

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 2000.10.06	(73) Titular(es): CORIXA CORPORATION
(30) Prioridade(s): 1999.10.07 US 158585 P	CSC THE UNITED STATES CORPORATION 2711
(43) Data de publicação do pedido: 2005.03.30	CENTERVILLE ROAD WILMINGTON DE 19808
(45) Data e BPI da concessão: 2007.02.21 005/2007	US
	(72) Inventor(es): YASIR SKEIKY US JEFFREY GUDERIAN US
	(74) Mandatário: PEDRO DA SILVA ALVES MOREIRA RUA DO PATROCÍNIO, N.º 94 1399-019 LISBOA PT

(54) Epígrafe: **SEQUÊNCIA CODIFICANTE DE MYCOBACTERIUM TURBERCULOSIS PARA EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS.**

(57) Resumo:

RESUMO

"SEQUÊNCIA CODIFICANTE DE *MYCOBACTERIUM TUBERCULOSIS* PARA EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS"

A presente invenção refere-se, de um modo geral, a sequências de ácidos nucleicos e de aminoácidos de um polipéptido de fusão compreendendo um polipéptido de *Mycobacterium tuberculosis* e um polipéptido heterólogo de interesse, vectores de expressão e células hospedeiras compreendendo esses ácidos nucleicos e métodos para produzir esses polipéptidos de fusão. Em particular, a invenção refere-se a materiais e métodos de utilização dessa sequência de *M. tuberculosis* como um parceiro de fusão para facilitar a expressão estável e rendimento elevado de polipéptidos heterólogos recombinantes tanto de origem eucariótica como procariótica.

DESCRIÇÃO

"SEQUÊNCIA CODIFICANTE DE *MYCOBACTERIUM TUBERCULOSIS* PARA EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS"

CAMPO TÉCNICO

A presente invenção refere-se, de um modo geral, a sequências de ácidos nucleicos e de aminoácidos de um polipéptido de fusão compreendendo um polipéptido de *Mycobacterium tuberculosis*, e um polipéptido heterólogo de interesse, vectores de expressão e células hospedeiras compreendendo esses ácidos nucleicos, e métodos para produzir esses polipéptidos de fusão. Em particular, a invenção refere-se a materiais e métodos de utilização dessa sequência de *M. tuberculosis* como um parceiro de fusão para facilitar a expressão estável e de produção elevada de polipéptidos heterólogos recombinantes tanto de origem eucariótica como procariótica.

ANTECEDENTES DA INVENÇÃO

O advento da tecnologia do ADN recombinante conduziu à clonagem molecular de um grande número de sequências codificantes ou genes de diversos tipos de células. De modo a estudar a função destes genes ou para produzir os produtos codificados por estas sequências, estes genes são inseridos em

vectores de expressão sob o controlo de sequências de regulação apropriadas. Esta transferência do vector de expressão para uma célula hospedeira eucariótica ou procariótica resulta geralmente na expressão do produto codificado que pode ser, subsequentemente, purificado. A produção em larga escala de muitos produtos génicos é particularmente importante nos casos em que esses produtos têm valor médico ou industrial.

Todavia, não obstante os avanços na expressão génica, certas sequências codificantes não produzem rapidamente os seus produtos de forma estável. Por exemplo, a expressão em *E. coli* de proteínas recombinantes pode ser problemática, particularmente para proteínas com domínios transmembranares ou sequências hidrofóbicas extensas. Além disso, as proteínas recombinantes podem não conter os resíduos de aminoácidos N-terminais com os enclausuramentos de códons apropriados. Assim, permanece uma necessidade para materiais e métodos melhorados para a expressão de proteínas recombinantes.

Os documentos WO99/42076 e WO99/51748 revelam o antigénio Ra12 de *M. tuberculosis* e as proteínas de fusão que o contêm.

SUMÁRIO DA INVENÇÃO

A presente invenção proporciona, pela primeira vez, métodos de utilização de moléculas de ácido nucleico recombinante que codificam polipéptidos de fusão compreendendo um polipéptido Ra12 e um polipéptido heterólogo, polipéptidos de fusão, vectores de expressão e células hospedeiras compreendendo as moléculas de ácido nucleico para produzir a expressão estável e de rendimento elevado de polipéptidos de fusão de interesse.

Num aspecto, a presente invenção proporciona um método para a expressão estável e rendimento elevado de um polinucleótido heterólogo recombinante, o método compreendendo a expressão, numa célula hospedeira, de uma molécula de ácido nucleico recombinante que codifica um polipéptido de fusão, compreendendo o polipéptido de fusão um polipéptido Ral2 e um polipéptido heterólogo, em que o polipéptido Ral2 é codificado por uma sequência do polinucleótido Ral2 que hibrida com uma sequência complementar a SEQ ID N°:3 em condições de restrição e em que a sequência do polinucleótido Ral2 está localizada a 5' da sequência do polinucleótido heterólogo que codifica o polipéptido heterólogo. Noutra forma de realização, as moléculas de ácido nucleico recombinante compreendem ainda uma sequência polinucleotídica que codifica um péptido ligando entre a sequência de polinucleótido Ral2 e a sequência de polinucleótido heteróloga, em que o péptido ligando pode compreender um sítio de clivagem. Ainda noutra forma de realização, as moléculas de ácido nucleico recombinante codificam um polipéptido de fusão compreendendo ainda uma marcação de afinidade. Ainda noutra forma de realização, as moléculas de ácido nucleico recombinantes codificam um péptido de fusão compreendendo um DPPD, um WT1, uma mamaglobina, ou um polipéptido heterólogo H9-32A. Ainda noutra forma de realização, as moléculas de ácido nucleico recombinante compreendem uma sequência de polinucleótido Ral2 compreendendo, pelo menos, cerca de 30 nucleótidos, pelo menos, cerca de 60 nucleótidos ou, pelo menos, cerca de 100 nucleótidos. Ainda noutra forma de realização, as moléculas de ácido nucleico recombinante compreendem uma sequência de polinucleótido Ral2 como apresentada na SEQ ID N°:3. Ainda noutra forma de realização, as moléculas de ácido nucleico recombinante compreendem uma sequência de

polinucleótido Ral2 que codifica um polinucleótido Ral2 como apresentado na SEQ ID N°:4, SEQ ID N°:17 ou SEQ ID N°:18.

Numa forma de realização preferida, a célula hospedeira é *E. coli*.

Numa forma de realização, o polipéptido Ral2 compreende, pelo menos, cerca de 10 aminoácidos, pelo menos, cerca de 30 aminoácidos ou, pelo menos, cerca de 100 aminoácidos. Noutra forma de realização, o polipéptido Ral2 possui uma sequência como apresentada na SEQ ID N°: 4, SEQ ID N°: 17, ou SEQ ID N°: 18.

Numa forma de realização, o método compreende ainda purificar polipéptidos de fusão após a sua expressão. Noutra forma de realização, o método compreende ainda clivar um polipéptido de fusão entre um polipéptido Ral2 e um polipéptido heterólogo.

Estes e outros aspectos da presente invenção tornar-se-ão óbvios em relação à seguinte descrição detalhada e desenhos anexos.

BREVE DESCRIÇÃO DOS DESENHOS

A Figura 1 ilustra uma sequência de nucleótidos (SEQ ID N°:1) e uma sequência de aminoácidos (SEQ ID N°:2) de MTB32A.

A Figura 2 ilustra uma sequência de nucleótidos (SEQ ID N°:3) e uma sequência de aminoácidos (SEQ ID N°:4) de Ral2.

A Figura 3 ilustra uma sequência de ácido nucleico recombinante compreendendo uma sequência de nucleótidos (SEQ ID N°:5) e uma sequência de aminoácidos (SEQ ID N°:6) do polipéptido de fusão Ral2-DPPD.

A Figura 4 ilustra uma sequência de ácido nucleico recombinante compreendendo uma sequência de nucleótidos (SEQ ID N°:7) e uma sequência de aminoácidos (SEQ ID N°:8) do polipéptido de fusão Ral2-WT1.

A Figura 5 ilustra uma sequência de ácido nucleico recombinante compreendendo uma sequência de nucleótidos (SEQ ID N°:9) e uma sequência de aminoácidos (SEQ ID N°:10) do polipéptido de fusão Ral2-mamaglobina.

A Figura 6 ilustra uma sequência de ácido nucleico recombinante compreendendo uma sequência de nucleótidos (SEQ ID N°:11) e uma sequência de aminoácidos (SEQ ID N°:12) do polipéptido de fusão Ral2-H9-32A.

A Figura 7 ilustra o polipéptido Ral2 (curto) (SEQ ID N°:17), que possui os aminoácidos 1-30 de SEQ ID N°:3.

A Figura 8 ilustra o polipéptido Ral2 (longo) (SEQ ID N°:18), que possui os aminoácidos 1-128 de SEQ ID N°:4.

A Figura 9 ilustra uma construção de polinucleótido Ral2 (curto) fundido com um gene da mamaglobina humana.

DESCRIÇÃO DETALHADA DA INVENÇÃO

Como referido acima, a presente invenção proporciona, pela primeira vez, métodos para utilizar sequências de Ra12 como um parceiro de fusão, para facilitar a expressão estável e elevado rendimento de polipéptidos heterólogos recombinantes, tanto de origem eucariótica como procariótica.

MTB32A é uma protease da serina de 32 KD de peso molecular codificada por um gene em estirpes virulentas e não virulentas de *M. tuberculosis*. A sequência de nucleótidos (SEQ ID N°:1) e a sequência de aminoácidos (SEQ ID N°:2) completas de MTB32A são reveladas na Figura 1. Ver, também, Skeiky et al., *Infection and Immun.* (1999) 67:3998-4007. Esta proteína é segregada naturalmente para o sobrenadante das culturas bacterianas. A grelha de leitura aberta da sequência codificante contém sinais de secreção hidrofóbica N-terminais. Estimula as células mononucleares do sangue periférico de doadores saudáveis positivos para derivado de proteína purificada (PPD), para proliferar e segregar interferão. Assim, MTB32A é um antígeno candidato para utilização no desenvolvimento de vacinas contra a tuberculose.

Surpreendentemente, foi descoberto pelos presentes requerentes que um fragmento C-terminal de 14 KD da sequência codificante de MTB32A se expressa a níveis elevados por si só e permanece como uma proteína solúvel ao longo do processo de purificação. Este fragmento C-terminal de 14 KD de MTB32A é aqui referido como Ra12 (possuindo os resíduos de aminoácidos 192 a 323 de MTB32A). As sequências de ácidos nucleicos e de aminoácidos de Ra12 nativo são apresentadas, e. g., nas Figuras 2-6. Como descrito em detalhe abaixo, o termo "polipéptido Ra12"

ou "polinucleótido Ral2" como aqui utilizado refere-se às sequências de Ral2 nativo (e. g., SEQ ID N°:3 ou SEQ ID N°:4), suas variantes, ou seus fragmentos (e. g., SEQ ID N°:17 ou SEQ ID N°:18). A presente invenção utiliza estas propriedades do polipéptidos Ral2 e proporciona moléculas de ácido nucleico recombinante, vectores de expressão, células hospedeiras, e métodos para expressão estável e de rendimento elevado de polipéptidos de fusão compreendendo um polipéptido Ral2 e um polipéptido heterólogo de interesse. Os métodos da presente invenção são particularmente úteis na expressão de certos polipéptidos heterólogos (e. g., DPPD) que outros métodos de expressão convencionais não conseguem expressar numa quantidade substancial.

Ácidos nucleicos de Fusão Recombinantes

Ácidos nucleicos recombinantes, que codificam um polipéptido de fusão compreendendo um polipéptido Ral2 e um polipéptido heterólogo de interesse, podem ser rapidamente construídos através de técnicas de engenharia genética convencionais. Os ácidos nucleicos recombinantes são construídos de modo a que a sequência de polinucleótido Ral2 esteja localizada a 5' em relação a uma sequência de polinucleótidos heteróloga seleccionada.

Na presente invenção, qualquer polinucleótido heterólogo adequado de interesse pode ser seleccionado como um parceiro de fusão para ácidos nucleicos de Ral2 para produzir um polipéptido de fusão. Uma "sequência heteróloga" ou um "ácido nucleico heterólogo", como aqui utilizado, é um que tem origem numa fonte estranha para a célula hospedeira particular, ou, se for da

mesma fonte, é modificado da sua forma original. Assim, um ácido nucleico heterólogo numa célula hospedeira procariótica inclui um ácido nucleico heterólogo que é endógeno para a célula hospedeira particular que foi modificada. A modificação da sequência heteróloga pode ocorrer, e. g., através do tratamento do ADN com uma enzima de restrição, para criar um fragmento de ADN que é capaz de ser ligado operacionalmente ao promotor. Técnicas, tais como mutagenese dirigida também são úteis na modificação de uma sequência heteróloga.

Um ácido nucleico heterólogo tanto de origem eucariótica como procariótica pode ser seleccionado como um parceiro de fusão. Estes ácidos nucleicos incluem, mas não estão limitados a, ácidos nucleicos que codificam antigénios patogénicos, antigénios bacterianos, antigénios virais, antigénios de cancro, antigénios tumorais e supressores tumorais. Exemplos de ácidos nucleicos heterólogos de interesse incluem ácidos nucleicos de DPPD, WT1, mamaglobina, H9-32A e outros ácidos nucleicos de *Mycobacterium tuberculosis* (ver, e. g., Cole et al. *Nature* (1999) 393:537-544; <http://www.sanger.ac.uk>; e <http://www.pasteur.fr/mycdb/> para as sequências do genoma completo de *M. tuberculosis*; ver, também os documentos WO98/53075 e WO98/53076, ambos publicados em 26 de Novembro de 1998 para as sequências de ácido nucleico que codificam proteínas de *M. tuberculosis*). Qualquer um dos ácidos nucleicos aqui revelados pode ser utilizado isoladamente ou em combinação como um ácido nucleico heterólogo que pode ser seleccionado como um parceiro de fusão.

Adicionalmente, qualquer polinucleótido de Ra12 adequado (e. g., polinucleótido Ra12 nativo possuindo SEQ ID N°:3, variantes ou fragmentes deste) pode ser utilizado na construção

de ácidos nucleicos de fusão recombinantes da presente invenção. Polinucleótidos de Ral2 preferidos compreendem, pelo menos, cerca de 15 nucleótidos consecutivos, pelo menos, cerca de 30 nucleótidos, pelo menos, cerca de 60 nucleótidos, pelo menos, cerca de 100 nucleótidos, pelo menos, cerca de 200 nucleótidos, ou pelo menos, cerca de 300 nucleótidos. Os polinucleótidos podem ser de cadeia simples ou cadeia dupla, e podem ser moléculas de ADN (genómico, ADNc ou sintético) ou de ARN.

Numa forma de realização, a sequência de polinucleótido Ral2 é como apresentada em SEQ ID N°:3. Noutra forma de realização, a sequência de polinucleótido Ral2 codifica um polipéptido Ral2 como apresentado em SEQ ID N°:4. Em algumas formas de realização, a sequência de polinucleótido Ral2 compreende uma porção de SEQ ID N°:3 ou codifica uma porção de SEQ ID N°:4. Por exemplo, um polinucleótido Ral2 compreendendo 90 nucleótidos (e. g., nucleótidos 1-90 de SEQ ID N°:3), ou um polinucleótido Ral2 compreendendo 384 nucleótidos (e. g., nucleótidos 1-384 de SEQ ID N°:3) pode ser utilizado como um parceiro de fusão. Ver Exemplos 2 e 3 abaixo.

Os polinucleótidos podem compreender uma sequência nativa (i. e., uma sequência endógena que codifica um polipéptido Ral2 SEQ ID N°:3, ou uma porção desta) ou pode compreender uma variante dessa sequência. As variantes do polinucleótido podem conter uma ou mais substituições, adições, deleções e/ou inserções, de modo a que a actividade biológica do polipéptido de fusão codificado não seja diminuída em relação a um polipéptido de fusão compreendendo um polipéptido Ral2 nativo. As variantes, de um modo preferido, apresentam, pelo menos, cerca de 70% de identidade, de um modo mais preferido, pelo menos, cerca de 80% de identidade e de um modo muito preferido,

pelo menos, cerca de 90% de identidade com uma sequência polinucleotídica que codifica um polipéptido Ral2 nativo (SEQ ID N°:4) ou uma porção deste. Opcionalmente, a identidade existe ao longo de uma região que é pelo menos, cerca de 25 até cerca de 50 aminoácidos ou nucleótidos de comprimento ou, opcionalmente, ao longo de uma região que tem 75-100 aminoácidos ou nucleótidos de comprimento.

Duas sequências de polinucleótido ou polipéptido dizem-se "idênticas" se a sequência de nucleótidos ou aminoácidos nas duas sequências é a mesma quando alinhadas para o máximo de correspondência, como descrito abaixo. As comparações entre as duas sequências são tipicamente realizadas comparando as sequências numa janela de comparação para identificar e comparar regiões locais de semelhança de sequências. Uma "janela de comparação", como aqui utilizado, refere-se a um segmento de, pelo menos, cerca de 20 posições contíguas, normalmente 30 até cerca de 75, e. g. 40 até cerca de 50, no qual a sequência pode ser comparada com uma sequência de referência do mesmo número de posições contíguas após as duas sequências serem alinhadas para um ótimo de identidade.

O alinhamento ótimo de sequências para comparação pode ser realizado utilizando o programa Megalign no conjunto de software bioinformático Lasergene (DNASTAR, Inc., Madison, WI), utilizando os parâmetros por defeito. Este programa realiza vários esquemas de alinhamento descritos nas referências seguintes: Dayhoff, M.O. (1978) *A model of evolutionary change in proteins - Matrices for detecting distant relationships*. Em Dayhoff, M.O. (ed.) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington DC Vol. 5, Supl. 3, pp. 345-358; Hein J. (1990) *Unified Approach to*

Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. e Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. e Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. e Sokal, R.R. (1973) *Numerical Taxonomy - the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. e Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA* 80:726-730.

Alternativamente, o alinhamento óptimo de sequências para comparação pode ser realizado através do algoritmo de identidade local de Smith e Waterman (1981) *Add. APL. Math* 2:482, através do algoritmo de alinhamento de identidade de Needleman e Wunsch (1970) *J. Mol. Biol.* 48:443, através dos métodos de pesquisa de semelhança de Pearson e Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85:2444, através de implementações computadorizadas destes algoritmos (GAP, BESTFIT, BLAST, FASTA, e TFASTA no Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), ou por inspecção.

