACTGAAGTAAATCATTTCAGCTTATAGTTCTTAAAGCATAACCCT 1288

TTACCCCATTTNATTTCTAGAGTCNAGAACGCAAGGATCTCTTGGAAATGACAAATGATAGGTACCTAAAAATGTAACATGA 1367

Figura 3B

AAATACTAGCTTATTTTCTGAAATGTACTATCTTAATGCTTAAATTTATATTTCCCTTTAGGCTGTGATAGTTTTTGAAA 1446
TAAAATTTAACAATTTAATATCATGAAATGKTATAAGTAGACATAAAAAAAGGGCCGCTAGA 1525
CTAG 1529

Figura 4

E F G T R V G R Y C H S P H Q G S S A C 20
 GAA TTC GGC ACG AGG GTT GGG AGG TAT TGC CAC AGT CCC CAC CAA GGA TCA TCG GCC TGC 60
 M V C R R K K R C H R D G M C C P S T 40
 ATG GTG TGT CGG AGA AAA AAG AAG CCG TGC CCA GAT GGC ATG TGC CCC AGT ACC 120
 R C N N G I C I P V T E S I L T P H I P 60
 CGC TGC AAT AAT GGC ATC TGT ATC CCA GTT ACT GAA AGC ATC TTA ACC CCT CAC ATC CCG 180
 A L D G T R H R D R N H G H Y S N H D L 80
 CCT CTG GAT GGT ACT CGG CAC AGA GAT CGA AAC CAC GGT CAT TAC TCA AAC CAT GAC TTG 240
 G W Q N L G R P H T K M S H I K G H E G 100
 GGA TGG CAG AAT CTA GGA AGA CCA CAC ACT AAG ATG TCA CAT ATA AAA GGG CAT GAA GGA 300
 D P C L R S S D C I E G F C A R H F W 120
 GAC CCC TGC CTA CGA TCA GAC TGC ATT GAA GGG TTT TGC TGT GCT CGT CAT TTC TGG 360
 T K I C K P V L H Q G E V C T K Q R K K 140
 ACC AAA ATC TGC AAA CCA GTG CTC CAT CAG GGG GAA GTC TGT ACC AAA CAA CGC AAG AAG 420
 G S H G L E I F Q R C D C A K G L S C K 160
 GGT TCT CAT GGG CTG GAA ATT TTC CAG CGT TGC GAC TGT GCG AAG GGC CTG TCT TGC AAA 480
 V W K D A T Y S S K A R L H V C Q K I * 180
 GTA TGG AAA GAT GCC ACC TAC TCC AAA GCC AGA CTC CAT GTG TGT CAG AAA ATT TGA 540
 TCACCATTGAGGAACATCATCAATTGCAGACTGTGGAAGTTGTGTATTTAATGCAATTATAGCATGCTGGAAATAGGTT 619
 CAGATGCAGAAAGTGGCTAAAAATAAGAAAACGGTGTATAGAAATATAGATGATCACAATAAAAAAAAAAAAAAAAAAGATGCCG 698
 CCGC 702

