

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

(22) Data de pedido: **2001.06.20**

(30) Prioridade(s): **2000.06.20 US 597796**
2001.02.01 US 265737 P

(43) Data de publicação do pedido: **2005.06.22**

(45) Data e BPI da concessão: **2009.09.16**
216/2009

(73) Titular(es):

CORIXA CORPORATION
CSC THE UNITED STATES CORPORATION 2711
CENTERVILLE ROAD WILMINGTON DE 19808
US

(72) Inventor(es):

YASIR SKEIKY **US**
STEVEN REED **US**
MARK ALDERSON **US**

(74) Mandatário:

ELSA MARIA MARTINS BARREIROS AMARAL CANHÃO
RUA DO PATROCÍNIO 94 1399-019 LISBOA **PT**

(54) Epígrafe: **PROTEÍNAS DE FUSÃO DE MYCOBACTERIUM TUBERCULOSIS**

(57) Resumo:

DESCRIÇÃO

"PROTEÍNAS DE FUSÃO DE *MYCOBACTERIUM TUBERCULOSIS*"

A presente invenção refere-se a proteínas de fusão que contêm, pelo menos, dois antigénios de *Mycobacterium* sp. Em particular, esta refere-se a ácidos nucleicos codificando proteínas de fusão que incluem dois ou mais antigénios individuais de *M. tuberculosis*, os quais aumentam a sensibilidade serológica de soros de indivíduos infectados com tuberculose e a sua utilização em diagnóstico, tratamento e prevenção da infecção da tuberculose.

ANTECEDENTES DA INVENÇÃO

A tuberculose é uma doença infecciosa crónica causada pela infecção com *M. tuberculosis* e outras espécies de *Mycobacterium*. Esta é uma doença grave em países em desenvolvimento, assim como é um problema crescente em áreas desenvolvidas do mundo, com cerca de 8 milhões de novos casos e 3 milhões de mortes cada ano. Embora a infecção possa ser assintomática durante um período de tempo considerável, a doença é mais frequentemente manifestada como uma inflamação aguda dos pulmões, resultando em febre e numa tosse não produtiva. Se não for tratada, resulta tipicamente em complicações graves e morte.

Embora a tuberculose possa ser, geralmente, controlada utilizando terapia prolongada com antibiótico, esse tratamento não é suficiente para prevenir a disseminação da doença. Os

indivíduos infectados podem ser assintomáticos, mas contagiosos, durante algum tempo. Além disso, embora a observância do regime de tratamento seja crucial, o comportamento do doente é difícil de monitorizar. Alguns doentes não concluem o período de tratamento, o que pode originar tratamento ineficaz e o desenvolvimento da resistência a fármacos.

A vacinação eficaz e o diagnóstico precoce exacto da doença são da maior importância para controlar a disseminação da tuberculose. Actualmente, a vacinação com bactérias vivas é o método mais eficaz para induzir imunidade protectora. A micobactéria mais comum utilizada com esta finalidade é o *Bacillus Calmette-Guerin* (BCG), uma estirpe avirulenta de *M. bovis*. No entanto, a segurança e a eficácia da BCG são uma fonte de controvérsia e alguns países, tal como os Estados Unidos, não vacinam o grande público com este agente.

O diagnóstico da tuberculose é normalmente realizado, utilizando um teste dérmico que envolve a exposição intradérmica a PPD (derivado proteico purificado) da tuberculina. As respostas de células T específicas para um antígeno resultam numa induração mensurável no local de injeção às 48-72 horas após a injeção, o que indica a exposição a antígenos micobacterianos. No entanto, a sensibilidade e a especificidade têm sido um problema com este teste e os indivíduos vacinados com o BCG não podem ser distinguidos de indivíduos infectados.