Exemplos preferidos de algoritmos que são adequados para determinar a percentagem de identidade de sequência e semelhança de sequências são os algoritmos BLAST e BLAST 2.0, que são descritos em Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 e Altschul et al. (1990) *J Mol. Biol.* 215:403-410, respectivamente. Podem ser utilizados BLAST e BLAST 2.0, por exemplo com os parâmetros aqui descritos, para determinar a percentagem da identidade da sequência para os polinucleótidos e polipéptidos da invenção. O software para realizar as análises BLAST está disponível ao público através do National Center for Biotechnology Information: Para as sequências de aminoácidos,

pode ser utilizada uma matriz de classificação para calcular a classificação cumulativa. A extensão do número de semelhanças encontradas em cada direcção são interrompidas quando: o valor de alinhamento cumulativo cai fora pela quantidade X do seu valor máximo alcançado; o valor cumulativo vai para zero ou abaixo, devido à acumulação de um ou mais alinhamentos de resíduos com classificações negativas; ou é atingido o fim de qualquer uma das sequências. Os parâmetros W, T e X do algoritmo BLAST determinam a sensibilidade e a velocidade do alinhamento.

Numa abordagem preferida, a "percentagem de identidade de sequência" é determinada comparando duas sequências optimamente alinhadas ao longo de uma janela de comparação de pelo menos, 20 posições, em que a porção da sequência de polipéptido na janela de comparação pode compreender adições ou deleções (*i. e.*, intervalos) de 20 por cento ou menos, normalmente 5 a 15 por cento, ou 10 a 12 por cento, em comparação com as sequências de referência (que não compreendem adições ou deleções) para alinhamento óptimo das duas sequências. A percentagem é calculada determinando o número de posições nas quais ocorre o resíduo de aminoácido idêntico em ambas as sequências para produzir o número de posições emparelhadas, dividindo o número de posições emparelhadas pelo número total de posições na sequência de referência (*i. e.*, o tamanho da janela) e multiplicando os resultados por 100 para produzir a percentagem de identidade de sequências.

As variantes também podem, ou alternativamente, ser substancialmente homólogas a um polinucleótido Ral2 nativo (*e. g.*, SEQ ID N°:3), ou uma porção ou complemento deste. Essas variantes de polinucleótido são capazes de hibridar em condições de restringência com uma sequência de ADN que ocorre

naturalmente codificando um polinucleótido Ra12 nativo (ou uma sequência complementar).

A frase "hibrida selectivamente (ou especificamente) com" refere-se à ligação, formação de duplexos, ou hibridação de uma molécula apenas com uma sequência de nucleótidos particular em condições de hibridação de restringência quando essa sequência está presente numa mistura complexa (e. g., ADN ou ARN celular total ou biblioteca).

A frase "condições de hibridação de restringência" refere-se a condições sob as quais uma sonda hibridará com a sua subsequência alvo, tipicamente numa mistura complexa de ácido nucleico, mas não com outras sequências. As condições restritivas são dependentes da sequência e serão diferentes em diferentes circunstâncias. Sequências mais longas hibridam especificamente a temperaturas mais elevadas. Um guia extensivo para a hibridação de ácidos nucleicos encontra-se em Tijssen, *Techniques in Biochemistry and Molecular Biology---Hybridization with Nucleic Probes*, "Overview of principles of Hybridization and the strategy of nucleic acid assays" (1993). Geralmente, as condições de restringência são seleccionadas para serem cerca de 5-10 °C mais baixas do que o ponto de fusão térmica (T_m) para a sequência específica a uma força iónica de pH definida. A T_m é a temperatura (sob força iónica definida, pH, e concentração nucleica) na qual 50% das sondas complementares ao alvo hibridam com a sequência alvo no equilíbrio (como as sequências alvo estão presentes em excesso, a T_m , 50% das sondas estão ocupadas no equilíbrio). As condições de restringência serão aquelas em que a concentração de sal é inferior a cerca de 1,0 M de ião sódio, tipicamente cerca de 0,01 a 1,0 M de concentração de ião sódio (ou outros sais) a pH 7,0 até 8,3 e a temperatura é pelo

menos, cerca de 30°C para sondas curtas (e. g., 10 até 50 nucleótidos) e pelo menos, cerca de 60 °C para sondas longas (e. g., superior a 50 nucleótidos). As condições de restringência também podem ser alcançadas com a adição de agentes desestabilizantes, tais como formamida. Para hibridação selectiva ou específica, um sinal positivo é, pelo menos, duas vezes o fundo, de um modo preferido 10 vezes o fundo da hibridação. Condições de hibridação de restringência exemplares podem ser como se segue: 50% de formamida, 5x SSC, e 1% SDS, incubando a 42 °C, ou, 5x SSC, 1% SDS, incubando a 65 °C, com uma lavagem em 0,2x SSC, e 0,1% SDS a 65 °C.

Deve ser entendido pelos especialistas na técnica que, como resultado da degenerescência do código genético, existem muitas sequências de nucleótidos que codificam um polipeptido Ral2 como aqui descrito. Alguns destes polinucleótidos possuem homologia mínima com a sequência de nucleótidos de qualquer gene nativo. No entanto, os polinucleótidos que variam devido às diferenças na utilização dos códons estão especificamente contemplados pela presente invenção. Além disso, os alelos dos genes que compreendem as sequências de polinucleótido aqui fornecidas estão dentro do âmbito da presente invenção. Os alelos são genes endógenos que são alterados como resultado de uma ou mais mutações, tais como deleções, adições e/ou substituições de nucleótidos. O ARNm e a proteína resultantes podem, mas não necessariamente, possuir uma estrutura ou função alterada. Os alelos podem ser identificados utilizando técnicas convencionais (tais como hibridação, amplificação e/ou comparação de sequências com as bases de dados).

Assim, os termos tais como "polinucleótido Ral2" ou "sequência de polinucleótido", como aqui utilizados, referem-se

a sequências de polinucleótido Ral2 nativo (e. g., SEQ ID N°:3), fragmentos deste, ou quaisquer variantes deste. Funcionalmente, qualquer polinucleótido Ral2 possui a capacidade para produzir uma proteína de fusão, e a sua capacidade para produzir uma proteína de fusão em células hospedeiras pode ser aumentada ou não carregada, em relação ao polinucleótido Ral2 nativo (e. g., SEQ ID N°:3), ou pode ser diminuída em menos de 50%, e de um modo preferido menos de 20%, em relação ao polinucleótido Ral2 nativo.

Ácidos nucleicos que codificam polipéptidos Ral2 desta invenção podem ser preparados por qualquer método adequado conhecido na técnica. Métodos exemplares incluem clonagem e restrição de sequências apropriadas ou síntese química directa através de métodos tais como o método de fosfotriéster de Narang et al. (1979) *Meth. Enzymol.* 68:90-99; o método de fosfodiéster de Brown et al. (1979) *Meth. Enzymol.* 68:109-151; o método da dietilfosforamidite de Beaucage et al. (1981) *Tetra. Lett.*, 22:1859-1862; e o método do suporte sólido da Patente U.S. N°. 4 458 066.

Numa forma de realização, um ácido nucleico codificando MTB32A ou Ral2 é isolado através de métodos de clonagem de rotina. As sequências de nucleótidos de MTB32A ou Ral2, como aqui proporcionadas, são utilizadas para proporcionar sondas que hibridam especificamente com outros ácidos nucleicos de MTB32A ou Ral2 numa amostra de ADN genómico, ou com um ARNm de MTB32A ou ARNm de Ral2 numa amostra de ARN total (e. g., numa transferência de Southern ou de Northern). Assim que os ácidos nucleicos alvo de MTB32A ou Ral2 estejam identificados, podem ser isolados de acordo com métodos convencionais conhecidos dos especialistas na técnica.

Os ácidos nucleicos desejados também podem ser clonados utilizando técnicas de amplificação bem conhecidas. Exemplos de protocolos suficientes para guiar pessoas especialistas através de métodos de amplificação *in vitro*, incluindo a reacção da polimerase em cadeia (PCR) a reacção da ligase em cadeia (LCR), amplificação por replicase Q β e outras técnicas mediadas pela polimerase de ARN encontram-se em Berger, Sambrook, e Ausubel, assim como em Mullis *et al.* (1987) Patente U.S. N° 4 683 202; *PCR Protocols A Guide to Methods and Applications* (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (1 de Outubro de 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3:81- 94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874; Lomell *et al.* (1989) *J. Clin. Chem.* 35:1826; Landegren *et al.* (1988) *Science* 241:1077-1080; Van Brunt (1990) *Biotechnology* 8:291-294; Wu e Wallace (1989) *Gene* 4:560; e Barringer *et al.* (1990) *Gene* 89:117. Métodos melhorados de clonagem de ácidos nucleicos amplificados *in vitro* são descritos em Wallace *et al.*, Patente U.S. N° 5 426 039. Iniciadores adequados para utilização na amplificação dos ácidos nucleicos da invenção podem ser concebidos com base nas sequências aqui fornecidas.

Os ácidos nucleicos de MTB32A ou Ral2 também podem ser clonados detectando o seu produto expresso através de ensaios baseados nas propriedades físicas, químicas, ou imunológicas da proteína expressa. Por exemplo, pode identificar-se um ácido nucleico de MTB32A ou Ral2 clonado através da capacidade de um polipéptido codificado pelo ácido nucleico para se ligar a anti-soros ou anticorpos purificados produzidos contra os polipéptidos MTB32A ou Ral2s aqui fornecidos, que também

reconhece e se liga selectivamente aos homólogos de MTB32A ou Ral2.

Em certas formas de realização, pode ser desejável modificar os ácidos nucleicos de MTB32A ou Ral2 da invenção. Sequências de nucleótidos alteradas que podem ser utilizadas de acordo com a invenção incluem deleções, adições ou substituições de diferentes resíduos de nucleótidos resultando numa sequência que codifica o mesmo produto génico ou um funcionalmente equivalente. O próprio produto génico pode conter deleções, adições ou substituições de resíduos de aminoácidos, que resultam numa alteração silenciosa, produzindo assim um epitopo antigénico funcionalmente equivalente. Essas substituições de aminoácidos conservativas podem ser realizadas com base na semelhança em polaridade, carga, solubilidade, hidrofobicidade, hidrofiliabilidade, e/ou a natureza anfipática dos resíduos envolvidos. De um modo preferido, podem ser utilizados os ácidos nucleicos de Ral2 que são mais curtos em comprimento do que SEQ ID N°:3 que codificam o parceiro de fusão biologicamente activo. Esses equivalentes funcionais mais pequenos de polipéptidos Ral2 podem ser desejáveis para aumentar a quantidade de recursos de célula hospedeira que estão disponíveis para a produção de polipéptidos heterólogo de interesse.

Um especialista reconhecerá muitas maneiras de criar alterações numa dada construção de ácido nucleico. Esses métodos bem conhecidos incluem mutagénese dirigida, amplificação por PCR utilizando oligonucleótidos degenerados, exposição de células contendo o ácido nucleico a agentes mutagénicos ou radiação, síntese química de um oligonucleótido desejado (e. g., em conjugação com ligação e/ou clonagem para criar ácidos nucleicos grandes) e outras técnicas bem conhecidas. Ver, e. g., Gilman e

Smith (1979) *Gene* 8:81-97, Roberts et al., (1987) *Nature* 328:731-734.

Os ácidos nucleicos recombinantes que codificam um polipéptido de fusão compreendendo um polipéptido Ral2 e um polipéptido heterólogo seleccionado podem ser preparados utilizando quaisquer métodos conhecidos na técnica. Como descrito acima, os ácidos nucleicos recombinantes são construídos de modo a que uma sequência de polinucleótido Ral2 esteja localizada a 5' de uma sequência de polinucleótido heterólogo seleccionado. As sequências de Ral2 e polinucleótido heterólogo também podem ser modificadas para facilitar a sua fusão e subsequente expressão de polipéptidos de fusão. Por exemplo, o codão de terminação a 3' da sequência de polinucleótido Ral2 pode ser substituído com uma sequência de ligando em grelha, que pode proporcionar sítios de restrição e/ou sítios de clivagem. Os ácidos nucleicos recombinantes podem ainda compreender outras sequências de nucleótidos, tais como sequências que codificam etiquetas de afinidade para facilitar o protocolo de purificação da proteína.

Vectores de Expressão e Células Hospedeiras

Os ácidos nucleicos recombinantes, como aqui descrito podem ser unidos a uma variedade de outras sequências de nucleótidos, utilizando técnicas de ADN recombinante estabelecidas. Por exemplo, um polinucleótido pode ser clonado em qualquer um de uma variedade de vectores de clonagem, incluindo plasmídeos, fagémídeos, derivados de fago lambda e cosmídeos. Vectores de particular interesse incluem vectores de expressão, vectores de replicação, vectores de criação de sonda e vectores de

sequenciação. Em geral, um vector conterá uma origem de replicação funcional em pelo menos, um organismo, sítios de endonuclease de restrição convenientes e um ou mais marcadores de selecção. Outros elementos dependerão da utilização desejada, e serão óbvios para os especialistas na técnica.

As sequências de ADN que codificam os componentes polipeptídicos podem ser montados separadamente, e ligados num vector de expressão apropriado. A extremidade 3' da sequência de ADN que codifica um componente polipeptídico é ligado, com ou sem uma sequência polinucleotídica codificando um ligando peptídico, à extremidade 5' de uma sequência de ADN codificando o segundo componente polipeptídico de modo a que as grelhas de leitura das sequências estejam em fase. Isto permite a tradução numa única proteína de fusão que retém a actividade biológica de ambos os componentes polipeptídicos.

As sequências de ADN ligadas são ligadas operacionalmente a elementos de regulação de transcrição ou tradução adequados. Os elementos de regulação responsáveis pela expressão de ADN estão localizados apenas a 5' da sequência de ADN codificando os primeiros polipéptidos. De um modo semelhante, os codões de terminação necessários para terminar a tradução e os sinais de terminação de transcrição estão apenas presentes a 3' da sequência de ADN codificando o segundo polipéptido.

Dependendo do sistema hospedeiro/vector utilizado, qualquer um de um número de elementos adequados de transcrição e tradução, incluindo promotores constitutivos e indutíveis, pode ser utilizado no vector de expressão. Por exemplo, quando se clona em sistemas bacterianos, podem ser utilizados promotores indutíveis, tais como pL de bacteriófago λ , plac, ptrp, ptac

(promotor híbrido ptrp-lac; promotor de citomegalovírus) e semelhantes; quando se clona em sistemas de células de leveduras, podem ser utilizados promotores tais como ADHI, PGK, PHO5, ou o promotor do factor α ; quando se clona em sistemas de células de insecto, podem ser utilizados promotores tais como o promotor de poliedro de baculovírus; quando se clona em sistemas de células vegetais, podem ser utilizados promotores derivados do genoma de células vegetais (e. g., promotores de choque térmico; o promotor para a subunidade pequena da RUBISCO; o promotor para a proteína de ligação a clorofila α/β) ou de vírus de plantas (e. g., o promotor do ARN 35S de CaMV; o promotor da proteína da capa de TMV); quando se clona em sistemas de células de mamíferos, podem ser utilizados promotores derivados do genoma de células de mamíferos (e. g., promotor da metalotioneína) ou de vírus de mamíferos (e. g., o promotor tardio de adenovírus; o promotor do vírus vaccinia de 7,5K); quando se criam linhas celulares que contêm cópias múltiplas de uma sequência de antígeno codificante, podem ser utilizados vectores à base de SV40-, BPV- e EBV- com um marcador de selecção apropriado.

Pode ser utilizada uma variedade de sistemas de hospedeiro-vector de expressão para expressar sequências codificantes da proteína de fusão Ral2. Estes incluem, mas não estão limitados a, microrganismos tais como bactérias (e. g., *E. coli*, *B. subtilis*) transformados com vectores de expressão de ADN de bacteriófago recombinante, de ADN de plasmídeo ou de ADN de cosmídeo contendo uma sequência codificante; leveduras (e. g., *Saccharomyces*, *Pichia*) transformados com vectores de expressão de levedura recombinantes contendo uma sequência codificante; sistemas de células de insectos infectadas com vectores de expressão de vírus recombinantes (e. g., baculovírus) contendo

uma sequência codificante; sistemas de células vegetais infectadas com vectores de expressão de vírus recombinantes (e. g., vírus do mosaico da couve-flor, CaMV; vírus do mosaico do tabaco, TMV) ou transformadas com vectores de expressão de plasmídeo recombinantes (e. g., plasmídeo Ti) contendo uma sequência codificante; ou sistemas de células de mamífero (e. g., células COS, CHO, BHK, 293, 3T3). Os elementos de expressão destes sistemas variam na sua força e especificidades.

São preferidos os sistemas bacterianos para a expressão de polipéptidos de fusão Ral2. Sequências de controlo procarióticas normalmente utilizadas, que são aqui definidas para incluir promotores para iniciação de transcrição, opcionalmente com um operador, juntamente com sequências de sítio de ligação a ribossoma, incluem os promotores normalmente utilizados, tais como os sistemas de promotor da beta-lactamase (penicilinase) e da lactose (*lac*) (Change et al., *Nature* (1977) 198:1056), o sistema de promotor do triptofano (*trp*) (Goeddel et al., *Nucleic Acids Res.* (1980) 8:4057), o promotor *tac* (DeBoer et al., *Proc. Natl. Acad. Sci. U.S.A.* (1983) 80:21-25); e o promotor P_L derivado de lambda e sítio de ligação ao ribossoma do gene N (Shimatake et al., *Nature* (1981) 292:128). O sistema particular de promotor não é crítico para a invenção; pode ser utilizado qualquer promotor disponível que funcione em procariotas.

Podem ser utilizados na presente invenção promotores constitutivos ou regulados. Os promotores regulados podem ser vantajosos porque as células hospedeiras podem ser cultivadas em densidades elevadas antes da expressão dos polipéptidos de fusão Ral2 ser induzida. A expressão de proteínas heterólogas em níveis elevados retarda o crescimento celular em algumas situações. Promotores regulados especialmente adequados para

utilização em *E. coli* incluem o promotor P_L do bacteriófago lambda, o promotor híbrido *trp-lac* (Amann et al., *Gene* (1983) 25:167; de Boer et al., *Proc. Natl. Acad. Sci. USA* (1983) 80:21, e o promotor do bacteriófago T7 (Studier et al., *J Mol. Biol.* (1986); Tabor et al., (1985). Estes promotores e a sua utilização são discutidos em Sambrook et al., (1989) *Molecular Cloning: A Laboratory Manual*, 2ª Ed., Vols. 1-3, Cold Spring Harbor Laboratory.