Figura 5

CTCGAGGCCAAAATTCGGCACGAGCCGGGCTGTGGTCTAGCATAAAGCGGAGCCAGAAAGAAGGGCGGGGT	M	1
	ATG	77
G E A S P P A P A R R H L L V L L L L L		21
GGA GAA GCC TCC CCA CCT GCC CCC GCA AGG CGG CAT CTG CTG GTC CTG CTG CTG CTC CTC		137
S T L V I P S A A A P I H D A D A Q E S		41
TCT ACC CTG GTG ATC CCC TCC GCT GCA GCT CCT ATC CAT GAT GCT GAC GCC CAA GAG AGC		197
S L G L T G L Q S L L Q G F S R L F L K		61
TCC TTG GGT CTC ACA GGC CTC CAG AGC CTA CTC CAA GGC TTC AGC CGA CTT TTC CTG AAA		257
G N L L R G I D S L F S A P M D F R G L		81
GGT AAC CTG CTT CGG GGC ATA GAC AGC TTA TTC TCT GCC CCC ATG GAC TTC CGG GGC CTC		317
P G N Y H K E E N Q E H Q L G N N T L S		101
CCT GGG AAC TAC CAC AAA GAG GAG AAC CAG GAG CAC CAG CTG GGG AAC AAC ACC CTC TCC		377
S H L Q I D K M T D N K T G E V L I S E		121
AGC CAC CTC CAG ATC GAC AAG ATG ACC GAC AAC AAG ACA GGA GAG GTG CTG ATC TCC GAG		437
N V V A S I Q F A E G S F E G D L K V P		141
AAT GTG GTG GCA TCC ATT CAA CCA GCG GAG GGG AGC TTC GAG GGT GAT TTG AAG GTA CCC		497
R M E E K E A L V P I Q K A T D S F H T		161
AGG ATG GAG GAG AAG GAG GCC CTG GTA CCC ATC CAG AAG GCC ACG GAC AGC TTC CAC ACA		557
E L H P R V A F W I I K L P R R R S H Q		181
GAA CTC CAT CCC CGG GTG GCC TTC TGG ATC ATT AAG CTG CCA CGG CGG AGG TCC CAC CAG		617
D A L E G G H W L S E K R H R L Q A I R		201
GAT GCC CTG GAG GGC GGC CAC TGG CTC AGC GAG AAG CGA CAC CGC CTG CAG GCC ATC CGG		677
D G L R K G T H K D V L E E G T E S S S		221
GAT GGA CTC CGC AAG GGG ACC CAC AAG GAC GTC CTA GAA GAG GGG ACC GAG AGC TCC TCC		737
H S R L S P R K T H L L Y I L R P S R Q		241
CAC TCC AGG CTG TCC CCC CGA AAG ACC CAC TTA CTG TAC ATC CTC AGG CCC TCT CGG CAG		797
L *		243
CTG TAG		803
GGGTGGGACCGGGAGCACCTGCCTGTAGCCCCATCAGACCCTGCCCAAGCACCATATGGAATAAAGTCTTTCT		882
TACATCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGGCGGCCGC		928