Embora os macrófagos tenham demonstrado actuar como os principais efectores da imunidade contra *Mycobacterium*, as células T são os indutores predominantes dessa imunidade. O papel essencial de células T na protecção contra a infecção com *Mycobacterium* é ilustrado pela ocorrência frequente de infecção

com *Mycobacterium* em doentes com SIDA, devido ao esgotamento de células T CD4⁺ associado à infecção com o vírus da imunodeficiência humana (HIV). As células T CD4⁺ reactivas a *Mycobacterium* demonstraram ser potentes produtoras de γ -interferão (IFN- γ), o qual, por sua vez, demonstrou provocar efeitos anti-micobacterianos de macrófagos em murganhos. Embora o papel do IFN- γ em humanos seja menos evidente, os estudos mostraram que a 1,25-di-hidroxi-vitamina D3, isoladamente ou em combinação com IFN- γ ou factor alfa da necrose tumoral, activa os macrófagos humanos para inibirem a infecção com *M. tuberculosis*. Além disso, é conhecido que o IFN- γ estimula os macrófagos humanos a produzirem 1,25-di-hidroxi-vitamina D3. De um modo semelhante, a interleucina 12 (IL-12) demonstrou desempenhar um papel no estímulo da resistência à infecção com *M. tuberculosis*. Para uma revisão da imunologia da infecção com *M. tuberculosis*, ver Chan e Kaufmann, *Tuberculosis: Pathogenesis, Protection and Control* (Bloom ed., 1994) e Harrison's Principles of Internal Medicine, volume 1, páginas 1004-1014 e 1019-1023 (14^a ed., Fauci et al., eds., 1998).

Consequentemente, verifica-se uma necessidade de reagentes de diagnóstico melhorados e de métodos melhorados para diagnóstico, prevenindo e tratando a tuberculose.

SUMÁRIO DA INVENÇÃO

Consequentemente, a presente invenção proporciona um polipéptido de fusão compreendendo (i) um antigénio MTB39 (SEQ ID N° 12 ou 14), uma variante do referido antigénio MTB39 tendo, pelo menos, de 90% de identidade este ou um fragmento imunogénico do referido antigénio MTB39 e (ii) um antigénio

MTB32A (SEQ ID N° 2 ou 4), uma variante do referido antigénio MTB32A tendo, pelo menos, 90% de identidade com este ou um fragmento imunogénico do referido antigénio MTB32A, **caracterizado por** o referido antigénio MTB32A, variante ou fragmento ter uma substituição de aminoácido correspondente à posição 183 da SEQ ID N° 4 ou à posição de aminoácido 208 da SEQ ID N° 2, em que o resíduo de serina normalmente encontrado nesta posição foi substituído por outro aminoácido.

Numa forma de realização, a presente invenção proporciona a proteína de fusão MTB72FMutSA, na qual o componente Ra35 da proteína de fusão tem uma mutação de serina em alanina na posição de aminoácido 710 da sequência MTB72F. Numa outra forma de realização, a presente invenção proporciona um ácido nucleico codificando a proteína de fusão MTB72F, na qual o ácido nucleico codificando o componente Ra35 foi mutado por alteração de uma T numa G, resultando numa mutação de serina em alanina na posição de aminoácido 710 da sequência MTB72F.

A presente invenção é baseada, em parte, na verificação da requerente de que os polinucleótidos de fusão ou os polipéptidos de fusão que contêm, pelo menos, duas sequências codificantes ou antigénios heterólogos de *M. tuberculosis* são altamente antigénicos e, após administração a um doente, aumentam a sensibilidade dos soros da tuberculose. Além disso, os polipéptidos e os polinucleótidos de fusão são úteis como instrumentos de diagnóstico em doentes que podem ter sido infectados com o *Mycobacterium*.

Os polipéptidos de fusão e os ácidos nucleicos da invenção são utilizados em ensaios *in vitro* e *in vivo* para detectar anticorpos humorais ou imunidade mediada por células contra

M. tuberculosis para o diagnóstico de infecção ou monitorização da progressão de doença. Por exemplo, os polipéptidos podem ser utilizados como um agente de diagnóstico *in vivo* sob a forma de um teste dérmico intradérmico. Os polipéptidos podem ser também utilizados em testes *in vitro*, tal como num ELISA com soro do doente. Alternativamente, os ácidos nucleicos e os polipéptidos de fusão podem ser utilizados para produzir anticorpos anti-*M. tuberculosis* num animal não-humano. Os anticorpos podem ser utilizados para detectar os antígenos alvo *in vivo* e *in vitro*.