Para a expressão de polipéptidos de fusão Ral2 em células procarióticas diferentes de *E. coli*, é necessário um promotor que funcione na espécie procariótica particular. Esses promotores podem ser obtidos dos genes que foram clonados da espécie, ou podem ser utilizados promotores heterólogos. Por exemplo, o promotor híbrido *trp-lac* funciona em *Bacillus* adicionalmente a *E. coli*.

Um sítio de ligação ao ribossoma (RBS) é incluído convenientemente nas cassetes de expressão da invenção. Um RBS em *E. coli*, por exemplo, consiste numa sequência de nucleótidos de 3-9 nucleótidos de comprimento localizada 3-11 nucleótidos a montante do codão de iniciação (Shine e Dalgarno, *Nature* (1975) 254:34; Steitz, *In Biological regulation and development: Gene expression* (ed. R. F. Goldberger), vol. 1, p. 349, 1979, Plenum Publishing, NY).

Quando se querem produzir grandes quantidades da proteína de fusão Ral2, podem ser desejáveis os vectores que dirigem a expressão de níveis elevados de produtos de proteína de fusão que são rapidamente purificados. Esses vectores incluem, mas não estão limitados a, o vector de expressão de *E. coli* pUR278 (Ruther et al. (1983) *EMBO J.* 2:1791), no qual uma sequência

codificante pode ser ligada ao vector em grelha com a região codificante de *lacZ* de modo a que seja produzida uma proteína híbrida; vectores pIN (Inouye e Inouye (1985) *Nucleic Acids Res.* 13:3101-3109; Van Heeke e Schuster (1989) *J. Biol. Chem.* 264:5503-5509); e semelhantes. Os vectores pGEX também podem ser utilizados para expressar polipéptidos estranhos como proteínas de fusão com glutathione S-transferase (GST). Em geral, essas proteínas de fusão são solúveis e podem ser purificados facilmente de células lisadas por adsorção a esferas de glutathione-agarose seguido por eluição na presença de glutathione livre. Para certas aplicações, pode ser desejável clivar o polipéptido heterólogo de interesse do polipéptido de fusão Ral2 após purificação. Isto pode ser realizado por qualquer um de vários métodos conhecidos na técnica. Por exemplo, os vectores pGEX são concebidos para incluir sítios de clivagem da trombina ou protease de factor Xa de modo a que o polipéptido de fusão clonado de interesse possa ser libertado da unidade de GST. Ver, e. g., Sambrook et al., *supra.*; Itakura et al., *Science* (1977) 198:1056; Goeddel et al., *Proc. Natl. Acad. Sci. USA* (1979) 76:106; Nagai et al., *Nature* (1984) 309:810; Sung et al., *Proc. Natl. Acad. Sci. USA* (1986) 83:561. Os sítios de clivagem podem ser construídos nos ácidos nucleicos recombinantes para as proteínas de fusão no ponto de clivagem desejado.

Polipéptidos de fusão

No contexto da presente invenção, um polipéptido de "fusão" compreende pelo menos, duas partes: um polipéptido Ral2, como aqui descrito, e um polipéptido heterólogo de interesse. Num polipéptido de fusão, um polipéptido Ral2 é, de um modo preferido, fundido, directamente ou indirectamente, ao terminal

amino de um polipéptido heterólogo de interesse, embora a fusão ao terminal carboxilo do polipéptido heterólogo ou inserção do polipéptido heterólogo num sítio dentro de um polipéptido Ral2 também possa ser apropriado.

Qualquer polipéptido heterólogo de interesse, seja de origem eucariótica ou procariótica, pode ser seleccionado como um parceiro de fusão a um polipéptido Ral2. Estes polipéptidos heterólogos incluem, mas não estão limitados a, antigénios patogénicos, antigénios bacterianos, antigénios virais, antigénios de cancro, antigénios tumorais, e supressores tumorais. Polipéptidos heterólogos exemplares incluem polipéptidos DPPD, Wt1, mamaglobina, H9-32A, ou outras proteínas de *M. tuberculosis*. Qualquer um destes polipéptidos pode ser utilizado isoladamente ou em combinação como um polipéptido heterólogo que pode ser seleccionado como um parceiro de fusão.

Como referido acima, um polipéptido de fusão pode compreender um polipéptido Ral2 nativo (e. g., SEQ ID N°:4), uma variante deste, ou um fragmento deste. Uma "variante" de polipéptido, como aqui utilizado, é um polipéptido que difere de um polipéptido Ral2 nativo em uma ou mais substituições, deleções, adições e/ou inserções, tais como a actividade biológica do polipéptido não é substancialmente diminuída. Por outras palavras, a capacidade de uma variante para produzir polipéptido de fusão em células hospedeiras pode ser aumentada ou não alterada, em relação à proteína Ral2 nativa, ou pode ser diminuída em menos de 50%, e de um modo preferido menos do que 20%, em relação à proteína Ral2 nativa. Essas variantes podem geralmente ser identificadas modificando uma das sequências de polipéptido acima e avaliando o nível de produção de polipéptido de fusão em células hospedeiras, tais como em *E. coli*. Variantes

exemplares incluem aquelas em que uma pequena porção (e. g., 1-30 aminoácidos, de um modo preferido 5-15 aminoácidos) foi removida do terminal N e/ou C dos polipéptidos Ral2 nativos. Numa forma de realização, as variantes do polipéptido Ral2 nativos compreendem pelo menos, cerca de 5 aminoácidos, pelo menos, cerca de 10 aminoácidos, pelo menos, cerca de 30 aminoácidos, pelo menos, cerca de 50 aminoácidos, ou pelo menos, cerca de 100 aminoácidos.

Numa forma de realização, a sequência de polipéptido Ral2 é como apresentada na SEQ ID N°:4. Noutra forma de realização, a sequência do polipéptido Ral2 compreende uma porção da SEQ ID N°:4. Por exemplo, um polipéptido Ral2 compreendendo 30 aminoácidos (e. g., aminoácidos 1-30 de SEQ ID N°:4) ou um polipéptido Ral2 compreendendo 128 aminoácidos (e. g., aminoácidos 1-128 de SEQ ID N°:4) podem ser utilizados como um parceiro de fusão. Ver Exemplos 2 e 3 abaixo.

As variantes de polipéptido apresentam, de um modo preferido, pelo menos, cerca de 70%, de um modo mais preferido, pelo menos, cerca de 80% ou, pelo menos, cerca de 90%, e de um modo muito preferido, pelo menos, cerca de 95% de identidade (determinado como descrito acima) com os polipéptidos identificados. Opcionalmente, a identidade existe numa região que é, pelo menos, cerca de 20 até cerca de 50 aminoácidos de comprimento ou, opcionalmente numa região que tem 75-100 aminoácidos de comprimento.

De um modo preferido, uma variante contém substituições conservativas. Uma "substituição conservativa" é uma em que um aminoácido é substituído por outro aminoácido que possui propriedades semelhantes, de modo a que um especialista na

técnica de química de péptidos esperaria que a estrutura secundária e a natureza hidropática do polipéptido fosse substancialmente não alterada. As substituições de aminoácido podem ser geralmente realizadas com base na semelhança em polaridade, carga, solubilidade, hidrofobicidade, hidrofiliabilidade e/ou a natureza anfipática dos resíduos. Por exemplo, aminoácidos carregados negativamente incluem ácido aspártico e ácido glutâmico; aminoácidos carregados positivamente incluem lisina e arginina; e aminoácidos com grupos de cabeça polar não carregados possuindo valores de hidrofiliabilidade semelhantes incluem leucina, isoleucina e valina; glicina e alanina; asparagina e glutamina; e serina, treonina, fenilalanina e tirosina. Outros grupos de aminoácidos que podem representar cargas conservativas incluem: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; e (5) phe, tyr, trp, his. Uma variante pode também, ou alternativamente, conter alterações não conservativas. Numa forma preferida de realização, os polipéptidos variantes diferem de uma sequência nativa por substituição, deleção ou adição de cinco aminoácidos ou menos. As variantes podem também (ou alternativamente) ser modificados por, por exemplo, a deleção ou adição de aminoácidos que possuem influência mínima na imunogenicidade, estrutura secundária e natureza hidropática do polipéptido.

Assim, os termos, tais como "polipéptido Ral2" ou "sequência de polipéptido Ral2", como aqui utilizados, referem-se a sequências de polinucleótido Ral2 nativos (e. g., SEQ ID N°:4), fragmentos destes (e. g., SEQ ID N°:17 ou 18), ou quaisquer variantes destes. Funcionalmente, um polipéptido Ral2 possui a capacidade para produzir uma proteína de fusão, e a sua capacidade para produzir uma proteína de fusão nas células

hospedeiras pode ser aumentada ou não carregada, em relação ao polipéptido Ral2 nativo (e. g., SEQ ID N°:4), ou pode ser diminuído em menos de 50%, e de um modo preferido menos de 20%, em relação ao polipéptido Ral2 nativo.

Como referido acima, os polipéptidos de fusão podem ser conjugados a um ligando ou outra sequência para facilidade de síntese, purificação ou identificação do polipéptido ou para aumentar a ligação do polipéptido a um suporte sólido. Por exemplo, uma sequência de ligando peptídico pode ser empregue para separar um polipéptido Ral2 e um polipéptido heterólogo de interesse por uma distância suficiente para assegurar que cada polipéptido se dobra nas suas estruturas secundárias e terciárias. Uma tal sequência de ligando peptídico é incorporada na proteína de fusão utilizando técnicas convencionais bem conhecidas na técnica. Sequências de ligando peptídico adequadas podem ser escolhidas com base nos seguintes factores: (1) a sua capacidade para adoptar uma conformação flexível estendida; (2) a sua incapacidade para adoptar uma estrutura secundária que possa interactuar com epitopos funcionais no primeiro e segundo polipéptidos; e (3) a ausência de resíduos hidrofóbicos ou carregados que possam reagir com os epitopos funcionais do polipéptido. Em certas formas de realização, as sequências de ligando peptídico podem conter resíduos Gly, Asn e Ser. Outros aminoácidos próximos do neutro, tais como Thr e Ala também podem ser utilizados na sequência do ligando. As sequências de aminoácidos que podem ser empregues com utilidade como ligandos incluem as reveladas em Maratea et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; Patente U.S. N° 4 935 233 e Patente U.S. N° 4 751 180. A sequência de ligando pode ser geralmente desde 1 até cerca de 50 aminoácidos de comprimento. As sequências de ligando não são

necessárias quando o primeiro e segundo polipéptidos possuem regiões de aminoácidos não essenciais N-terminais que podem ser utilizadas para separar os domínios funcionais e prevenir interferência estérica.

Numa forma de realização preferida, um ligando pode fornecer um sítio de clivagem específico entre um polipéptido Ra12 e um polipéptido heterólogo de interesse. Um tal sítio de clivagem pode conter um alvo para enzima proteolítica que inclui, por exemplo, enterocinase, Factor Xa, tripsina, colagenase, trombina, hidrolase de ubiquitina; ou para agentes de clivagem química, tais como, por exemplo, brometo de cianogénio ou hidroxiamina.

Um polipéptido de fusão pode conter opcionalmente uma etiqueta de afinidade que está ligada ao polipéptido de fusão de modo a que a purificação de polipéptidos recombinantes pode ser simplificada. Por exemplo, resíduos múltiplos de histidina codificados pela etiqueta permitem a utilização de métodos de cromatografia de afinidade com quelato de metal para a purificação de polipéptidos de fusão. Outros exemplos de moléculas de etiqueta de afinidade incluem, Strep-tag, PinPoint, proteína de ligação à maltose, glutathione S-transferase, etc. Ver, e. g., Glick e Pasternak (1999) *Molecular Biotechnology Principles and Applications of Recombinant DNA*, 2ª Ed., American Society for Microbiology, Washington, DC.

Os polipéptidos de fusão podem ser preparados utilizando qualquer uma de uma variedade de técnicas bem conhecidas. Os polipéptidos de fusão recombinantes codificados pelas sequências de ADN como descrito acima podem ser rapidamente preparados a partir das sequências de ADN utilizando qualquer um de uma

variedade de vectores de expressão conhecidos dos especialistas na técnica. A expressão pode ser alcançada em qualquer célula hospedeira apropriada que foi transformada ou transfectada com um vector de expressão contendo uma molécula de ADN que codifica um polipéptido recombinante. Células hospedeiras adequadas incluem células de procariotas, levedura e eucariotas superiores, como descrito acima. De um modo preferido, a célula hospedeira empregue é *E. coli*. Os sobrenadantes de sistemas de hospedeiro/vector adequados que segregam proteína ou polipéptido recombinante no meio de cultura podem ser primeiro concentrados utilizando um filtro comercialmente disponível. Após concentração, o concentrado pode ser aplicado a uma matriz de purificação adequada, tal como uma matriz de afinidade ou uma resina de permuta iónica. Finalmente, podem ser empregues um ou mais passos de HPLC de fase reversa para purificar ainda mais um polipéptido recombinante.

Porções e outras variantes possuindo menos do que cerca de 100 aminoácidos, e geralmente menos do que cerca de 50 aminoácidos, também podem ser criados por meios sintéticos, utilizando técnicas bem conhecidas dos especialistas na técnica. Por exemplo, esses polipéptidos podem ser sintetizados utilizando qualquer uma das técnicas de fase sólida comercialmente disponíveis, tais como o método de síntese de fase sólida de Merrifield, em que os aminoácidos são adicionados sequencialmente a uma cadeia de aminoácidos em crescimento. Ver Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. O equipamento para síntese automatizada de polipéptidos está comercialmente disponível em fornecedores, tais como Perkin Elmer/Applied BioSystems Division (Foster City, CA), e pode ser operado de acordo com as instruções do fabricante.

Em geral, os polipéptidos (incluindo proteínas de fusão) e polinucleótidos como aqui descritos são isolados. Um polipéptido ou polinucleótido "isolado" é aquele que é removido do seu ambiente original. Por exemplo, uma proteína que ocorre naturalmente é isolada se for separada do mesmo ou de todos os materiais coexistentes no sistema natural. De um modo preferido, esses polipéptidos são, pelo menos, cerca de 90% puros, de um modo preferido pelo menos, cerca de 95% puros e de um modo muito preferido pelo menos, cerca de 99% puros. Um polinucleótido é considerado como estando isolado se, por exemplo, for clonado num vector que não é uma parte do ambiente natural.

Adicionalmente a proporcionar expressão estável e rendimento elevado de polipéptidos de fusão de interesse, os ácidos nucleicos de fusão recombinante e os polipéptidos de fusão da invenção podem ser utilizados em vários outros métodos. Por exemplo, a sequência codificante do polipéptido de fusão da invenção pode ser utilizada para codificar um produto de proteína para utilização como um antigénio para detectar anticorpos no soro. Por exemplo, a presença de anticorpos no soro contra antigénios de *M. tuberculosis* num indivíduo indica que o indivíduo está infectado com *M. tuberculosis*. Em testes de diagnóstico convencionais, os anticorpos do soro contra *M. tuberculosis* são detectados monitorizando a ligação de anticorpos do soro contra proteínas de *M. tuberculosis*. Os polipéptidos de fusão da invenção são úteis como fontes de proteínas para monitorizar a ligação de anticorpos do soro a proteínas de fusão.

Alternativamente, o polipéptido de fusão pode ser utilizado como um imunogénio para induzir e/ou aumentar respostas imunitárias. Essas sequências codificantes podem ser ligadas à

sequência codificante de outra molécula, tal como um antigénio de *M. tuberculosis*, uma citocina ou um adjuvante. Esses polinucleótidos podem ser utilizados *in vivo* como uma vacina de ADN (Patentes U.S. Nºs 5 589 466; 5 679 647; e 5 703 055). Alternativamente, polipéptidos de fusão purificados ou parcialmente purificados ou fragmentos podem ser utilizados como vacinas ou composições terapêuticas. Qualquer um de uma variedade de métodos conhecidos na técnica pode ser empregue para produzir vacinas ou composições terapêuticas compreendendo os polipéptidos de fusão da presente invenção.

Purificação e Preparações de Proteína

Assim que uma proteína recombinante é expressa, pode ser identificada através de ensaios baseados nas propriedades físicas ou funcionais do produto, incluindo marcação radioactiva do produto, seguido por análise através de electroforese em gel, radioimunoensaio, ELISA, bioensaios, etc.

Assim que a proteína codificada é identificada, pode ser isolada e purificada através de métodos convencionais incluindo cromatografia (e. g., cromatografia líquida de elevado desempenho, permuta iónica, afinidade, e cromatografia em coluna de separação por tamanho), centrifugação, solubilidade diferencial, ou através de qualquer outra técnica convencional para a purificação de proteínas. Ver, de um modo geral, R. Scopes, *Protein Purification*, Springer-Verlag, N.Y. (1982), Deutscher, *Methods in Enzymology* Vol. 182: *Guide to Protein Purification*, Academic Press, Inc. N.Y. (1990). As condições actuais utilizadas dependerão, em parte, de factores tais como carga líquida, hidrofobicidade, hidrofiliicidade, etc., e serão

evidentes para os especialistas na técnica. As propriedades funcionais podem ser avaliadas utilizando quaisquer ensaios adequados.

As propriedades funcionais da proteína de fusão podem ser avaliadas utilizando qualquer ensaio adequado, tal como ligação de anticorpo, indução da proliferação de células T, estimulação da produção de citocina, tais como IL2, IL-4 e IFN- γ . Para a prática da presente invenção, é preferido que cada proteína de fusão seja pelo menos, 80% purificada de outras proteínas. É mais preferido que sejam pelo menos, 90% purificadas. Para administração *in vivo*, é preferido que as proteínas sejam mais do que 95% purificadas.

As proteínas purificadas podem ser ainda processadas antes da utilização. Por exemplo, as proteínas podem ser digeridas com uma enzima específica para separar o polipéptido Ra12 do polipéptido heterólogo.

Um especialista reconheceria que podem ser realizadas modificações aos ácidos nucleicos recombinantes e polipéptidos de fusão sem diminuir a sua actividade biológica. Algumas modificações podem ser realizadas para facilitar a clonagem, expressão, ou incorporação da molécula de etiqueta num polipéptido de fusão. Essas modificações são bem conhecidas dos especialistas na técnica e incluem, por exemplo, uma metionina adicionada ao terminal amino para proporcionar um sítio de iniciação, ou aminoácidos adicionais (e. g., poli His) colocados em qualquer dos terminais para criar sítios de restrição ou codões de terminação ou sequências de purificação localizados convenientemente.

Os seguintes Exemplos são oferecidos a título de ilustração e não como meio de limitação.

EXEMPLOS

Os exemplos seguintes descrevem experiências que ilustram que as construções de fusão de Ral2 produziram expressão estável e de rendimento elevado dos polipéptidos de fusão. Os exemplos seguintes também ilustram que várias sequências de Ral2 podem ser utilizadas como um parceiro de fusão.