FIGURA 6

```

crsp-2h .....
crsp-3h .....
crsp-4h .....
tango59 MQRLLGATLLCLLLAAAVPTAFAPAPTATSAPVKFGPALS Y
Consensus ..... 40

crsp-2h .....
crsp-3h ..... MMALGAGATRVFV MVA
crsp-4h .....
tango59 PQEEATLNEMFREVEELHEDTQHLLRS AVEEMEAE EAAK
Consensus ..... a a 80

crsp-2h ..... MVAAVLLGLSWLCS
crsp-3h M ALGGKPLLGVSATL N SVLNSNAIK N LPPPLGGAAACRPGS
crsp-4h .....
tango59 M SSEVNLANLPPSYH TETNTDTNVGN TTIHVHREIHKITN
Consensus ..... a n n v s 120

crsp-2h PLGALVLDFNIR S ADLHCA R KGS Q CLS D DCM T R K F C L
crsp-3h AVS A APGILYPPGNXY Q TIDNY Q P Y P C A E D E E C G D E Y C A
crsp-4h ..... EFGTRVGRVCH
tango59 NQTGQHVFSSETVITS V G D E E G R S H E C I I D E D C C P S N Y C Q
Consensus ..... a s x C D e d C g t Y C 160

crsp-2h Q P R D E K F F . . . . C A T C R G L R A R C Q R D A M C C P G T L C V N D V C
crsp-3h S P T R G G D A G V Q I C L A C R K R R R C M R H A M C C P G N Y C R N G I C
crsp-4h S P H Q Q S S A . . . . C M V C R R K R R C H R D M C C P S T R C N V C I C
tango59 F A S P Q Y T . . . . C Q P C R C Q R N L C T R D S E C C G D Q L C V W G H C
Consensus ..... s p g a C C R g R k R C R D a M C C P g t l c v n g i c 200

crsp-2h T T M E D A T P I L L E R Q L D S Q D G H A E C T T G H P V Q E N Q . . . P K R
crsp-3h V S S D Q N . . H F R G E X E T I . E S F C N D H S T L D . . . G Y S R I
crsp-4h I P V T E S . . H I T P H I P A L D C T R H R D R N E G H Y S N H D L G W Q N L
tango59 K K M . . . . . i l i a d g t . . . . . g h . . . . . g r
Consensus ..... t m . . . . . i l i a d g t . . . . . g h . . . . . g r 240

crsp-2h R P S I K S Q G R K G Q E G E S C L A T F M C G P G L C C A R H F W T K . . I
crsp-3h T T L S S E M Y H T K G O E G S V C L R S D C A S G L C C A R H F W S K . . I
crsp-4h G R P H T M S H I E G H E G D P C L R S D C T E G P C C A R H F W T K . . I
tango59 . . . . . A R G S N G T I C D N Q R D C G P G L C C A F O R G L L F P V
Consensus ..... K m h t K G q e g C L R s s d C p G L C C A R H F W t K I 280

crsp-2h K P V L E G Q V C S R R G N K D T . . . . . A Q A P E I F O R C D C G P G
crsp-3h K P V L E G Q V C T K H R K G . . . . . S H G L E I F O R C Y C G E G
crsp-4h K P V L Q G E V C T K Q R K K G . . . . . S H G L E I F O R C D C A K G
tango59 L T D L F V E C E L C H D P A S R L L D L I T W E L F P D G A L D R C P C A S G
Consensus ..... C K P V L E G q v C t k r K g . . . . . s h g l e i f o r c d c a g 320

crsp-2h L A C R S Q L T S N R . . Q H A R L R V C Q K I E K L . . . . .
crsp-3h L S C R I Q D H H Q A S N S S R L H T C Q R H . . . . .
crsp-4h L S K V W E D A T Y S K A R L H V C Q K I . . . . .
tango59 L L C Q P H S H L . . . . . V Y V C K P T F V G S R D Q D G E I L L P R E
Consensus ..... L s C r q k d s . . . . . s a R L h V C Q k i 360

crsp-2h .....
crsp-3h .....
crsp-4h .....
tango59 V P D E Y E V G S F M E E V R Q E L E D L E R S L T E E M . . . R E P A A A A A A
Consensus ..... 400

crsp-2h .....
crsp-3h .....
crsp-4h .....
tango59 L L G R E E I
Consensus ..... 406