Os polipéptidos de fusão e os ácidos nucleicos podem ser utilizados como imunogénios para produzir ou promover uma resposta imunitária protectora num doente. Os polinucleótidos isolados ou purificados são utilizados para produzir antígenos polipeptídicos de fusão recombinantes *in vitro*, os quais são então administrados como uma vacina. Alternativamente, os polinucleótidos podem ser administrados directamente num indivíduo como vacinas de ADN para promover a expressão de antígeno no indivíduo e a indução subsequente de uma resposta imunitária anti-*M. tuberculosis*. Deste modo, os polipéptidos de *M. tuberculosis* e os ácidos nucleicos da invenção, isolados ou purificados, podem ser formulados como composições farmacêuticas para administração a um indivíduo na prevenção e/ou no tratamento da infecção com *M. tuberculosis*. A imunogenicidade da proteína de fusão ou dos antígenos pode ser intensificada pela inclusão de um adjuvante, assim como de polipéptidos de fusão adicionais, de *Mycobacterium* ou de outros organismos, tais como polipéptidos bacterianos, virais, de mamífero. Podem ser também incluídos polipéptidos adicionais nas composições, ligados ou não ligados ao polipéptido de fusão.

BREVE DESCRIÇÃO DOS DESENHOS

A Figura 1 mostra a percentagem de sobrevivência de cobaios vacinados com a poliproteína MTB72F.

A Figura 2 mostra CFU de células do baço (Fig. 2A) e de células pulmonares após a imunização com ADN de MTB72F, MTB59F, MTB72F ou com uma composição compreendendo os antígenos Ra12, TbH9 e Ra35.

A Figura 3 mostra um diagrama esquemático do MTB72F.

A Figura 4 mostra a sequência de nucleótidos e de aminoácidos de Ra35 (195 aminoácidos da porção do terminal N de MTB32A).

A Figura 5 mostra um alinhamento das sequências de aminoácidos de MTB72F e a versão mutada de MTB72FMutSA.

A Figura 6 mostra um alinhamento das sequências de aminoácidos da Ra35/MTB32A madura (completa) e a versão mutada de Ra35FLMutSA.

A Figura 7 mostra a sobrevivência a longo prazo de cobaios vacinados com formulações Mtb72F.

DESCRIÇÃO DAS FORMAS DE REALIZAÇÃO ESPECÍFICAS

A presente invenção refere-se a polipéptidos de fusão úteis para o diagnóstico e tratamento de infecção com *Mycobacterium* e polinucleótidos codificando esses antígenos.

Os antígenos da presente invenção são polipéptidos de fusão de antígenos de *Mycobacterium* e seus fragmentos imunogénicos. Mais especificamente, um polipéptido de fusão compreendendo (i) um antígeno MTB39 (SEQ ID N° 12 ou 14), uma variante do referido antígeno MTB39 tendo, pelo menos, 90% de identidade com este ou um fragmento imunogénico do referido antígeno MTB39 ou variante e (ii) um antígeno MTB32A (SEQ ID N° 2 ou 4), uma variante do referido antígeno MTB32A tendo, pelo menos, 90% de identidade com este ou um fragmento imunogénico do referido antígeno MTB32A ou variante, **caracterizada por** o referido antígeno MTB32A, variante ou fragmento ter uma substituição de aminoácido correspondente à posição 183 da SEQ ID N° 4 ou à posição de aminoácido 208 da SEQ ID N° 2, em que o resíduo de serina normalmente encontrado nesta posição foi substituído por outro aminoácido.

Estes polipéptidos de fusão e os ácidos nucleicos que os codificam são, consequentemente, úteis para promover uma resposta protectora em doentes e em aplicações de diagnóstico.