EXEMPLO 1: A Sequência de Ral2 de Comprimento Total (SEQ ID N°:4) como um Parceiro de Fusão A. *Construção de Vectors de Expressão*

As sequências codificantes de antígenos de *M. tuberculosis* foram modificadas por PCR de modo a facilitar a sua fusão e subsequente expressão de proteína de fusão. O vector pET 17b (Novagen) foi modificado para incluir Ral2, um fragmento de 14 kDa C-terminal do antígeno da protease de serina MTB32A de *M. tuberculosis*. O codão de terminação de 3' da sequência Ral2 foi substituído com um sítio *EcoRI* em grelha e a extremidade N-terminal foi construída para codificar seis resíduos de etiqueta His imediatamente após o iniciador Met para facilitar um protocolo de purificação simples de um passo das proteínas recombinantes de Ral2 por cromatografia de afinidade ao longo da matriz Ni-NTA.

Especificamente, o fragmento C-terminal do antígeno MTB32A foi amplificado através de métodos de PCR convencionais

utilizando os iniciadores oligonucleotídicos 5' CAA TTA CAT ATG CAT CAC CAT CAC CAT CAC ACG GCC GCG TCC GAT AAC TTC (SEQ ID N°:13) e a sequência oligonucleotídica 3' é 5'--CTA ATC GAA TTC GGC CGG GGG TCC CTC GGC CAA (SEQ ID N°:14). O produto de 450 pb foi digerido com *NdeI* e *EcoRI* e clonado no vector de expressão pET 17b de um modo semelhante digerido com as mesmas enzimas. A expressão da proteína Ral2 recombinante foi realizada após transformação nas células hospedeiras de *E. coli* BL-21 (pLysE) (Novagen) e indução com IPTG. Após a lise das células de *E. coli* e centrifugação a 10K rpm, o Ral2 recombinante foi encontrado na fracção solúvel do sobrenadante. A proteína do sobrenadante solúvel foi purificada por cromatografia de afinidade numa coluna de Ni-NTA que permaneceu solúvel após diálise em 1 x PBS. A quantidade de proteína purificada obtida foi, de rotina, na gama de 60 a 100 mg por litro.

A sequência de DPPD foi construída para expressão como uma proteína de fusão com Ral2 concebendo iniciadores oligonucleotídicos para amplificar especificamente a forma madura segregada. O oligonucleótido 5' contendo um sítio de reconhecimento da enterocinase (DDDK) possui as sequências 5'--CAA TTA GAA TTC GAC GAC GAC GAC AAG GAT CCA CCT GAC CCG CAT CAG-3' (SEQ ID N°:15) e a sequência de oligonucleótidos 3' é 5'CAA TTA GAA TTC TCA GGG AGC GTT GGG CTG CTC (SEQ ID N°:16). O produto amplificado por PCR resultante foi digerido com *EcoRI* e subclonado no sítio *EcoRI* do vector pET-Ral2. Após a transformação na estirpe hospedeira de *E. coli* (XL1-blue; Stratagene), os clones contendo a inserção de tamanho correcto foram submetidos a sequenciação de modo a identificar aqueles que estavam em grelha com a fusão Ral2. Subsequentemente, o ADN de interesse (Fig. 3) foi transformado no hospedeiro bacteriano

BL-21 (plyse) e a proteína de fusão foi expressa após indução da cultura com IPTG.

B. Expressão e Purificação de Proteínas de Fusão

A proteína de fusão recombinante (His-etiqueta) Ral2-DPPD foi purificada a partir de culturas descontínuas de 500 mL induzidas com IPTG do sobrenadante solúvel por cromatografia de afinidade utilizando a matriz de um passo QIAexpress Ni-NTA Agarose matrix (QIAGEN, Chatsworth, CA) na presença de 8 M de ureia. Resumidamente, 20 mL de uma cultura de um dia para o outro, saturada, de BL21 contendo a construção pET foi adicionada a 500 mL de meio 2xYT contendo 50 µg/mL de ampicilina e 34 µg/mL de cloranfenicol, cultivado a 37 °C com agitação. As culturas bacterianas foram induzidas com IPTG a 2 mM até uma DO 560 de 0,3 e cultivada durante mais 3 h (DO 1,3 a 1,9). As células foram recolhidas a partir de culturas descontínuas de 500 mL por centrifugação e ressuspensas em 20 mL de tampão de ligação (fosfato de sódio a 0,1 M, pH 8,0; Tris-HCl a 10 mM, pH 8,0) contendo PMSF a 2 mM e 20 µg/mL de leupeptina. *E. coli* foi lisada adicionando 15 mg de lisozima e agitando em sistema de vaivém durante 30 min a 4 °C após sonicação (4 x 30 seg). As células lisadas foram centrifugadas a 12 k rpm durante 30 min e foi adicionada ureia directamente ao sobrenadante para uma concentração final de 8 M.

O sobrenadante foi ligado de modo descontínuo a resina de Ni-NTA agarose (5 mL de resina por induções de 500 mL) agitando em vaivém à temperatura ambiente durante 1 h e a matriz foi passada por uma coluna. O fluxo de passagem foi passado duas vezes pela mesma coluna, seguida por três lavagens com 30 mL

cada de tampão de lavagem (fosfato de sódio a 0,1 M e Tris-HCl a 10 mM, pH 6,3) também contendo ureia a 8 M. A proteína ligada foi eluída com 30 mL de imidazole a 100 mM em tampão de lavagem e foram recolhidas fracções de 5 mL. As fracções contendo o antigénio recombinante foram reunidas, dialisadas contra Tris-HCl a 10 mM (pH 8,0) ligadas mais uma vez a matriz Ni-NTA, eluídas e dialisadas em 1xPBS (pH 7,4) ou Tris-HCl a 10 mM (pH 7,8). O rendimento da proteína de fusão recombinante purificada estava em 50 a 75 mg por litro de cultura bacteriana induzida com mais de 95% de pureza, representando uma banda única. As proteínas recombinantes foram ensaiadas para contaminação de endotoxina utilizando o ensaio de Limulus (BioWhittaker) e demonstraram conter < 10 E.U./mg (< 1 ng LPS/mg).

C. Criação de Anti-soro

A proteína de fusão purificada (100 µg) foi misturada com 100 µg de dipéptido de muramilo, reconstituído até 1 mL com 1 x PBS e emulsificado com 1 mL de adjuvante IFA (Freunds incompleto; Life Technologies). A emulsão foi injectada em múltiplos sítios s.c. em coelhos fêmeas New Zealand (R&R Rabbitry, Stanwood, WA). Foram administrados ao coelho dois reforços subsequentes (100 µg de antigénio em IFA) a 6 semanas de distância e uma injeção final i.v. com 100 µg da proteína recombinante administrada novamente após 6 semanas. Uma semana após o reforço final, o coelho foi sacrificado e o soro foi recolhido e armazenado a -20 °C.

D. Análise de Imunotransferência

Lisado total de *M. tuberculosis* (estirpe H37Rv) ou PPD (2,5 µg cada) e 25 ng da proteína de fusão recombinante purificada Ral2-DPPD foram separados por electroforese em géis de SDS-PAGE a 16% e transferidos para nitrocelulose utilizando um dispositivo de transferência semi-seco (BioRad). As transferências, em duplicado, foram bloqueadas durante um mínimo de 1 h com PBS/Tween a 0,1% e rastreados com sonda com soros policlonais do mesmo coelho antes da imunização ou pós imunização com a proteína de fusão recombinante purificada (diluída 1:500 em PBS/Tween 20 a 0,1%). A reactividade foi avaliada como anteriormente utilizando [¹²⁵I]-proteína A, seguido por autorradiografia.

E. Resultados

Foram inicialmente avaliados vários sistemas de expressão para a expressão de DPPD em *E. coli*. Isto incluiu a sub-clonagem da sequência codificante de DPPD como construções de não fusão em 1) pET 17b (Novagen) e pQ30 (Qiagen, Santa Clarita, CA) ou 2) como construções de fusão utilizando pET32A (Novagen, Madison, WI) ou pGEX-2T (Pharmacia Biotech, Piscataway, NJ). Em todos esses sistemas, foi expresso e purificado muito pouco, ou nenhum, DPPD.

Em contraste, quando a sequência codificante de DPPD foi inserida a 3' da sequência de Ral2 num vector de expressão e transformada em *E. coli*, foi produzida uma grande quantidade de proteína de fusão Ral2-DPPD. A sequência de nucleótidos (SEQ ID N°:5) e sequência de aminoácidos (SEQ ID N°:6) de Ral2-DPPD são

reveladas na Figura 3. A imunogenicidade de DPPD foi mantida como evidenciado pela capacidade do anti-soro para reagir com a proteína purificada em análise de imunotransferência. Adicionalmente, três outras proteínas de origem eucariótica ou procariótica (ver Figuras 4-6) foram também expressas com sucesso pelas construções de fusão de Ral2. Assim, a sequência codificante de Ral2 é útil como um parceiro de fusão numa construção de expressão para facilitar a expressão de uma sequência heteróloga.

EXEMPLO 2: Polipéptido de Ral2 Curto (SEQ ID N°:17) como um Parceiro de Fusão

Neste exemplo, um polipéptido Ral2 compreendendo os aminoácidos 1-30 de SEQ ID N°:4 foi utilizado como um parceiro de fusão para ligação com o gene de comprimento total da mamaglobina humana. Esta forma curta de polipéptido Ral2 possui a sequência de aminoácidos apresentada em SEQ ID N°:17, e é aqui referido como "Ral2 (curto)".

Como apresentado na Figura 9, a extremidade 3' da sequência de Ral2 (curto) é fundida com o gene de comprimento total da mamaglobina humana. Especificamente, o gene da mamaglobina humana foi amplificado por métodos de PCR convencionais utilizando os seguintes iniciadores oligonucleotídicos: o iniciador 5', sítio *HindIII*: 5'-gcgaagcattATGAAGTTGCTGATGGTCCTCATGC-3' (SEQ ID N°:19); o iniciador 3', sítio *XhoI*: 5'-cggctcgagTTAAATAAATCACAAAGACTGCTGTC-3' (SEQ ID N°:20). Os sítios 5' *HindIII* e 3' *XhoI* foram adicionados para auxiliar a subclonagem num vector. A extremidade N-terminal da construção

de fusão foi construída para codificar seis resíduos His-etiqueta imediatamente após a Met para facilitar os protocolos de purificação. A expressão da construção de fusão foi realizada após a transformação em *E. coli* utilizando processos semelhantes aos descritos no Exemplo 1. Comparado com uma construção sem uma sequência de Ral2 (curto), a construção de fusão com uma sequência de Ral2 (curto) aumentou substancialmente a expressão da proteína de fusão Ral2 (curto)-mamaglobina.

EXEMPLO 3: polipéptido Ral2 mais Longo (SEQ ID N°:18) como um Parceiro de Fusão

Neste exemplo, um polipéptido Ral2 compreendendo os aminoácidos 1-128 de SEQ ID N°:4 foi utilizado como um parceiro de fusão para ligação com o gene de comprimento total da mamaglobina humana. Esta forma longa de polipéptido Ral2 possui a sequência de aminoácidos apresentada em SEQ ID N°:18, e é aqui referida como "Ral2 (longo)". Foram utilizados processos de clonagem e expressão semelhantes aos descritos no Exemplo. Em comparação com uma construção sem uma sequência de Ral2 (longa), a construção de fusão com uma sequência de Ral2 (longa) aumentou substancialmente a expressão da proteína de fusão Ral2 (longa)-mamaglobina.

A presente invenção não deve ser limitada no âmbito pelas formas de realização exemplificadas que pretendem ser ilustrações de aspectos da invenção, e quaisquer clones, nucleótidos ou sequências de aminoácidos que são funcionalmente equivalentes estão dentro do âmbito da invenção. Na verdade, várias modificações da invenção, adicionalmente aquelas aqui

descritas, tornar-se-ão óbvias para o especialista na técnica a partir da descrição anterior e dos desenhos acompanhantes. Também deve ser entendido que todos os tamanhos em pares de bases apresentados para nucleótidos são aproximados e são utilizados para efeitos de descrição.

LISTAGEM DE SEQUÊNCIAS

<110> Corixa Corporation

<120> Métodos de Utilização de uma Sequência Codificante de *Mycobacterium Tuberculosis* para facilitar a Expressão Estável e de Elevado rendimento de Proteínas Heterólogas

<130> AHB/FP6050272

<140> EP 00970619.3

<141> 2000-10-06

<150> PCT/US00/27652

<151> 2060-10-06

<150> US 60/158 585

<151> 1999-10-07

<160> 22

<170> Patentin Ver. 2.1

<210> 1

<211> 1872

<212> ADN

<213> *Mycobacterium tuberculosis*

<220>

<223> protease de serina MTB32A de 32 KD

<220>

<221> CDS

<222> (89) .. (1156)

<223> MTB32A

<220>

<221> sig_péptido

<222> (89) .. (184)

<223> sequência de sinal de secreção hidrofóbica N-terminal

<220>

<221> mat_péptido

<222> (185) .. (1153)

<400> 1

gactacgttg.gtgtagaaaa atcctgccgc ccggaccctt aaggctggga caatttctga 60

tagctacccc gacacaggag gttacggg atg agc aat tcg cgc cgc cgc tca 112
Met Ser Asn Ser Arg Arg Arg Ser
-30 -25

ctc agg tgg tca tgg ttg ctg agc gtg ctg gct gcc gtc ggg ctg ggc 160
Leu Arg Trp Ser Trp Leu Leu Ser Val Leu Ala Ala Val Gly Leu Gly
-20 -15 -10

ctg gcc acg gcg ccg gcc cag gcg gcc ccg ccg gcc ttg tcg cag gac 208
Leu Ala Thr Ala Pro Ala Gln Ala Ala Pro Pro Ala Leu Ser Gln Asp
-5 -1 1 5

cgg ttc gcc gac ttc ccc gcg ctg ccc ctc gac ccg tcc gcg atg gtc 256
Arg Phe Ala Asp Phe Pro Ala Leu Pro Leu Asp Pro Ser Ala Met Val
10 15 20

gcc caa gtg ggg cca cag gtg gtc aac atc aac acc aaa ctg ggc tac 304
Ala Gln Val Gly Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr
25 30 35 40

aac aac gcc gtg ggc gcc ggg acc ggc atc gtc atc gat ccc aac ggt 352
Asn Asn Ala Val Gly Ala Gly Thr Gly Ile Val Ile Asp Pro Asn Gly
45 50 55

gtc gtg ctg acc aac aac cac gtg atc gcg ggc gcc acc gac atc aat 400
Val Val Leu Thr Asn Asn His Val Ile Ala Gly Ala Thr Asp Ile Asn
60 65 70

gcg ttc agc gtc ggc tcc gcc caa acc tac ggc gtc gat gtg gtc ggg 448
Ala Phe Ser Val Gly Ser Gly Gln Thr Tyr Gly Val Asp Val Val Gly
75 80 85

tat gac cgc acc cag gat gtc gcg gtg ctg cag ctg cgc ggt gcc ggt 496
Tyr Asp Arg Thr Gln Asp Val Ala Val Leu Gln Leu Arg Gly Ala Gly
90 95 100

ggc ctg ccg tcg gcg gcg atc ggt ggc ggc gtc gcg gtt ggt gag ccc	544
Gly Leu Pro Ser Ala Ala Ile Gly Gly Gly Val Ala Val Gly Glu Pro	
105 110 115 120	
gtc gtc gcg atg ggc aac agc ggt ggc cag ggc gga acg ccc cgt gcg	592
Val Val Ala Met Gly Asn Ser Gly Gly Gln Gly Gly Thr Pro Arg Ala	
125 130 135	
gtg cct ggc agg gtg gtc gcg ctc ggc caa acc gtg cag gcg tcg gat	640
Val Pro Gly Arg Val Val Ala Leu Gly Gln Thr Val Gln Ala Ser Asp	
140 145 150	
tcg ctg acc ggt gcc gaa gag aca ttg aac ggg ttg atc cag ttc gat	688
Ser Leu Thr Gly Ala Glu Glu Thr Leu Asn Gly Leu Ile Gln Phe Asp	
155 160 165	
gcc gcg atc cag ccc ggt gat tcg ggc ggc ccc gtc gtc aac ggc cta	736
Ala Ala Ile Gln Pro Gly Asp Ser Gly Gly Pro Val Val Asn Gly Leu	
170 175 180	
gga cag gtg gtc ggt atg aac acg gcc gcg tcc gat aac ttc cag ctg	784
Gly Gln Val Val Gly Met Asn Thr Ala Ala Ser Asp Asn Phe Gln Leu	
185 190 195 200	
tcc cag ggt ggc cag gga ttc gcc att ccg atc ggc cag gcg atg gcg	832
Ser Gln Gly Gly Gln Gly Phe Ala Ile Pro Ile Gly Gln Ala Met Ala	
205 210 215	
atc gcg ggc cag atc cga tcg ggt ggc ggc tca ccc acc gtt cat atc	880
Ile Ala Gly Gln Ile Arg Ser Gly Gly Gly Ser Pro Thr Val His Ile	
220 225 230	
ggg cct acc gcc ttc ctc ggc ttg ggt gtt gtc gac aac aac ggc aac	928
Gly Pro Thr Ala Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly Asn	
235 240 245	
ggc gca cga gtc caa cgc gtg gtc ggc agc gct ccg gcg gca agt ctc	976
Gly Ala Arg Val Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser Leu	
250 255 260	

ggc atc tcc acc ggc gac gtg atc acc gcg gtc gac ggc gct ccg atc 1024
 Gly Ile Ser Thr Gly Asp Val Ile Thr Ala Val Asp Gly Ala Pro Ile
 265 270 275 280
 aac tcg gcc acc gcg atg gcg gac gcg ctt aac ggg cat cat ccc ggt 1072
 Asn Ser Ala Thr Ala Met Ala Asp Ala Leu Asn Gly His His Pro Gly
 285 290 295
 gac gtc atc tcg gtg acc tgg caa acc aag tcg ggc ggc acg cgt aca 1120
 Asp Val Ile Ser Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg Thr
 300 305 310
 ggg aac gtg aca ttg gcc gag gga ccc ccg gcc tga ttccgtccg 1166
 Gly Asn Val Thr Leu Ala Glu Gly Pro Pro Ala
 315 320
 gataccaccc gccggccggc caattggatt ggccccagcc gtgattgccg cgtgagcccc 1226
 cgagttccgt ctcccgtagc cgtggcatcg tggaaagcaat gaacgaggca gaacacagcg 1286
 tcgagcacc tcccgtgcag ggcagtcacg tcgaaggcgg tctgggtcag catccggatg 1346
 ccaaggactt cggcagcgcc gccgccctgc ccgccgatcc gacctggttt aagcacgccg 1406
 tcttctacga ggtgctggtc cgggcgttct tcgacgccag cgggacggg tccggcgatt 1466
 tgcgtggact catcgatccc ctcgactacc tgcagtggct tggcatcgac tgcattcggg 1526
 tgcgcgcgtt ctacgactcg ccgctgcgcg acggcggtta cgacattcgc gacttctaca 1586
 aggtgctgcc cgaattccgc accgtcgacg atttcgtcgc cctggtcgac gccgctcacc 1646
 ggcgaggtat ccgcattcgc accgacctgg tgatgaatca caactcggag tcgcacccct 1706
 ggtttcagga gtcccgccgc gaccagacg gaccgtacgg tgactattac gtgtggagcg 1766
 acaccagcga gcgtacacc gacgcccggg tctcttctgt cgacaccgaa ggtcgaact 1826
 ggteattcga tctgtccgc cgacagttct actggcaccg attctt 1872