```

FIGURA 7
A Família da CRSP Humana

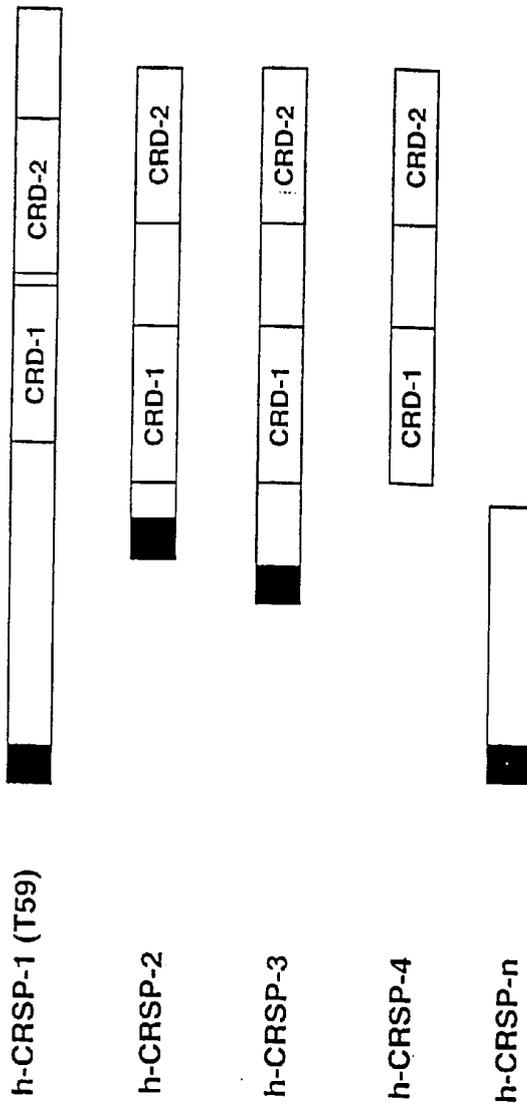


Figura 8A

```

FOTCGACCCACGGCTCCGCTGTGGCAGCCAGCTACCGGTGCGTACCAGATCCAGCTTGCAGCTCAGCTTTGTTTCATTC 79
                                     M Q R L G G I L L C T L      12
GAATTGGGCGGCGGCCAGCGCGSAACAAAC ATG CAG CGG CTC GGG GGT ATT TTG CTG TGT ACA CTG      145
L A A A V P T A P A P S P T V T W T P A      32
CTG GCG GCG GCG GTC CCC ACT GCT CCT GCT CCT TCC CCG ACG GTC ACT TGG ACT CCG GCG      205
E P G P A L N Y P Q E E A T L N E M F R      52
GAG CCG GGC CCA GCT CTC AAC TAC CCT CAG GAG GAA GCT ACG CTC AAT GAG ATG TTT CGA      265
E V E E L M E D T Q H K L R S A V E E M      72
GAG GTG GAG GAG CTG ATG GAA GAC ACT CAG CAC AAA CTG CGC AGT GCC CTG GAG GAG ATG      325
E A E E A A A K T S S E V N L A S L P P      92
GAG GCG GAA GAA GCA GCT GCT AAA ACG TCC TCT GAG GTG AAC CTG GCA AGC TTA CCT CCC      385
N Y H N E T S T E T R V G N N T V H V H      112
AAC TAT CAC AAT GAG ACC AGC ACG GAG ACC AGG GTG GGA AAT AAC ACA GTC CAT GTG CAC      445
Q E V H K I T N N Q S G Q V V F S E T V      132
CAG GAA GTT CAC AAG ATA ACC AAC AAC CAG AGT GGA CAG GTG GTC TTT TCT GAG ACA GTC      505
I T S V G D E E G K R S H E C I I D E D      152
ATT ACA TCT GTA GGG GAT GAA GAA GGC AAG AGG AGC CAT GAA TGT ATC ATT GAT GAA GAC      565
C G P T R Y C Q F S S F K Y T C Q P C R      172
TGT GGG CCC ACC AGG TAC TGC CAG TTC TCC AGC TTC AAG TAC ACC TGC CAG CCA TGC CCG      625
D Q Q M L C T R D S E C C G D Q L C A W      192
GAC CAG CAG ATG CTA TGC ACC CGA GAC AGT GAG TGC TGT GGA GAC CAG CTG TGT GCC TGG      685
G H C T Q K A T K G G N G T I C D N Q R      212
GGT CAC TGC ACC CAA AAG GCC ACC AAA GGT GGC AAT GGG ACC ATC TGT GAC AAC CAG AGG      745
D C Q P G L C C A F Q R G L L F P V C T      232
GAT TGC CAG CCT GGC CTG TGT TGT GCC TTC CAA AGA GGC CTG CTG TTC CCC GTG TGC ACA      805
P L P V E G E L C H D P T S Q L L D L I      252
CCC CTG CCC GTG GAG GGA GAG CTC TGC CAT GAC CCC ACC AGC CAG CTG CTG GAT CTC ATC      865
T W E L E P E G A L D R C P C A S G L L      272
ACC TGG GAA CTG GAG CCT GAA GGA GCT TTG GAC CGA TGC CCC TGC GCC AGT GGC CTC CTA      925
C Q P H S H S L V Y M C K P A F V G S H      292
TGC CAG CCA CAC AGC CAC AGT CTG GTG TAC ATG TGC AAG CCA GCC TTC GTG GGC AGC CAT      985
D H S E E S Q P R E A P D E Y E D V G      312
GAC CAC AGT GAG GAG AGC CAG CTG CCC AGG GAG GCC CCG GAT GAG TAC GAA GAT GTT GGC      1045
F I G E V R Q E L E D L E R S L A Q E M      332
TTC ATA GGG GAA GTG CCG CAG GAG CTG GAA GAC CTG GAG CCG AGC CTA GCC CAG GAG ATG      1105