Os antígenos da presente invenção podem ainda compreender outros componentes concebidos para intensificar a antigenicidade dos antígenos ou melhorar estes antígenos em outros aspectos, por exemplo, o isolamento destes antígenos através da adição de um prolongamento de resíduos de histidina numa extremidade do antígeno. Os polipéptidos de fusão e os ácidos nucleicos da

invenção podem compreender cópias adicionais de antígenos ou polipéptidos heterólogos adicionais de *Mycobacterium* sp., tais como antígeno MTB8.4, antígeno MTB9.8, antígeno MTB9.9, antígeno MTB40, antígeno MTB41, 38-1, TbRa3, 38 kD, DPEP, TbH4, DPPD, antígeno ESAT-6, antígeno do complexo MTB85 (e. g., MTB85b) ou antígeno α -cristalino e Erd14. Os polipéptidos de fusão e os ácidos nucleicos da invenção podem também compreender polipéptidos heterólogos adicionais provenientes de outras fontes não-*Mycobacterium*. Por exemplo, as proteínas de fusão da invenção podem incluir polipéptidos ou ácidos nucleicos codificando polipéptidos, em que o polipéptido intensifica a expressão do antígeno, e. g., NS 1, uma proteína do vírus influenza ou uma sua porção imunogénica (ver e. g., documentos W099/4018 e W093/04175). Os ácidos nucleicos da invenção podem ser concebidos com base na preferência de codão de uma espécie seleccionada, e. g., humanos.

As composições da invenção podem ser ADN nu ou as composições, e. g., os polipéptidos podem também compreender adjuvantes, e. g., MPL, 3D-MPL, IFA, adjuvantes AS, tais como AS2, AS2', AS2'', AS4, AS6, ENHANZYN (Detox), QS21, CWS, TDM, AGP, CPG, Leif, saponina e miméticos da saponina e os seus derivados. Além disso, as composições da invenção podem compreender BCG ou Pvac como um adjuvante.

Na nomenclatura do pedido, Ra35 refere-se ao terminal N de MTB32A (Ra35FL), compreendendo, pelo menos, cerca de 195 a 205 aminoácidos de MTB32A de *M. tuberculosis* ou à região correspondente de outra espécie de *Mycobacterium*. Ra12 refere-se ao terminal C de MTB32A (Ra35FL), compreendendo, pelo menos, cerca de 132 dos últimos aminoácidos de MTB32A de *M. tuberculosis* ou a região correspondente de outra espécie de *Mycobacterium*.

O seguinte proporciona sequências de alguns antígenos utilizados nas proteínas de fusão da invenção:

SEQ ID N°: 1-4: MTB32A (Ra35FL ou Ra35 maduro), cuja sequência é também divulgada como SEQ ID N°: 17 (ADNc) e SEQ ID N°: 79 (proteína) nos pedidos de patente U.S. N° 08/523436, 08/523435, N° 08/658800, N° 08/659683, N°. 08/818112, N° 09/056556 e N° 08/818111 e nos pedidos WO97/09428 e WO97/09429, ver também Skeiky et al., *Infection and Immunity* **67**:3998-4007 (1999). O termo MTB32A inclui também sequências de aminoácidos MTB32A, nas quais o resíduo de serina na posição de aminoácido 208 na SEQ ID N°: 2 ou na posição de aminoácido 183 na SEQ ID N°: 4, foi modificado para outro aminoácido (e. g., alanina, Ra35FLMutSA, ver e. g., Figura 6 e SEQ ID N°: 6).

SEQ ID N°: 5 e 6: Ra35FLMut SA, a versão madura de RA35FL na qual o resíduo de serina na posição de aminoácido 183 da SEQ ID N°: 4 foi modificada para um resíduo alanina.

SEQ ID N°: 7 e 8: Ra35, o terminal N de MTB32A (Ra35FL), compreendendo, pelo menos, cerca de 195 aminoácidos do terminal N de MTB32A de *M. tuberculosis*, cujas sequências nucleotídicas e de aminoácidos são divulgadas na Figura 4 (ver também aminoácidos 33-227 da SEQ ID N°: 2 e aminoácidos 8 a 202 da SEQ ID N°: 4). O termo Ra35 (N-term) inclui também sequências de aminoácidos Ra35, nas quais o local activo Ser foi modificado como descrito acima.