<210> 2

<211> 355

<212> PRT

<213> *Mycobacterium tuberculosis*

<223> protease de serina MTB32A de 32 KD

<400> 2

Met	Ser	Asn	Ser	Arg	Arg	Arg	Ser	Leu	Arg	Trp	Ser	Trp	Leu	Leu	Ser
1				5				10					15		
Val	Leu	Ala	Ala	Val	Gly	Leu	Gly	Leu	Ala	Thr	Ala	Pro	Ala	Gln	Ala
			20					25					30		
Ala	Pro	Pro	Ala	Leu	Ser	Gln	Asp	Arg	Phe	Ala	Asp	Phe	Pro	Ala	Leu
			35				40					45			
Pro	Leu	Asp	Pro	Ser	Ala	Met	Val	Ala	Gln	Val	Gly	Pro	Gln	Val	Val
			50			55					60				
Asn	Ile	Asn	Thr	Lys	Leu	Gly	Tyr	Asn	Asn	Ala	Val	Gly	Ala	Gly	Thr
65				70						75					80
Gly	Ile	Val	Ile	Asp	Pro	Asn	Gly	Val	Val	Leu	Thr	Asn	Asn	His	Val
				85				90						95	
Ile	Ala	Gly	Ala	Thr	Asp	Ile	Asn	Ala	Phe	Ser	Val	Gly	Ser	Gly	Gln
			100					105					110		
Thr	Tyr	Gly	Val	Asp	Val	Val	Gly	Tyr	Asp	Arg	Thr	Gln	Asp	Val	Ala
		115					120					125			
Val	Leu	Gln	Leu	Arg	Gly	Ala	Gly	Gly	Leu	Pro	Ser	Ala	Ala	Ile	Gly
	130				135						140				
Gly	Gly	Val	Ala	Val	Gly	Glu	Pro	Val	Val	Ala	Met	Gly	Asn	Ser	Gly
145					150					155				160	
Gly	Gln	Gly	Gly	Thr	Pro	Arg	Ala	Val	Pro	Gly	Arg	Val	Val	Ala	Leu
				165				170						175	
Gly	Gln	Thr	Val	Gln	Ala	Ser	Asp	Ser	Leu	Thr	Gly	Ala	Glu	Glu	Thr
			180					185					190		
Leu	Asn	Gly	Leu	Ile	Gln	Phe	Asp	Ala	Ala	Ile	Gln	Pro	Gly	Asp	Ser
	195					200					205				
Gly	Gly	Pro	Val	Val	Asn	Gly	Leu	Gly	Gln	Val	Val	Gly	Met	Asn	Thr
	210				215					220					
Ala	Ala	Ser	Asp	Asn	Phe	Gln	Leu	Ser	Gln	Gly	Gly	Gln	Gly	Phe	Ala
225				230						235				240	
Ile	Pro	Ile	Gly	Gln	Ala	Met	Ala	Ile	Ala	Gly	Gln	Ile	Arg	Ser	Gly
			245					250					255		
Gly	Gly	Ser	Pro	Thr	Val	His	Ile	Gly	Pro	Thr	Ala	Phe	Leu	Gly	Leu
			260					265					270		
Gly	Val	Val	Asp	Asn	Asn	Gly	Asn	Gly	Ala	Arg	Val	Gln	Arg	Val	Val
	275					280					285				
Gly	Ser	Ala	Pro	Ala	Ala	Ser	Leu	Gly	Ile	Ser	Thr	Gly	Asp	Val	Ile
	290				295					300					
Thr	Ala	Val	Asp	Gly	Ala	Pro	Ile	Asn	Ser	Ala	Thr	Ala	Met	Ala	Asp
305				310						315				320	
Ala	Leu	Asn	Gly	His	His	Pro	Gly	Asp	Val	Ile	Ser	Val	Thr	Trp	Gln
			325					330					335		
Thr	Lys	Ser	Gly	Gly	Thr	Arg	Thr	Gly	Asn	Val	Thr	Leu	Ala	Glu	Gly
			340					345					350		
Pro	Pro	Ala													
		355													

<210> 3

<211> 396

<212> ADN

<213> *Mycobacterium tuberculosis*

<220>

<223> fragmento do terminal C de MTB32A Ra12 de 14 KD

<220>

<221> CDS

<222> ((1)..(396))

<223> Ra12

<400>

```
acg gcc gcg tcc gat aac ttc cag ctg tcc cag ggt ggg cag gga ttc 48
Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe
  1           5           10          15

gcc att ccg atc ggg cag gcg atg gcg atc gcg ggc cag atc cga tgc 96
Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
          20          25          30

ggt ggg ggg tca ccc acc gtt cat atc ggg cct acc gcc ttc ctc ggc 144
Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
          35          40          45

ttg ggt gtt gtc gac aac aac ggc aac ggc gca cga gtc caa cgc gtg 192
Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
          50          55          60

gtc ggg agc gct ccg gcg gca agt ctc ggc atc tcc acc ggc gac gtg 240
Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val
          65          70          75          80

atc acc gcg gtc gac ggc gct ccg atc aac tcc gcc acc gcg atg gcg 288
Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala
          85          90          95

gac gcg ctt aac ggg cat cat ccc ggt gac gtc atc tcc gtg acc tgg 336
Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Thr Trp
          100          105          110

caa acc aag tcc ggc ggc acg cgt aca ggg aac gtg aca ttg gcc gag 384
Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu
          115          120          125

gga ccc ccg gcc 396
Gly Pro Pro Ala
          130
```

<210> 4
 <211> 132
 <212> PRT
 <213> *Mycobacterium tuberculosis*
 <223> fragmento do terminal C de MTB32A Ra12 de 14 KD

<400> 4

```

    Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe
      1           5           10           15
    Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
           20           25           30
    Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
           35           40           45
    Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
           50           55           60
    Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val
           65           70           75           80
    Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala
           85           90           95
    Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Thr Trp
           100          105          110
    Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu
           115          120          125
    Gly Pro Pro Ala
           130
  
```

<210> 5
 <211> 702
 <212> ADN
 <213> Sequência Artificial

<220>
 <221> CDS
 <222> (4)..(696)
 <223> polipéptido de fusão Ra12-DPPD

<220>
 <223> Descrição da Sequência Artificial: polipéptido de fusão
 Ra12-DPPD

<400> 5

cat atg cat cac cat cac cat cac acg gcc gcg tcc gat aac ttc cag	48
Met His His His His His His Thr Ala Ala Ser Asp Asn Phe Gln	
1 5 10 15	
ctg tcc cag ggt ggg cag gga ttc gcc att ccg atc ggg cag gcg atg	96
Leu Ser Gln Gly Gly Gln Gly Phe Ala Ile Pro Ile Gly Gln Ala Met	
20 25 30	
gcg atc gcg ggc cag atc cga tcc ggt ggg ggg tca ccc acc gtt cat	144
Ala Ile Ala Gly Gln Ile Arg Ser Gly Gly Gly Ser Pro Thr Val His	
35 40 45	
atc ggg cct acc gcc ttc ctc ggc ttg ggt gtt gtc gac aac aac ggc	192
Ile Gly Pro Thr Ala Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly	
50 55 60	
aac ggc gca cga gtc caa cgc gtg gtc ggg agc gct ccg gcg gca agt	240
Asn Gly Ala Arg Val Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser	
65 70 75	
ctc ggc atc tcc acc ggc gac gtg atc acc gcg gtc gac ggc gct ccg	288
Leu Gly Ile Ser Thr Gly Asp Val Ile Thr Ala Val Asp Gly Ala Pro	
80 85 90 95	
atc aac tcc gcc acc gcg atg gcg gac gcg ctt aac ggg cat cat ccc	336
Ile Asn Ser Ala Thr Ala Met Ala Asp Ala Leu Asn Gly His His Pro	
100 105 110	
ggc gac gtc atc tcc gtg acc tgg caa acc aag tcc ggc ggc acg cgt	384
Gly Asp Val Ile Ser Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg	
115 120 125	
aca ggg aac gtg aca ttg gcc gag gga ccc ccg gcc gaa ttc gac gac	432
Thr Gly Asn Val Thr Leu Ala Glu Gly Pro Pro Ala Glu Phe Asp Asp	
130 135 140	
gac gac aag gat cca cct gac ccg cat cag ccg gac atg acg aaa ggc	480
Asp Asp Lys Asp Pro Pro Asp Pro His Gln Pro Asp Met Thr Lys Gly	
145 150 155	
tat tgc ccg ggt ggc cga tgg ggt ttt ggc gac ttg gcc gtg tgc gac	528
Tyr Cys Pro Gly Gly Arg Trp Gly Phe Gly Asp Leu Ala Val Cys Asp	
160 165 170 175	
ggc gag aag tac ccc gac ggc tcc ttt tgg cac cag tgg atg caa acg	576
Gly Glu Lys Tyr Pro Asp Gly Ser Phe Trp His Gln Trp Met Gln Thr	
180 185 190	
tgg ttt acc ggc cca cag ttt tac ttc gat tgt gtc agc ggc ggt gag	624
Trp Phe Thr Gly Pro Gln Phe Tyr Phe Asp Cys Val Ser Gly Gly Glu	
195 200 205	
ccc ctc ccc ggc ccg ccg cca ccg ggt ggt tgc ggt ggg gca att ccg	672
Pro Leu Pro Gly Pro Pro Pro Pro Gly Gly Cys Gly Gly Ala Ile Pro	

210	215	220	
tcc gag cag ccc aac gct ccc tga gaattc			702
Ser Glu Gln Pro Asn Ala Pro			
225	230		

<210> 6

<211> 230

<212> PRT

<213> Sequência Artificial

<223> Descrição da Sequência Artificial: polipéptido de fusão
Ra12-DPPD

<400> 6

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

<210> 7
 <211> 1746
 <212> ADN
 <213> Sequência Artificial

 <220>
 <223> Descrição da Sequência Artificial: fusão Ral2-WT1

 <220>
 <221> CDS
 <222> (4)..(1740)
 <223> polipéptido de fusão Ral2-WT1

 <400> 7

```

cat atg cat cac cat cac cat cac acg gcc gcg tcc gat aac ttc cag 48
Met His His His His His His Thr Ala Ala Ser Asp Asn Phe Gln
  1             5             10             15

ctg tcc cag ggt ggg cag gga ttc gcc att ccg atc ggg cag gcg atg 96
Leu Ser Gln Gly Gly Gln Gly Phe Ala Ile Pro Ile Gly Gln Ala Met
             20             25             30

gcg atc gcg ggc cag atc cga tgg ggt ggg ggg tca ccc acc gtt cat 144
Ala Ile Ala Gly Gln Ile Arg Ser Gly Gly Gly Ser Pro Thr Val His
             35             40             45

atc ggg cct acc gcc ttc ctg gcc ttg ggt gtt gtc gac aac aac ggc 192
Ile Gly Pro Thr Ala Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly
             50             55             60

aac ggc gca cga gtc caa cgc gtg gtc ggg agc gct ccg gcg gca agt 240
Asn Gly Ala Arg Val Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser
             65             70             75

ctc ggc atc tcc acc gcc gac gtg atc acc gcg gtc gac ggc gct ccg 288
Leu Gly Ile Ser Thr Gly Asp Val Ile Thr Ala Val Asp Gly Ala Pro
             80             85             90             95

atc aac tgg gcc acc gcg atg gcg gac gcg ctt aac ggg cat cat ccc 336
Ile Asn Ser Ala Thr Ala Met Ala Asp Ala Leu Asn Gly His His Pro
             100             105             110

ggg gac gtc atc tgg gtg acc tgg caa acc aag tgg ggc ggc acg cgt 384
Gly Asp Val Ile Ser Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg
             115             120             125

```

aca ggg aac gtg aca ttg gcc gag gga ccc ccg gcc gaa ttc ccg ctg	432
Thr Gly Asn Val Thr Leu Ala Glu Gly Pro Pro Ala Glu Phe Pro Leu	
130 135 140	
gtg ccg cgc gcc agc ccg atg ggc tcc gac gtt ccg gac ctg aac gca	480
Val Pro Arg Gly Ser Pro Met Gly Ser Asp Val Arg Asp Leu Asn Ala	
145 150 155	
ctg ctg ccg gca gtt ccg tcc ctg ggt ggt ggt ggt ggt tgc gca ctg	528
Leu Leu Pro Ala Val Pro Ser Leu Gly Gly Gly Gly Cys Ala Leu	
160 165 170 175	
ccg gtt agc ggt gca gca cag tgg gct ccg gtt ctg gac ttc gca ccg	576
Pro Val Ser Gly Ala Ala Gln Trp Ala Pro Val Leu Asp Phe Ala Pro	
180 185 190	
ccg ggt gca tcc gca tac ggt tcc ctg ggt ggt ccg gca ccg ccg ccg	624
Pro Gly Ala Ser Ala Tyr Gly Ser Leu Gly Gly Pro Ala Pro Pro Pro	
195 200 205	
gca ccg ccg ccg ccg ccg ccg ccg ccg ccg cac tcc ttc atc aaa cag	672
Ala Pro Pro Pro Pro Pro Pro Pro Pro Pro His Ser Phe Ile Lys Gln	
210 215 220	
gaa ccg agc tgg ggt ggt gca gaa ccg cac gaa gaa cag tgc ctg agc	720
Glu Pro Ser Trp Gly Gly Ala Glu Pro His Glu Glu Gln Cys Leu Ser	
225 230 235	
gca ttc acc gtt cac ttc tcc ggc cag ttc act ggc aca gcc gga gcc	768
Ala Phe Thr Val His Phe Ser Gly Gln Phe Thr Gly Thr Ala Gly Ala	
240 245 250 255	
tgt cgc tac ggg ccc ttc ggt cct cct ccg ccc agc cag gcg tca tcc	816
Cys Arg Tyr Gly Pro Phe Gly Pro Pro Pro Ser Gln Ala Ser Ser	
260 265 270	
ggc cag gcc agg atg ttt cct aac ggc ccc tac ctg ccc agc tgc ctc	864
Gly Gln Ala Arg Met Phe Pro Asn Ala Pro Tyr Leu Pro Ser Cys Leu	
275 280 285	
gag agc cag ccc gct att cgc aat cag ggt tac agc acg gtc acc ttc	912
Glu Ser Gln Pro Ala Ile Arg Asn Gln Gly Tyr Ser Thr Val Thr Phe	
290 295 300	
gac ggg acg ccc agc tac ggt cac acg ccc tgc cac cat gcg gcg cag	960
Asp Gly Thr Pro Ser Tyr Gly His Thr Pro Ser His His Ala Ala Gln	
305 310 315	
ttc ccc aac cac tca ttc aag cat gag gat ccc atg ggc cag cag ggc	1008
Phe Pro Asn His Ser Phe Lys His Glu Asp Pro Met Gly Gln Gln Gly	
320 325 330 335	
tgc ctg ggt gag cag cag tac tgc gtg ccg ccc ccg gtc tat ggc tgc	1056
Ser Leu Gly Glu Gln Gln Tyr Ser Val Pro Pro Pro Val Tyr Gly Cys	
340 345 350	
cac acc ccc acc gac agc tgc acc ggc agc cag gct ttg ctg ctg agg	1104
His Thr Pro Thr Asp Ser Cys Thr Gly Ser Gln Ala Leu Leu Leu Arg	
355 360 365	

acg ccc tac agc agt gac aat tta tac caa atg aca tcc cag ctt gaa	1152
Thr Pro Tyr Ser Ser Asp Asn Leu Tyr Gln Met Thr Ser Gln Leu Glu	
370 375 380	
tgc atg acc tgg aat cag atg aac tta gga gcc acc tta aag ggc cac	1200
Cys Met Thr Trp Asn Gln Met Asn Leu Gly Ala Thr Leu Lys Gly His	
385 390 395	
agc aca ggg tac gag agc gat aac cac aca acg ccc atc ctc tgc gga	1248
Ser Thr Gly Tyr Glu Ser Asp Asn His Thr Thr Pro Ile Leu Cys Gly	
400 405 410 415	
gcc caa tac aga ata cac acg cac ggt gtc ttc aga ggc att cag gat	1296
Ala Gln Tyr Arg Ile His Thr His Gly Val Phe Arg Gly Ile Gln Asp	
420 425 430	
gtg cga cgt gtg cct gga gta gcc ccg act ctt gta cgg tcg gca tct	1344
Val Arg Arg Val Pro Gly Val Ala Pro Thr Leu Val Arg Ser Ala Ser	
435 440 445	
gag acc agt gag aaa cgc ccc ttc atg tgt gct tac tca ggc tgc aat	1392
Glu Thr Ser Glu Lys Arg Pro Phe Met Cys Ala Tyr Ser Gly Cys Asn	
450 455 460	
aag aga tat ttt aag ctg tcc cac tta cag atg cac agc agg aag cac	1440
Lys Arg Tyr Phe Lys Leu Ser His Leu Gln Met His Ser Arg Lys His	
465 470 475	
act ggt gag aaa cca tac cag tgt gac ttc aag gac tgt gaa cga agg	1488
Thr Gly Glu Lys Pro Tyr Gln Cys Asp Phe Lys Asp Cys Glu Arg Arg	
480 485 490 495	
ctt ttt cgt tca gac cag ctc aaa aga cac caa agg aga cat aca ggt	1536
Phe Phe Arg Ser Asp Gln Leu Lys Arg His Gln Arg Arg His Thr Gly	
500 505 510	
gtg aca cca ttc cag tgt aca act tgt cag cga aag ttc tcc cgg tcc	1584
Val Lys Pro Phe Gln Cys Lys Thr Cys Gln Arg Lys Phe Ser Arg Ser	
515 520 525	
gac cac ctg aag acc cac acc agg act cat aca ggt gaa aag ccc ttc	1632
Asp His Leu Lys Thr His Thr Arg Thr His Thr Gly Glu Lys Pro Phe	
530 535 540	
agc tgt cgg tgg cca agt tgt cag aaa aag ttt gcc cgg tca gat gaa	1680
Ser Cys Arg Trp Pro Ser Cys Gln Lys Lys Phe Ala Arg Ser Asp Glu	
545 550 555	
tta gtc cgc cat cac aac atg cat cag aga aac atg acc aca ctc cag	1728
Leu Val Arg His His Asn Met His Gln Arg Asn Met Thr Lys Leu Gln	
560 565 570 575	
ctg gcg ctt tga gaattc	1746
Leu Ala Leu	