```

Figura 8B

A F E G P A P V E S L G G E E I * 350
 GCA TTT GAG GGG CCT GCC CCT GIG GAG TCA GGC GGA GAG GAG GAG ATT TAG 1159
 GCCCAGACCCAGCTGAGTCACTGGTAGATGCAATAGAAATGGTAATTTATTTCCAGAGTGTCCCCNAGTGTGG 1238
 AATGGCCGACGCTCCTTCCAGTAGCTTTTCCCTTGGCTTGACAAGGTACAGTGCAGTACATTTCTTCCAGCCGCCCTG 1317
 CTTCTGTACTTGGAAAGACAGGCATGGCGGTAAAGGGCAGCGGTGAGTGCCTCCCTCGCTGTTCCTAGAAAACGCTGIC 1396
 TTGTTCTTCATGGAATGGAAGATTTGTTTGAAGGAGAGATGGAAAGGGTCAAGTCTCTCATGATGATGATTTGGGGGA 1475
 TACAGGGAGGAGGATGCCCTGGCTTGACAGCTGAGCTTGGCAAAATGTAACCTTTGCTTTTGGCTTGGCCCGCTCCCAT 1554
 GGGCTGAGGCAGTGGCTACACAAGAGCTATGCTGCTCTGTGGCCCTCCACATATTCATCCCTGTGTTTCAGCTCCTACC 1633
 TCACGTGCAGCACAGCCCTTCATAGCCACGCCCCCTCTTGGCTCACCCACAGCCTTAGGAGGGGACCAGAGGGGACTTCTCT 1712
 CAGAGCCCCATGCTCTCTCTCAACCCCAATACCAGCCCTCTGTGCCAGGACAGTCCCTTCCAAATGGAGGGAGTGAAT 1791
 CCTTTGGTTTAAATTTATTTCTCCTTCAAGGCACGCTGCCACTAAGGTCAGGCTGACTTGCATYGTCCCTCTAACGTTCCG 1870
 TAGCAGTGTGGTGGACACTGTCTTCCACCCGACTGCTTCAATACCTCTGAAAGCCAGTGTCCGGAGTGCAGTTCGTGTAA 1949
 ATTAATTTGCAGGAGTATACTTGGCTAATTTGTAGGGCTAGGATTTGTAATGAAATTTGCCAAGTCCGCTTAGCACAAT 2028
 GGAAAGCCTTTCTCAGTCAACCCGAGAGTCAACAACCAAGCCAGGTTGTGTAGAGTACAGCTGTGACATACAGACAGAA 2107
 GAAGGCTGGGCTGGATGTCAGGCCCTCAGATGACGGTTTTCAGGTGCCAGGAACTATTACCATTCTGTATCTATCCAGAGT 2186
 TATTTAAATTTGAAAGTTGCACACATTTGTATAAGCATGCCCTTCTCCTGAGTTTAAATTTATATGATACACAAACATG 2265
 TGGCCCTCAAGATCATGCAACAACCACTACTCTTTTGCTAATTTCTTGGACTTTTCTCTTTGATTTTCAATAAATACAAA 2344
 TCCCTTTCATGCAAAAAANAANAANAAGGGCGGCGCGC 2381

FIGURA 9