SEQ ID N°: 9 e 10: MTBRa12, o terminal C de MTB32A (Ra35FL), compreendendo, pelo menos, cerca de 132

aminoácidos do terminal C de MTB32A de *M. tuberculosis* (ver, e. g., os aminoácidos 224 a 355 da SEQ ID N°: 2 e os aminoácidos 199 a 330 da SEQ ID N°: 4), cuja sequência é divulgada como SEQ ID N°: 4 (ADN) e SEQ ID N°: 66 (sequência de aminoácidos prevista) no pedido de patente U.S. N° 09/072967.

SEQ ID N°: 11, 12, 13 e 14: MTB39 (TbH9), cuja sequência é divulgada como SEQ ID N°: 106 (ADNc completo) e SEQ ID N°: 107 (proteína completa) nos pedidos de patente U.S. N° 08/658800, N° 08/659683, N° 08/818112 e N° 08/818111 e nos pedidos WO97/09428 e WO97/09429. A sequência é também divulgada como SEQ ID N°: 33 (ADN) e SEQ ID N°: 91 (aminoácidos) no pedido de patente U.S. N° 09/056559.

O seguinte proporciona sequências de algumas proteínas de fusão

SEQ ID N°: 15 e 16: MTB72F (Ra12-TbH9-Ra35), cuja sequência é divulgada como SEQ ID N°: 1 (ADN) e SEQ ID N°: 2 (proteína) no pedido de patente U.S. N° 09/223040, N° 09/223040 e no pedido PCT/US99/07717. O termo MTB372F inclui também sequências de aminoácidos MTB72F, nas quais qualquer dos três aminoácidos da tríade do local activo em Ra35FL (*i. e.*, His, Asp ou Ser), foi modificado como descrito acima (ver e. g., MTB72FMutSA, Figura 5).

SEQ ID N°: 17 e 18: MTB72FMutSA (Ra12-TbH9-Ra35MutSA), em que, no componente Ra35 da proteína de fusão, a serina na posição 710 foi modificada para uma alanina.

SEQ ID N°: 19 e 20: TbH9-Ra35 (MTB59F), cuja sequência é divulgada como SEQ ID N°: 23 (ADNc) e SEQ ID N°: 24 (proteína) no pedido de patente U.S. N° 09/287849 e no pedido PCT/US99/07717.

O seguinte proporciona sequências de alguns antigénios adicionais utilizados nas composições e proteínas de fusão da invenção:

SEQ ID N° 21 e 22: MTB8.4 (DPV), cuja sequência é divulgada como SEQ ID N°: 101 (ADNc) e SEQ ID N°: 102 (proteína) nos pedidos de patente U.S. N° 08/658800, N° 08/659683, N° 08/818112 e N° 08/818111 e nos pedidos WO97/09428 e WO97/09429.

SEQ ID N°: 23 e 24: MTB9.8 (MSL), cuja sequência é divulgada como SEQ ID N°: 12 (ADN), SEQ ID N°: 109 (sequência de aminoácidos prevista) e SEQ ID N°: 110 a 124 (péptidos) nos pedidos de patente U.S. N° 08/859381, N° 08/858998, N° 09/073009 e N° 09/073010 e nos pedidos PCT/US98/10407 e PCT/US98/10514.

SEQ ID N°: 25, 26 e 27: MTB9.9A (MTI, também conhecido como MTI-A), cuja sequência é divulgada como SEQ ID N°: 3 e SEQ ID N°: 4 (ADN) e SEQ ID N°: 29 e SEQ ID N°: 51 a 66 (péptido da ORF de MTI) nos pedidos de patente U.S. N° 08/859381, N°. 08/858998, N° 09/073009 e 09/073010 e nos pedidos PCT/US98/10407 e PCT/US98/10514. Existem também duas outras variantes MTI designadas MTI-B e MTI-C.

SEQ ID N°: 28 e 29: MTB40 (HTCC#1), cuja sequência é divulgada como SEQ ID N°: 137 (ADNc) e 138 (sequência de

aminoácidos prevista) nos pedidos de patente U.S. N° 09/073009 e N° 09/073010 e nos pedidos PCT/US98/10407 e PCT/US98/10514.

SEQ ID N°: 30 e 31: MTB41 (MTCC#2), cuja sequência é divulgada como SEQ ID N°: 140 (ADNc) e SEQ ID N°: 142 (sequência de aminoácidos prevista) nos pedidos de patente U.S. N° 09/073009 e N° 09/073010 e nos pedidos PCT/US98/10407 e PCT/US98/10514.