<210> 8

<211> 578

<212> Sequência Artificial

<223> Descrição da Sequência Artificial: polipéptido de fusão
Ra12-WT1

<400> 8

```
Met His His His His His His Thr Ala Ala Ser Asp Asn Phe Gln Leu
 1          5          10          15
Ser Gln Gly Gly Gln Gly Phe Ala Ile Pro Ile Gly Gln Ala Met Ala
 20          25          30
Ile Ala Gly Gln Ile Arg Ser Gly Gly Gly Ser Pro Thr Val His Ile
 35          40          45
Gly Pro Thr Ala Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly Asn
 50          55          60
Gly Ala Arg Val Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser Leu
 65          70          75          80
Gly Ile Ser Thr Gly Asp Val Ile Thr Ala Val Asp Gly Ala Pro Ile
 85          90          95
Asn Ser Ala Thr Ala Met Ala Asp Ala Leu Asn Gly His His Pro Gly
100          105          110
Asp Val Ile Ser Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg Thr
115          120          125
Gly Asn Val Thr Leu Ala Glu Gly Pro Pro Ala Glu Phe Pro Leu Val
130          135          140
Pro Arg Gly Ser Pro Met Gly Ser Asp Val Arg Asp Leu Asn Ala Leu
145          150          155          160
Leu Pro Ala Val Pro Ser Leu Gly Gly Gly Gly Cys Ala Leu Pro
165          170          175
Val Ser Gly Ala Ala Gln Trp Ala Pro Val Leu Asp Phe Ala Pro Pro
180          185          190
Gly Ala Ser Ala Tyr Gly Ser Leu Gly Gly Pro Ala Pro Pro Pro Ala
195          200          205
Pro Pro Pro Pro Pro Pro Pro Pro His Ser Phe Ile Lys Gln Glu
210          215          220
Pro Ser Trp Gly Gly Ala Glu Pro His Glu Glu Gln Cys Leu Ser Ala
225          230          235          240
Phe Thr Val His Phe Ser Gly Gln Phe Thr Gly Thr Ala Gly Ala Cys
245          250          255
Arg Tyr Gly Pro Phe Gly Pro Pro Pro Pro Ser Gln Ala Ser Ser Gly
260          265          270
Gln Ala Arg Met Phe Pro Asn Ala Pro Tyr Leu Pro Ser Cys Leu Glu
275          280          285
Ser Gln Pro Ala Ile Arg Asn Gln Gly Tyr Ser Thr Val Thr Phe Asp
290          295          300
Gly Thr Pro Ser Tyr Gly His Thr Pro Ser His His Ala Ala Gln Phe
305          310          315          320
Pro Asn His Ser Phe Lys His Glu Asp Pro Met Gly Gln Gln Gly Ser
325          330          335
Leu Gly Glu Gln Gln Tyr Ser Val Pro Pro Pro Val Tyr Gly Cys His
340          345          350
Thr Pro Thr Asp Ser Cys Thr Gly Ser Gln Ala Leu Leu Leu Arg Thr
355          360          365
```

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

Ala Leu

<210> 9

<211> 672

<212> ADN

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: fusão de Ral2-mamaglobina humana

<220>

<221> CDS

<222> (4)..(666)

<223> polipéptido de fusão Ral2-mamaglobina humana

<400> 9

cat atg cat cac cat cac cat cac acg gcc gcg tcc gat aac ttc cag	48
Met His His His His His His Thr Ala Ala Ser Asp Asn Phe Gln	
1 5 10 15	
ctg tcc cag ggt ggg cag gga ttc gcc att ccg atc ggg cag gcg atg	96
Leu Ser Gln Gly Gly Gln Gly Phe Ala Ile Pro Ile Gly Gln Ala Met	
20 25 30	
gcg atc gcg ggc cag atc cga tgc ggt ggg ggg tca ccc acc gtt cat	144
Ala Ile Ala Gly Gln Ile Arg Ser Gly Gly Gly Ser Pro Thr Val His	
35 40 45	
atc ggg cct acc gcc ttc ctc ggc ttg ggt gtt gtc gac aac aac gcc	192
Ile Gly Pro Thr Ala Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly	
50 55 60	
aac ggc gca cga gtc caa cgc gtg gtc ggg agc gct ccg gcg gca agt	240
Asn Gly Ala Arg Val Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser	
65 70 75	
ctc ggc atc tcc acc ggc gac gtg atc acc gcg gtc gac ggc gct ccg	288
Leu Gly Ile Ser Thr Gly Asp Val Ile Thr Ala Val Asp Gly Ala Pro	
80 85 90 95	
atc aac tgc gcc acc gcg atg gcg gac gcg ctt aac ggg cat cat ccc	336
Ile Asn Ser Ala Thr Ala Met Ala Asp Ala Leu Asn Gly His His Pro	
100 105 110	
ggt gac gtc atc tgc gtg acc tgg caa acc aag tgc ggc ggc acg cgt	384
Gly Asp Val Ile Ser Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg	
115 120 125	
aca ggg aac gtg aca ttg gcc gag gga ccc ccg gcc gaa ttc atc gag	432
Thr Gly Asn Val Thr Leu Ala Glu Gly Pro Pro Ala Glu Phe Ile Glu	
130 135 140	
gga agg ggc tct ggc tgc ccc tta ttg gag aat gtg att tcc aag aca	480
Gly Arg Gly Ser Gly Cys Pro Leu Leu Glu Asn Val Ile Ser Lys Thr	
145 150 155	
atc aat cca caa gtg tct aag act gaa tac aaa gaa ctt ctt caa gag	528
Ile Asn Pro Gln Val Ser Lys Thr Glu Tyr Lys Glu Leu Leu Gln Glu	
160 165 170 175	
ttc ata gac gac aat gcc act aca aat gcc ata gat gaa ttg aag gaa	576
Phe Ile Asp Asp Asn Ala Thr Thr Asn Ala Ile Asp Glu Leu Lys Glu	
180 185 190	
tgt ttt ctt aac caa acg gat gaa act ctg agc aat gtt gag gtg ttt	624
Cys Phe Leu Asn Gln Thr Asp Glu Thr Leu Ser Asn Val Glu Val Phe	
195 200 205	
atg caa tta ata tat gac agc agt ctt tgt gat tta ttt taa gaattc	672
Met Gln Leu Ile Tyr Asp Ser Ser Leu Cys Asp Leu Phe	
210 215 220	

<210> 10
 <211> 220
 <212> PRT
 <213> Sequência Artificial
 <223> Descrição da Sequência Artificial: polipéptido de fusão
 Ra12-mamaglobina humana

 <400> 10

```

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

```

<210> 11
 <211> 2191
 <212> ADN
 <213> Sequência Artificial

 <220>

<223> Descrição da Sequência Artificial: fusão Ra12-H9-32A
(fusão Ra12-MTB39-MTB32A(N-ter))

<220>

<221> CDS

<222> (1)..(2190)

<223> polipéptido de fusão Ra12-H9-32A (Ra12-MTH32A(N-ter))

<400> 11

```

atg cat cac cat cac cat cac acg gcc gcg tcc gat aac ttc cag ctg   48
Met His His His His His His Thr Ala Ala Ser Asp Asn Phe Gln Leu
  1             5             10             15

tcc cag ggt ggg cag gga ttc gcc att ccg atc ggg cag gcg atg gcg   96
Ser Gln Gly Gly Gln Gly Phe Ala Ile Pro ile Gly Gln Ala Met Ala
      20             25             30

atc gcg ggc cag atc cga tcg ggt ggg ggg tca ccc acc gtt cat atc   144
Ile Ala Gly Gln Ile Arg Ser Gly Gly Gly Ser Pro Thr Val His Ile
      35             40             45

ggg cct acc gcc ttc ctc ggc ttg ggt gtt gtc gac aac aac ggc aac   192
Gly Pro Thr Ala Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly Asn
      50             55             60

ggc gca cga gtc caa cgc gtg gtc ggg agc gct ccg gcg gca agt ctc   240
Gly Ala Arg Val Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser Leu
      65             70             75             80

ggc atc tcc acc ggc gac gtg atc acc gcg gtc gac ggc gct ccg atc   288
Gly ile Ser Thr Gly Asp Val ile Thr Ala Val Asp Gly Ala Pro ile
      85             90             95

aac tcg gcc acc gcg atg gcg gac gcg ctt aac ggg cat cat ccc ggt   336
Asn Ser Ala Thr Ala Met Ala Asp Ala Leu Asn Gly His His Pro Gly
      100            105            110

gac gtc atc tcg gtg acc tgg caa acc aag tcg ggc ggc acg cgt aca   384
Asp Val ile Ser Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg Thr
      115            120            125

ggg aac gtg aca ttg gcc gag gga ccc ccg gcc gaa ttc atg gtg gat   432
Gly Asn Val Thr Leu Ala Glu Gly Pro Pro Ala Glu Phe Met Val Asp
      130            135            140

```

ttc ggg gcg tta cca ccg gag atc aac tcc gcg agg atg tac gcc ggc	480
Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met Tyr Ala Gly	
145 150 155 160	
ccg ggt tcg gcc tcg ctg gtg gcc gcg gct cag atg tgg gac agc gtg	528
Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp Asp Ser Val	
165 170 175	
gcg agt gac ctg ttt tcg gcc gcg tcg gcg ttt cag tcg gtg gtc tgg	576
Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser Val Val Trp	
180 185 190	
ggc ctg acg gtg ggg tcg tgg ata ggt tcg tcg gcg ggt ctg atg gtg	624
Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly Leu Met Val	
195 200 205	
gcg gcg gcc tcg ccg tat gtg gcg tgg atg agc gtc acc gcg ggg cag	672
Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr Ala Gly Gln	
210 215 220	
gcc gag ctg acc gcc gcc cag gtc cgg gtt gct gcg gcg gcc tac gag	720
Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala Ala Tyr Glu	
225 230 235 240	
acg gcg tat ggg ctg acg gtg ccc ccg ccg gtg atc gcc gag aac cgt	768
Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala Glu Asn Arg	
245 250 255	
gct gaa ctg atg att ctg ata gcg acc aac ctc ttg ggg caa aac acc	816
Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly Gln Asn Thr	
260 265 270	
ccg gcg atc gcg gtc aac gag gcc gaa tac gcc gag atg tgg gcc caa	864
Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln	
275 280 285	

gac gcc gcc gcg atg ttt ggc tac gcc gcg gcg acg gcg acg gcg acg	912
Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr	
290 295 300	
gcg acg ttg ctg ccg ttc gag gag gcg ccg gag atg acc agc gcg ggt	960
Ala Thr Leu Leu Pro Phe Glu Glu Ala Pro Glu Met Thr Ser Ala Gly	
305 310 315 320	
ggg ctc ctc gag cag gcc gcc gcg gtc gag gag gcc tcc gac acc gcc	1008
Gly Leu Leu Glu Gln Ala Ala Ala Val Glu Glu Ala Ser Asp Thr Ala	
325 330 335	
gcg gcg aac cag ttg atg aac aat gtg ccc cag gcg ctg caa cag ctg	1056
Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu Gln Leu	
340 345 350	
gcc cag ccc acg cag gcc acc acg cct tct tcc aag ctg ggt ggc ctg	1104
Ala Gln Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu	
355 360 365	
tgg aag acg gtc tcg ccg cat cgg tcg ccg atc agc aac atg gtg tcg	1152
Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser Asn Met Val Ser	
370 375 380	
atg gcc aac aac cac atg tcg atg acc aac tcg ggt gtg tcg atg acc	1200
Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gly Val Ser Met Thr	
385 390 395 400	
aac acc ttg agc tcg atg ttg aag ggc ttt gct ccg gcg gcg gcc gcc	1248
Asn Thr Leu Ser Ser Met Leu Lys Gly Phe Ala Pro Ala Ala Ala Ala	
405 410 415	
cag gcc gtg caa acc gcg gcg caa aac ggg gtc cgg gcg atg agc tcg	1296
Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala Met Ser Ser	
420 425 430	
ctg ggc agc tcg ctg ggt tct tcg ggt ctg ggc ggt ggg gtg gcc gcc	1344
Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly Gly Val Ala Ala	
435 440 445	
aac ttg ggt cgg gcg gcc tcg gtc ggt tcg ttg tcg gtg ccg cag gcc	1392
Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val Pro Gln Ala	
450 455 460	
tgg gcc gcg gcc aac cag gca gtc acc ccg gcg gcg ccg gcg ctg ccg	1440
Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg Ala Leu Pro	
465 470 475 480	
ctg acc agc ctg acc agc gcc gcg gaa aga ggg ccc ggg cag atg ctg	1488
Leu Thr Ser Leu Thr Ser Ala Ala Glu Arg Gly Pro Gly Gln Met Leu	
485 490 495	
ggc ggg ctg ccg gtg ggg cag atg ggc gcc agg gcc ggt ggt ggg ctc	1536
Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly Gly Gly Leu	
500 505 510	
agt ggt gtg ctg cgt gtt ccg ccg cga ccc tat gtg atg ccg cat tct	1584
Ser Gly Val Leu Arg Val Pro Pro Arg Pro Tyr Val Met Pro His Ser	
515 520 525	

ccg gca gcc ggc gat atc gcc ccg ccg gcc ttg tcg cag gac cgg ttc	1632
Pro Ala Ala Gly Asp Ile Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe	
530 535 540	
gcc gac ttc ccc gcg ctg ccc ctc gac ccg tcc gcg atg gtc gcc caa	1680
Ala Asp Phe Pro Ala Leu Pro Leu Asp Pro Ser Ala Met Val Ala Gln	
545 550 555 560	
gtg ggg cca cag gtg gtc aac atc aac acc aaa ctg ggc tac aac aac	1728
Val Gly Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn	
565 570 575	
gcc gtg ggc gcc ggg acc ggc atc gtc atc gat ccc aac ggt gtc gtg	1776
Ala Val Gly Ala Gly Thr Gly Ile Val Ile Asp Pro Asn Gly Val Val	
580 585 590	
ctg acc aac aac cac gtg atc gcg ggc gcc acc gac atc aat gcg ttc	1824
Leu Thr Asn Asn His Val Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe	
595 600 605	
agc gtc ggc tcc ggc caa acc tac ggc gtc gat gtg gtc ggg tat gac	1872
Ser Val Gly Ser Gly Gln Thr Tyr Gly Val Asp Val Val Gly Tyr Asp	
610 615 620	
cgc acc cag gat gtc gcg gtg ctg cag ctg cgc ggt gcc ggt ggc ctg	1920
Arg Thr Gln Asp Val Ala Val Leu Gln Leu Arg Gly Ala Gly Gly Leu	
625 630 635 640	
ccg tcg gcg gcg atc ggt ggc ggc gtc gcg gtt ggt gag ccc gtc gtc	1968
Pro Ser Ala Ala Ile Gly Gly Gly Val Ala Val Gly Glu Pro Val Val	
645 650 655	
gcg atg ggc aac agc ggt ggg cag ggc gga acg ccc cgt gcg gtg cct	2016
Ala Met Gly Asn Ser Gly Gly Gln Gly Gly Thr Pro Arg Ala Val Pro	
660 665 670	
ggc agg gtg gtc gcg ctc ggc caa acc gtg cag gcg tcg gat tcg ctg	2064
Gly Arg Val Val Ala Leu Gly Gln Thr Val Gln Ala Ser Asp Ser Leu	
675 680 685	
acc ggt gcc gaa gag aca ttg aac ggg ttg atc cag ttc gat gcc gcg	2112
Thr Gly Ala Glu Glu Thr Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala	
690 695 700	
atc cag ccc ggt gat tcg ggc ggg ccc gtc gtc aac ggc cta gga cag	2160
Ile Gln Pro Gly Asp Ser Gly Gly Pro Val Val Asn Gly Leu Gly Gln	
705 710 715 720	
gtg gtc ggt atg aac acg gcc gcg tcc tag g	2191
Val Val Gly Met Asn Thr Ala Ala Ser	
725 730	

<210> 12

<211> 729

<212> PRT

<213> Sequência Artificial

<223> Descrição da Sequência Artificial: polipéptido de fusão
Ra12-H9-32A (polipéptido de fusão Ra12-MTB39-MTB32A(N-ter))

<400> 12

Met	His	His	His	His	His	His	Thr	Ala	Ala	Ser	Asp	Asn	Phe	Gln	Leu
1				5					10					15	
Ser	Gln	Gly	Gly	Gln	Gly	Phe	Ala	Ile	Pro	Ile	Gly	Gln	Ala	Met	Ala
		20						25					30		
Ile	Ala	Gly	Gln	Ile	Arg	Ser	Gly	Gly	Gly	Ser	Pro	Thr	Val	His	Ile
		35					40					45			
Gly	Pro	Thr	Ala	Phe	Leu	Gly	Leu	Gly	Val	Val	Asp	Asn	Asn	Gly	Asn
	50					55					60				
Gly	Ala	Arg	Val	Gln	Arg	Val	Val	Gly	Ser	Ala	Pro	Ala	Ala	Ser	Leu
	65				70					75					80
Gly	Ile	Ser	Thr	Gly	Asp	Val	Ile	Thr	Ala	Val	Asp	Gly	Ala	Pro	Ile
				85					90					95	
Asn	Ser	Ala	Thr	Ala	Met	Ala	Asp	Ala	Leu	Asn	Gly	His	His	Pro	Gly
			100					105					110		
Asp	Val	Ile	Ser	Val	Thr	Trp	Gln	Thr	Lys	Ser	Gly	Gly	Thr	Arg	Thr
	115						120					125			
Gly	Asn	Val	Thr	Leu	Ala	Glu	Gly	Pro	Pro	Ala	Glu	Phe	Met	Val	Asp
	130					135					140				
Phe	Gly	Ala	Leu	Pro	Pro	Glu	Ile	Asn	Ser	Ala	Arg	Met	Tyr	Ala	Gly
145					150					155					160
Pro	Gly	Ser	Ala	Ser	Leu	Val	Ala	Ala	Ala	Gln	Met	Trp	Asp	Ser	Val
				165					170					175	
Ala	Ser	Asp	Leu	Phe	Ser	Ala	Ala	Ser	Ala	Phe	Gln	Ser	Val	Val	Trp
		180						185					190		
Gly	Leu	Thr	Val	Gly	Ser	Trp	Ile	Gly	Ser	Ser	Ala	Gly	Leu	Met	Val
	195						200					205			
Ala	Ala	Ala	Ser	Pro	Tyr	Val	Ala	Trp	Met	Ser	Val	Thr	Ala	Gly	Gln
	210					215					220				
Ala	Glu	Leu	Thr	Ala	Ala	Gln	Val	Arg	Val	Ala	Ala	Ala	Tyr	Glu	
225					230					235				240	
Thr	Ala	Tyr	Gly	Leu	Thr	Val	Pro	Pro	Pro	Val	Ile	Ala	Glu	Asn	Arg
				245					250					255	
Ala	Glu	Leu	Met	Ile	Leu	Ile	Ala	Thr	Asn	Leu	Leu	Gly	Gln	Asn	Thr
		260						265					270		
Pro	Ala	Ile	Ala	Val	Asn	Glu	Ala	Glu	Tyr	Gly	Glu	Met	Trp	Ala	Gln
	275					280						285			
Asp	Ala	Ala	Ala	Met	Phe	Gly	Tyr	Ala	Ala	Ala	Thr	Ala	Thr	Ala	Thr
	290					295					300				
Ala	Thr	Leu	Leu	Pro	Phe	Glu	Glu	Ala	Pro	Glu	Met	Thr	Ser	Ala	Gly
305					310					315					320
Gly	Leu	Leu	Glu	Gln	Ala	Ala	Ala	Val	Glu	Glu	Ala	Ser	Asp	Thr	Ala
				325					330					335	

Ala	Ala	Asn	Gln	Leu	Met	Asn	Asn	Val	Pro	Gln	Ala	Leu	Gln	Gln	Leu	340	345	350
Ala	Gln	Pro	Thr	Gln	Gly	Thr	Thr	Pro	Ser	Ser	Lys	Leu	Gly	Gly	Leu	355	360	365
Trp	Lys	Thr	Val	Ser	Pro	His	Arg	Ser	Pro	Ile	Ser	Asn	Met	Val	Ser	370	375	380
Met	Ala	Asn	Asn	His	Met	Ser	Met	Thr	Asn	Ser	Gly	Val	Ser	Met	Thr	385	390	400
Asn	Thr	Leu	Ser	Ser	Met	Leu	Lys	Gly	Phe	Ala	Pro	Ala	Ala	Ala	Ala	405	410	415
Gln	Ala	Val	Gln	Thr	Ala	Ala	Gln	Asn	Gly	Val	Arg	Ala	Met	Ser	Ser	420	425	430
Leu	Gly	Ser	Ser	Leu	Gly	Ser	Ser	Gly	Leu	Gly	Gly	Gly	Val	Ala	Ala	435	440	445
Asn	Leu	Gly	Arg	Ala	Ala	Ser	Val	Gly	Ser	Leu	Ser	Val	Pro	Gln	Ala	450	455	460
Trp	Ala	Ala	Ala	Asn	Gln	Ala	Val	Thr	Pro	Ala	Ala	Arg	Ala	Leu	Pro	465	470	480
Leu	Thr	Ser	Leu	Thr	Ser	Ala	Ala	Glu	Arg	Gly	Pro	Gly	Gln	Met	Leu	485	490	495
Gly	Gly	Leu	Pro	Val	Gly	Gln	Met	Gly	Ala	Arg	Ala	Gly	Gly	Gly	Leu	500	505	510
Ser	Gly	Val	Leu	Arg	Val	Pro	Pro	Arg	Pro	Tyr	Val	Met	Pro	His	Ser	515	520	525
Pro	Ala	Ala	Gly	Asp	Ile	Ala	Pro	Pro	Ala	Leu	Ser	Gln	Asp	Arg	Phe	530	535	540
Ala	Asp	Phe	Pro	Ala	Leu	Pro	Leu	Asp	Pro	Ser	Ala	Met	Val	Ala	Gln	545	550	555
Val	Gly	Pro	Gln	Val	Val	Asn	Ile	Asn	Thr	Lys	Leu	Gly	Tyr	Asn	Asn	565	570	575
Ala	Val	Gly	Ala	Gly	Thr	Gly	Ile	Val	Ile	Asp	Pro	Asn	Gly	Val	Val	580	585	590
Leu	Thr	Asn	Asn	His	Val	Ile	Ala	Gly	Ala	Thr	Asp	Ile	Asn	Ala	Phe	595	600	605
Ser	Val	Gly	Ser	Gly	Gln	Thr	Tyr	Gly	Val	Asp	Val	Val	Gly	Tyr	Asp	610	615	620
Arg	Thr	Gln	Asp	Val	Ala	Val	Leu	Gln	Leu	Arg	Gly	Ala	Gly	Gly	Leu	625	630	635
Pro	Ser	Ala	Ala	Ile	Gly	Gly	Gly	Val	Ala	Val	Gly	Glu	Pro	Val	Val	645	650	655
Ala	Met	Gly	Asn	Ser	Gly	Gly	Gln	Gly	Gly	Thr	Pro	Arg	Ala	Val	Pro	660	665	670
Gly	Arg	Val	Val	Ala	Leu	Gly	Gln	Thr	Val	Gln	Ala	Ser	Asp	Ser	Leu	675	680	685
Thr	Gly	Ala	Glu	Glu	Thr	Leu	Asn	Gly	Leu	Ile	Gln	Phe	Asp	Ala	Ala	690	695	700
Ile	Gln	Pro	Gly	Asp	Ser	Gly	Gly	Pro	Val	Val	Asn	Gly	Leu	Gly	Gln	705	710	715
Val	Val	Gly	Met	Asn	Thr	Ala	Ala	Ser								725		

<210> 13

<211> 51

<212> ADN

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: iniciador de oligonucleótido para amplificação por PCR do fragmento do terminal C de MTB32A

<400> 13

caattacata tgcatacaca tcaccatcac acggccgcgt ccgataactt c 51

<210> 14

<211> 33

<212> ADN

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: iniciador de oligonucleótido 3' para amplificação por PCR do fragmento do terminal C Ra12-C de MTB32A

<400> 14

ctaatacgaat tcggccgggg gtccctcggc caa 33

<210> 15

<211> 48

<212> ADN

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: iniciador de oligonucleótido 5' contendo o sítio de reconhecimento da enterocinase para amplificação por PCR da forma segregada madura de DPPD

<400> 15

caattagaat tcgacgacga cgacaaggat ccacctgacc cgcatcag 48

<210> 16

<211> 33

<212> ADN

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: iniciador de oligonucleótido 3' contendo o sítio de reconhecimento da enterocinase para amplificação por PCR da forma segregada madura de DPPD

<400> 16

caattagaat tctcagggag cgttgggctg ctc 33

<210> 17

<211> 30

<212> PRT

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: polipéptido Ra12(curto)

<400> 17

Thr	Ala	Ala	Ser	Asp	Asn	Phe	Gln	Leu	Ser	Gln	Gly	Gly	Gln	Gly	Phe
1				5					10					15	
Ala	Ile	Pro	Ile	Gly	Gln	Ala	Met	Ala	Ile	Ala	Gly	Gln	Ile		
			20					25					30		

<210> 18

<211> 128

<212> PRT

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: polipeptido Ra12(longo)

<400> 18

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

<210> 19

<211> 33

<212> ADN

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: iniciador de oligonucleótido 5', sítio *Hind*III, para amplificação por PCR da mamoglobina humana

<400> 19

gcgaagctta tgaagttgct gatggctcctc atgc

34

<210> 20

<211> 36

<212> ADN

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: iniciador de oligonucleótido 3', sítio *Xho*I, para amplificação por PCR da mamoglobina humana

<400> 20

cggctcgagt taaaataaat cacaagact gctgtc

36

<210> 21

<211> 7

<212> PRT

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: Met-His tag 6aa

<400> 21

Met His His His His His His
1 5

<210> 22

<211> 4

<212> PRT

<213> Sequência Artificial

<220>

<223> Descrição da Sequência Artificial: sítio de reconhecimento
de enterocinase

<400> 22

Asp Asp Asp Lys
1

Lisboa, 16 de Maio de 2007

REIVINDICAÇÕES

1. Método para a expressão estável e rendimento elevado de um polipéptido heterólogo recombinante, o método compreendendo expressar, numa célula hospedeira, uma molécula de ácido nucleico recombinante que codifica um polipéptido de fusão, compreendendo o polipéptido de fusão um polipéptido Ral2 e um polipéptido heterólogo, em que o polipéptido Ral2 é codificado por uma sequência de polinucleótido Ral2 que hibrida com uma sequência complementar à SEQ ID N°:3 em condições de restringência e em que a sequência de polinucleótido Ral2 está localizada a 5' da sequência do polinucleótido heterólogo que codifica o polipéptido heterólogo.
2. Método de acordo com a reivindicação 1, em que o polipéptido de fusão compreende ainda uma etiqueta de afinidade que está ligada ao polipéptido de fusão.
3. Método de acordo com a reivindicação 1 ou reivindicação 2, em que o polipéptido de fusão é purificado a partir da célula hospedeira.
4. Método de acordo com qualquer uma das reivindicações 1 a 3, em que a célula hospedeira é *E. coli*.
5. Método de acordo com qualquer das reivindicações 1 a 4, em que a molécula de ácido nucleico recombinante compreende ainda uma sequência de polinucleótido que codifica um péptido ligando entre a sequência de polinucleótido Ral2 e a sequência de polinucleótido heterólogo.

6. Método de acordo com a reivindicação 5, em que o péptido ligando compreende um sítio de clivagem.
7. Método de acordo com a reivindicação 1, em que a sequência de ácido nucleico heterólogo codifica um DPPD, um Wt1, uma mamoglobina ou um polipéptido H9-32A.
8. Método de acordo com qualquer uma das reivindicações 1 a 7, em que a sequência de polinucleótido Ral2 compreende pelo menos, 30 nucleótidos.
9. Método de acordo com a reivindicação 8, em que a sequência de polinucleótido Ral2 compreende, pelo menos, 60 nucleótidos.
10. Método de acordo com a reivindicação 8, em que a sequência de polinucleótido Ral2 compreende, pelo menos, 100 nucleótidos.
11. Método de acordo com qualquer uma das reivindicações 1 a 10, em que a sequência de polinucleótido Ral2 codifica um polipéptido Ral2 como apresentado em SEQ ID N°:17.
12. Método de acordo com qualquer uma das reivindicações 1 a 10, em que a sequência de polinucleótido Ral2 codifica um polipéptido Ral2 como apresentado em SEQ ID N°:18.
13. Método de acordo com qualquer uma das reivindicações 1 a 10, em que a sequência de polinucleótido Ral2 é como apresentado na SEQ ID N°:3.

14. Método de acordo com qualquer uma das reivindicações 1 a 10, em que a sequência de polinucleótido Ral2 codifica um polipéptido Ral2 como apresentado em SEQ ID N°:4 ou uma variante possuindo, pelo menos, 90% de identidade de sequência de aminoácidos com esta.
15. Método de acordo com a reivindicação 14, em que a sequência de polinucleótido Ral2 codifica um polipéptido Ral2 como apresentado em SEQ ID N°:4 ou uma variante possuindo, pelo menos, 95% de identidade de sequência de aminoácidos com esta.
16. Método de acordo com qualquer uma das reivindicações 1 a 7, em que o polipéptido Ral2 compreende, pelo menos, 10 aminoácidos.
17. Método de acordo com a reivindicação 16, em que o polipéptido Ral2 compreende, pelo menos, 30 aminoácidos.
18. Método de acordo com a reivindicação 16, em que o polipéptido Ral2 compreende, pelo menos, 100 aminoácidos.
19. Método de acordo com qualquer uma das reivindicações 1 a 7, em que o polipéptido Ral2 possui uma sequência como apresentado na SEQ ID N°:4.
20. Método de acordo com qualquer uma das reivindicações 1 a 7, em que o polipéptido Ral2 possui uma sequência como apresentado na SEQ ID N°:17.

21. Método de acordo com qualquer uma das reivindicações 1 a 7, em que o polipéptido Ra12 possui uma sequência como apresentado na SEQ ID N°:18.
22. Método de acordo com qualquer uma das reivindicações 1 a 21, compreendendo ainda o método clivar o polipéptido de fusão entre o polipéptido Ra12 e o polipéptido heterólogo.
23. Utilização de uma sequência de polinucleótido Ra12 que hibrida com uma sequência complementar à SEQ ID N°:3 em condições de restringência, num método para a expressão estável e rendimento elevado de um polipéptido heterólogo recombinante, compreendendo o método expressar, numa célula hospedeira, uma molécula de ácido nucleico recombinante que codifica um polipéptido de fusão, compreendendo o polipéptido de fusão um polipéptido Ra12 codificado pelo referido polinucleótido Ra12 e um polipéptido heterólogo, em que a referida sequência de polinucleótido Ra12 está localizada a 5' da sequência de polinucleótido heterólogo que codifica o polipéptido heterólogo.

Lisboa, 16 de Maio de 2007

GACTACGTTGGTGTAGAAAAATCCTGCCGCCCGGACCCCTTAAGCGTGGGACAATTTCGATAGCTACCCCGACACAGGAGTTACGGGATGAGCA	95
SEQUÊNCIA SINAL MTB32A	
N S R R R S L R W S W L L S V L A A V G L G L A T A P A Q N A P	190
CGGGCCTTGTGCGAGGACCGGTTCCGCCACTTCCCGCGCTGCCCTCGACCCGTCGGGATGGTCCGCCAAGTGGGGCCACAGGTGGTCAACAT	285
SEQUÊNCIA SINAL MTB32A	
P A L S Q D R F A D P F A L P L D P S A M V A Q V G P Q V V N I	380
CAACACCAAACCTGGGCTACAACAACCGCGTGGGGCGCGGACCGGCATCGTCATCGATCCCAACGGTSPCGTCTGACCAACACCAAGTGTATCG	475
SEQUÊNCIA SINAL MTB32A	
N T K I G Y N N A V G A G T G I V I D P N G V V L T N N H V I	570
CGGGCGCCACCGACATCAATGCGTTACGCGTCCGGCTCCGGCCAAACCTACCGCGTCGATGTGCTCGGGTATGACCGCACCCAGGATGTCGCGGTG	665
SEQUÊNCIA SINAL MTB32A	
A G A T D I N A F S V G S G Q T Y G V D V V G Y Q R T Q D V A V	760
CTGCAAGTGCSCGGTSCCGGTGGCCTGCCCTCGGCGGCGATCGGTGGCGGCGTCCGCGTTGCTGAGCCGCTCGTCCGATGGGCAACACCGGTGG	
SEQUÊNCIA SINAL MTB32A	
L Q L R G A F F L P S A A I G G G V A V G E P V V A N G N S G G	
GCAGGGCGGAACGCCCGTGCCTGGCAGGTTGGTCCGCTCCGGCCAAACCGTGCAGCGCTCGGATTGCTGACCGGTGCCGAAGAGACAT	
SEQUÊNCIA SINAL MTB32A	
Q G G T P R A V P G R V V A L G Q T V Q A S D S L T G A E E T	
TGAACGGGTTGATCCAGTTGATGCCGCGATCCAGCCCGGTGATTCCGGCGGGCCCGTCTCAACGCGCTAGGACAGGTGGTCCGTTATGAACACG	
SEQUÊNCIA SINAL MTB32A	
L N G L I Q F D A A I Q P G D S G G P V V N G L G Q V V G M N T	

```

GCGCGCTCCGATAACTTCCAGCTGTCCCAAGGTTGGGCAGGGATTCCGCATTCCGATCGGGCAGGCSATGSCGATCCCGGGCCAGATCCGATCGGG 855
-----MTB322A-----
A A S D N F Q L S Q G G Q G F A I P I G Q A M A I A G Q I R S G
TGGGGGGTCAOCCACCGTTCATATCGSGGCTACCGCCTTCCTCGGCTTGGGTGTGTGTCGACAACAACSGCAACGGCGCAGGAGTCCAACCGCTGG 950
-----MTB322A-----
G G S P T V H I G P T A F L G L G V V D N N G N F A R V Q R V
TCSGGAGCGCTCCGGCGGCAAGTCTCGGCAATCTCCACCGCGGACGTSATCACCGCGGTCGACGCGCTCCGATCAACTCGGCCACCGCGATGGCG 1045
-----MTB322A-----
V G S A P A A S L G I S T G D V I T A V D G A P I N S A T A M A
GACCGCCTTAACGGGCATCATCCCGGTGACGTCACTCTCGGTGACCTGGCAAAACCAAGTCGSGCGGCACCGGTACAGGGAACGTGACATTGGCCGA 1140
-----MTB322A-----
D A L N G H H P G D V I S V T W Q T K S G G T R T G N V T L A
GGGACCCCGGGCCTGATTTCCGTGCGGATACACCCCGCGGCGGSCAAATTGGATTGGCGCCAGCCGTGATTGCCCGCTGAGCCCGGAGTTCCG 1235
----MTB322A-->
E G P P A
TCTCCCGTGGCGCTGGCATCGTGGAGCAATGAACGAGGCGAGAACACAGCGTCGAGCACCCTCCCGTGCAGGSCAGTCACSTCGAAGGCGGTGTG 1330
GTCGAGCATCCGATGSCCAAGGACTTCGGGAGCGCGCGCGCCCTGCCCCCGGATCCGACCTGGTTAAGCAGCGCGTCTTCTACGAGGTGCTGGT 1425
CCSGGCGTTCTTCGACGCCAGCGCGGACGGTTCGGGCGATCTGCGTGGACTCATCGATCGCTCGACTACCTSCAGTGGCTTGGCATCGACTGCA 1520
TCTGGTTCGCGCGCTTCTACGACTCSCCGGTGCGGACGCGGTTACGACATTCGCGACTTCTACAAGGTGCTGCCCGAATTCGGCACCGTTCGAC 1615
GATTTCCGTGCGCCTGGTTCGACCGCGCTCACCGGAGGATTCGCGATCATCACCGACCTGGTGAATGAAATCACACCTCGGAGTCGCAOCCCTSGTT 1710
TCAGGAGTCCCGCGCGACCCAGACGACCGTACGGTGACTATTACGTGTGAGCGCACACCGAGCGGTACACCGACGCCCGATCATCTTCG 1805
TCGACACCGAAGAGTCGAACTGGTCATTTCGATCCTGTCCCGGACACTTCTACTGGCACCGATTCTT 1872

```

FIG. 1B.

```

ACGSCCGCGTCCGATAACTTCCAGCTGTCCAGGGTGGGCAGGGATTCCGCATTCCGATCGGGCAGGCGATGGCGA
-----Ra12-----
T A A S D N F Q L S Q G G Q G F A I P I G G A M A

TCGCGGGCCAGATCCGATCGSGTGGGGGGTCACCCACCGTTCATATCGSGCCTACCGCCTTCCTCGGCTTGGGTGTTGTGACACAACAGGCACCGGCGC
-----Ra12-----
I A G Q I R S G G G S P T V H I G P T A F L G L G V V D N N G N G A

ACGAGTCCAACGCGTGGTGGGGAGCGTCCGGCGGCAAGTCTCGGCATCTCCACGGCGACGTGATCACCAGCGTCCGACGGCGCTCCGATCAACTCGGCC
-----Ra12-----
R V Q R V V G S A P A A S L G I S Y G D V I T A V D G A P I N S A

ACCGCGATGGCGGACGCGTTAACGGGCATCATCCGGGTGACGTCACTCTCGGTGACCTGGCAACCAAGTCGGGGGGCACGCGTACAGGGAACGTGACAT
-----Ra12-----
T A M A D A L N G H H P G D V I S V T W Q T K S G G T R T G N V T

TGGCCGAGGGACCCCGGCC
-----Ra12-----
L A E G P F A

```

FIG. 2.

```

CATATGCATCACCATCACCATCACACGCCCGTCCGATAACTTCCAGCTGTCCAGGCTGGGCGAGGATTCGCCATTCCGATCGGCGAGGCGAT 95
|-----Met (6xHis)-----|-----Ra12-----
N N N H H H H T A A S D N F Q L S Q G G Q G F A I P I G Q A M

GGCSATCGCGGCCAGNTCCGATCGGGTGGGGGTCACCACCGTTTCATATCGGGCTACCGGCTTCCTCGGCTTGGGTGTGTGCGACAACAACG 190
-----Ra12-----
A I A G Q I R S G G G S P T V H I G P T A F L G L G V V D N N

GCCACGSCGACAGTCCAACGCGTGGTGGGGGGCTCCGGCGGCAAGTCTCGGCATCTCCACCGGCGACGTSATCACCGCGSTGACGCGCGCT 285
-----Ra12-----
G N G A R V Q R V V G S A P A A S L G I S T G O V I T A V D G A

CCCATCAACTCGGCCACCGGATGGCGGACGGCTTAACGGGCATCATCCGGTGACCTCATCTCGGTGACCTGGCAACCAAGTGGGGCGGCAC 380
-----Ra12-----
P I N S A T A M A D A L N G H H P G O V I S V T W Q T K S G G T

GGGTACAGGGAACGTGACATTGGCGAGGGACCCCGGCCGAATTGGACGACGACGACAAGGATCCACCTGACCCGCGATCAGCCGGACATGACGA 475
-----Ra12-----|EcoRI|Enterocinase|-----DPPD-----
R T G N V T L A E G F F A E F G D D D K D P P D P H Q P D M T

AAGGCTATTGCCGGTGGCCGATGGGGTTTTGGCGACTTGGCGGTGTCGAGCGGAGAGTACCCCGACGGCTCGTTTTGGCACCAAGTGGATG 570
-----DPPD-----
K G Y C P G G R W G F G D L A V C D G E K Y P D G S P W H Q W M

CARACGTGGTTTACGGGCCACAGTTTTACTTCGATTGTGTCAGCGCGGTGAGCCCTCCCGGGCCGCGCCACCGGGTGGTGGCGTGGGGC 665
-----DPPD-----
Q T W F T G P Q F Y F D C V S G G E F L P G F P P P G G C G G A

AATTCGCTCGAGCAGCCCAACGCTCCCTGAGGATTC 702
-----DPPD----->
I F S E Q P N A P

```

FIG. 3.

FIG. 4A.

FIG. 4A.

```

CAAACAGGAACCGAGCTGGGGTGGTGCAGAACCGCACGAAGACAGTGCCTGAGGGCATTACCGTTCACTTCTCCGSCCAGTTCACTGGCACAG 760
-----WT1-----
K Q E P S W G G A E P H E E Q C L S A F T V H F S G Q F T G T
CCGGAGCCTSTCGCTACGGGCCCTTCGGTCCCTCCGCCAGCCAGCGCTCATCCGSCCAGGCCAGCATGTTTCTTAACCGGCCCTACCTGCCCG 855
-----WT1-----
A G A C R Y G P F G P P P P S Q A S S G Q A R M F P N A F Y L P
AGCTGCCTCGAGAGCCAGCCCGCTATTCCGAATCAGGGTTACAGCACGGTCACCTTCGACSSGACGCCAGCTACGGTCACACGCCCTCGCACCA 950
-----WT1-----
S C L E S Q P A I R N Q G Y S T V T F D G T P S Y G H T F S H H
TGCGGCGCAGTTCCCAACCACTCATTCAAGCATGAGGATCCCATGGGCCAGCAGGGCTCGCTGGGTGAGCAGCAGTACTCGGTGCCGCCCGCCG 1045
-----WT1-----
A A Q E P M H S F K H E D P M G - Q G S L G E Q Q Y S V P P P
TCTATGGCTGCCACACCCCGACAGCTGCACCGGCAGCCAGGCTTTGCTGCTAAGGACGCCCTACAGCAGTGACAATTTATACCAATGACA 1140
-----WT1-----
V Y G C H T P T D S C T G S Q A L L L R T P Y S S, O N L Y Q M Y
TCCCAGCTTGAATGCAATGACCTGGGAATCAGATGAACCTTAGGAGCCACCTTAAAGGGCCACAGCACAGGGTACGAGAGCGATAACACACACACGCC 1235
-----WT1-----
S Q L E C M T W N Q M N L S A T L K G H S T G Y E S D N H T T P
CATCCTCTGCGGAGCCCAATACAGAATACACACGCCACGGTGTCTTCAGAGGGCATTACAGGATGTGCAACGTGTGCCTGGAGTAGCCCCGACTCTTG 1330
-----WT1-----
I L C G A Q Y R I H T H G V F R G I Q D V A R V P S V A P T L

```

FIG. 4B.

```

TACGGTCGGGATCTGAGACCAGTGAGAAACGCCCTTCATSTGTSCCTTACTCAGGCTGCAATAAGAGATAATTTAAGCTGTCCCACTTACAGATG 1425
-----WT1-----
V R S A S E T S E K R P F M C A Y S G C N K R Y F K L S H L Q M
CACAGCAGGAAGCACACTGGTGAGAAACCATACCACTGTGACTTCAAGGACGCTGAACGAAGGTTTTTTCGTTCAAACCAAGCTCAAAAGACACCA 1520
-----WT1-----
H S R K H T G E K P Y Q C D F N D C E R R F E R S D Q L K R H Q
AAGGAGACATACAGGTSTGAAACCATTCAGTGTAAACTTGTGACGGAAGTTCTCCCGGTCCGACCACTGAAGACCCACACCAGGACTCATA 1615
-----WT1-----
R R H T G V K P F Q C K T C Q R K F S R S D H L K T A T R T H
CAGTGAAGAGCCCTTCAGCTSTCGGTGSCCAAGTTGTGAGAAAAAGTTTGCCCGGTGAGATGAATTATGTCGSCCATCACAACATACATCAGGA 1710
-----WT1-----
T G E K P F S C R W P S C Q K K F A R S D E L V R H H N M H Q R
AACATGACCAAACTCCAGCTGCCGCTTTGAGAAATTC
-----WT1-----→
N M T K L Q L A L

```

1746

FIG. 4C.

```

CATATGCATCACCATCACCATCACACGGCCGCGTCCGATAACTTCCAGCTGTCCCAGGGTGGCCAGGATTCGCCATTCCGATCGGCCAGGCGAT 95
|-Met (5xHis)-----||-----Ra12-----
M H H R H H H T A A S D N F Q L S Q G G Q G F A I P I G Q A M

GGCGATCGCGGGCCAGATCCGATCGGGTGGGGGTCAACCCACCGTTCATATCGGGCCCTACCGCCTTCCTCGGCTTGSGTGTTCGACAAACCG 190
-----Ra12-----
A I A G Q I R S G G G S P T V H I G P T A P L G L G V V D N N

GCAACGGCGCAGAGTCCAACSCGTSGTCCGGAGCGTCCGGCGGCAAGTCTCGGCATCTCCACCGGCGAGTGATCACCGCGGTGACAGGCGCT 285
-----Ra12-----
G N G A R V Q R V V G S A P A A S L G I S T S Q V I T A V D G A

CCGATCAACTCGGCCACCGCGATGGCGGACCGGCTTAACGGGCATCATCCGGTGACGTTCATCTCGGTGACCTGGCAACCAAGTCGGGCGGCAC 380
-----Ra12-----
P I N S A T A M A D A L N G H H P G O V I S V T W Q T K S G G T

GCGTACAGGGAACGTGACATTCGCCSAGGGACCCCGGCCGAATTCATCGAGGGAAGGGCTCTGGCTGCCCTTATTGGAGAATGTGATTTC 475
-----Ra12-----|EcoR|Factor Xa|-----MAMAGLOBINA
R T G N V T L A E G P P A E F I E G R G S G C P L L E N V I S

AGACAATCAATCCACAAGTGTCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTTCATACAGSACAATGCCACTACAAATGCCATAGTGAATTG 570
-----MAMAGLOBINA-----
K T I N F Q V S K T E Y K E L L Q E F T D D N A T T N A I D E L

AAGGAATGTTTTCTTAACCAACGGGTGAAGACTCTGAGCAATGTTGAGGTTTATGCAATTAAATATGACAGCAGTCTTTGTGATTATTTTA 685
-----MAMAGLOBINA----->
K E C F L N Q T D E T L S N V E V F M Q L I Y D S S L C D L F

AGAATTC 672

```

FIG. 5.

```

ATCCATCACCATCACCATCACACGGCCGCTCCGATACCTCCAGCTGTCCCAGGGTGGGAGGGATTCCGCCATTCGGATCGGGCAGGGGATGGC 95
|----Met (6xHis)-----||-----Ra12-----|
M H H H H H H T A A S D N F Q L S Q G G Q G F A I P I G Q A M A

GATCGCGGGCCAGATCCGATCGGGTGGGGSTCACCCACCGTTCATATCGGGCCTACCGCCTTCCTCGGCTTGGGTGTTCTCGACAACAACGGCA 190
-----Ra12-----
I A G Q I R S G G G S P T V H I G P T A F L G L G V V D N N G

ACGGGCGACGAGTCCACGCGSTGGTCGGGAGCGGTCGCGCGGCAAGTCTCGGCATCTCCACCGGCGAGTATCACCGCGGTCCGACGGCGCTCCG 285
-----Ra12-----
N G A R V Q R V V G S A P A A S L G I S T G O V I T A V D G A P

ATCAACTCGGCCACCGCATGGCGGACCGCTTAACGGGCATCATECCGGTGACGTCATCTCGGTGACCTGGCAAACCAAGTCCGGCGGCACGCG 380
-----Ra12-----
I N S A T A M A D A L N G H H P G G V I S V T W Q T K S G G T R

TACAGGGAACGTGACATTGGCCGAGGACCCCGCGCGAATTTCATGGTGGATTTCGGGGCGGTTACCAACCGGAGATCAACTCCGGCAGGATGTACG 475
-----Ra12-----EcoRI-----MTB39-----
T G N V T L A E G P P A E F M V D F G A L P P E I N S A R M Y

CCGGCCCCGGTTTCGGCCTCGCTGGTCCGCGGGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGTTTTCCGGCCGCGTCGGCGTTTCAGTCCGTG 570
-----MTB39-----
A G P G S A S L V A A A Q N W D S V A S D L F S A A S A F Q S V

GTCTGGGGTCTGACGGTGGGGTCGTGGATAGGTTGCTCGCGGGGCTGATGGTGGCGGGCCCTCGCCGTATGTGGCGTGSATGAGCGTCACCGC 665
-----MTB39-----
V W G L T V G S W I G S S A G L M V A R A S F Y V A W M S V T A

GGGGCAGGCCGAGCTGACCGCGCCAGGTCCGGCTTGTGCGCGGGCTACGAGACGGCGTATGGGCTGACGTTGCCCGCCGCGGTGATCGCCG 760
-----MTB39-----
G Q A E L T A A Q V R V A A A Y E T A Y G L T V P P P V I A

```

FIG. 6A.

AGAAACCGTGGTGAAGTGGATGATTTCTATAGCGACCAACCTCTTGGGGGCAAAACACCCCGGCGATCGCGGTCAACCGAGGCCGAATACGGCGAGATG 855
-----MTB39-----
E N R A E L M I L I A T N L L G Q N T P A I A V N E A E Y G E M
TGGGCCCAAGACGCCGCCCGATGTTTGGCTACGCCGCGGCGACGCGCGACGCGCGAGCTGTGCTGCCGTTCCGAGGAGCGCGCCGAGATGAC 950
-----MTB39-----
W A Q D A A A M F G Y A A A T A T A T A T L L P F E E A P E M T
CAGCGCGGCTGGGCTCTCGAGCAGGCCGCCGCGCTCGAGSAGGCGCTCGACACCGCGCGCGCAACCACTGTGATGAACAAATGTGCCCGAGGCGC 1045
-----MTB39-----
S A G G L L E Q A A A V E E A S D T A A A N Q L M N N V P Q A
TCAACAGCTGSGCCCGACCCAGCGAGGGCACCAGCCCTCTTCCAAAGCTGGGTGGCTGTGGAACACGGTCTCGCGGCATCGGTGCGCGATCAGC 1140
-----MTB39-----
L Q Q L A Q P T Q G T T P S S K L G G L W K T V S P H R S P I S
AACATGGTGTGATGGCCCAACCAACCATGTCGATGACCAACACCGGCTGTGTCGATGACCAACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCC 1235
-----MTB39-----
N M V S M A N N N H M S M T N S G V S M T N T L S S M L K G F A P
GGCGGCGCGCCCGCCAGGCGCTGCAAAACCGCGGCGCAAAACGGGGTCCGSGCGATGAGCTCGCTGGGCACTCGCTGGGTTCTTCCGGGCTCTGGGGC 1330
-----MTB39-----
A A A A Q A V Q T A A Q N G V R A M S S L G S S L G S S G L G
GTGGGGTGGCGGCCAACTTGGGTGGGGCGGCTCGGTGGGTTGGTTGTGGGTGCCGAGAGCTGGGCGCGGCCCAACAGGCGATCACCCTGGGGC 1425
-----MTB39-----
G G V A A N L G R A A S V G S L S V P Q A W A A A N Q A V T P A
GGGCGGGCGCTGCGCTGACCAAGCTGACCAAGCGCGCGGAAAGAGGCCCGGGCGAGATGCTGGGCGGGCTCGCGGTGGGCGAGATGGGCGCCAG 1520
-----MTB39-----
A R A L P L T S L T S A A E R G P G Q M L G G L P V G D M G A R

FIG. 6B.

```

GECGGGTGGTGGGCTCAGTGGTGTGCTGGGTGTTCCGCCGCCGACCTATGTGATGCCGCATTCTCCGGCAGCCGGCGATATCGCCCCGCCGGGCT 1615
-----MTB32A (N-ter)-----Eco-----
A G G S L S G V L R V P P R P Y V M P H S P A A G D I A P P A
TGTCGGCAGGACCGGTTCGCCGACTTCCCGSCGCTGCCCTCGACCCGTCGCCGATGGTCGCCCAAGTGGGGCCACAGGTGGTCAACATCAACACC 1710
-----MTB32A (N-ter)-----
L S Q D R F A D P P A L P L D P S A N V A Q V G P Q V V N I N T
AAACTGGGCTACACACGCGCTGGGGCGCCGGGACCGGCATCGTCATCGATCCCAACGGTGTGTGTGACCAACAACACACGTGATCGCGGGCGC 1805
-----MTB32A (N-ter)-----
K L G Y N N A V G A G T G I V I E P N C V V L T N N H V I A G A
CACCGACATCAATGCCGTCAGCGTCGGCTCCGCCCAACCTACGGCGTCGATGTGGTGGTGGTATGACCGCACCCAGGATGTGCGCGGTGCTGCAGC 1900
-----WT1-----MTB32A (N-ter)-----
T D I N A F S V G S G Q T Y G V D V V G Y D R T Q D V A V L Q
TGCGCGGTGGCGGTGGCTGCCGTCGGCGCGATCGGTGGCGCGCTCGCGGTTGGTGAGCCCGTCTGTCGCGATGGGCAACAGCGGTGGGCGAGGGC 1995
-----MTB32A (N-ter)-----
L R G A G G L P S A A I G G V A V G E P V V A M G N S G G Q G
GGAACGCCCCGTGCGGTGCCGTGGCAGGGTGGTGGCGTTCGGGCCAACCCTGCAGGCGTCCGATTCGCTGACCGGTGCCGAAGAGACATTGAACGG 2090
-----MTB32A (N-ter)-----
G T P R A V P G R V V A L S Q T V Q A S D S L T G A E E T L N G
GTTGATCCAGTTCGATGCCGCGATCCAGCCCGGTGATTCCGGCGGGGCCCGTCTCAACGGCCCTAGGACAGGTGGTCCGGTATGAACACCGCCCGGT 2185
-----WT1-----MTB32A (N-ter)-----
L I Q P D A A I Q P G D S G G P V V N G L G Q V V G M N T A A
CCTAGG
→
S

```

2191

FIG. 6C.

polipéptido Ra12(curto) (SEQ ID N°:17)

TAASDNFQLSQGGQGFAIPIGQAMAIAGQI

FIG. 7.

polipéptido Ra12(longo) (SEQ ID N°:18)

TAASDNFQLSQGGQGFAIPIGQAMAIAGQIKLPTVHIGPTAFLGLGVVDNNGNGARV
 QRVVGSAPAASLGISTGDVITAVDGAPINSATAMADALNGHHPGDVISVTWQTKSG
 GTRTGNTLAEGPPA

FIG. 8.

H ₃ N -	Met	Histag6aa	Ra12(curto) 30aa	HindIII 2aa	Mamaglobina humana (comprimento total) 93aa	-C00
--------------------	-----	-----------	------------------	-------------	--	------

FIG. 9.