SEQ ID N°: 32 e 33: ESAT-6, cuja sequência é divulgada como SEQ ID N°: 103 (ADN) e SEQ ID N°: 104 (sequência de aminoácidos prevista) no pedido de patente U.S. N° 09/072967. A sequência de ESAT-6 é também divulgada na Patente U.S. N° 5955077.

SEQ ID N°: 34 e 35: Tb38-1 ou 38-1 (MTb11), cuja sequência é divulgada na SEQ ID N°: 46 (ADN) e SEQ ID N°: 88 (aminoácidos previstos) nos pedidos de patente U.S. N° 09/07296; 08/523436; 08/523435; 08/818112; e 08/818111; e nos pedidos WO97/09428 e WO97/09429.

SEQ ID N°: 36 e 37: TbRa3, cuja sequência é divulgada naa SEQ ID N°: 15 (ADN) e SEQ ID N°: 77 (sequência de aminoácidos prevista) dos pedidos WO 97/09428 e WO97/09429.

SEQ ID N°: 38 e 39: 38 kD, cuja sequência é divulgada na SEQ ID N°: 154 (ADN) e SEQ ID N°: 155 (sequência de aminoácidos prevista) no pedido de patente U.S. N° 09/072967. 38 kD têm duas formas alternativas, com e sem o resíduo de cisteína do terminal N.

SEQ ID N°: 40 e 41: DPEP, cuja sequência é divulgada na SEQ ID N°: 52 (ADN) e SEQ ID N°: 53 (sequência de aminoácidos prevista) nas publicações WO97/09428 e WO97/09429.

SEQ ID N°: 42 e 43: TbH4, cuja sequência é divulgada como SEQ ID N°: 43 (ADN) e SEQ ID N°: 81 (sequência de aminoácidos prevista) nas publicações WO97/09428 e WO97/09429.

SEQ ID N°: 44 e 45: DPPD, cuja sequência é divulgada na SEQ ID N°: 240 (ADN) e SEQ ID N°: 241 (sequência de aminoácidos prevista) no documento USSN 09/072967 e nos pedidos PCT/US99/03268 e PCT/US99/03265. A forma segregada de DPPD é aqui mostrada na Figura 12 do documento PCT/US00/28095.

MTb82 (MTb867), cuja sequência é divulgada nas Figuras 8 (ADN) e 9 (aminoácidos) do documento PCT/US00/2809.

Erd14 (MTb16), cujo ADNc e sequências de aminoácidos são divulgados em Verbon *et al.*, *J. Bacteriology* **174**:1352-1359 (1992).

Antigénio α -cristalino, cuja sequência é divulgada em Verbon *et al.*, *J. Bact.* **174**:1352-1359 (1992);

Antigénio complexo 85, e. g., antigénio 85b, cuja sequência é divulgada no Content *et al.*, *Infect. & Immunol.* **59**:3205-3212 (1991).

O seguinte proporciona sequências de algumas proteínas de fusão adicionais utilizadas nas composições e proteínas de fusão da invenção:

SEQ ID N°: 48 e 49 DPV-MTI-MSL-MTCC#2 (MTb71F), cuja sequência é divulgada como SEQ ID N°: 15 (ácido nucleico) e na SEQ ID N°: 16: (proteína) no pedido de patente U.S. N° 09/287849 e no pedido PCT/US99/07717.

SEQ ID N°: 46 e 47: DPV-MTI-MSL (MTb31F), cuja sequência é divulgada na SEQ ID N°: 18 (ADNc) e SEQ ID N°: 19 (proteína) no pedido de patente U.S. N° 09/287,849 e no pedido PCT/US99/07717.

Cada das sequências acima é também divulgada em Cole *et al.* *Nature* **393**:537 (1998) e podem ser encontradas em, *e. g.*, <http://www.sanger.ac.uk> e <http://www.pasteur.fr/mycodb/>.

As sequências acima são divulgadas nos pedidos de patente U.S. N° 08/523435, 08/523436, 08/658800, 08/65968308/81811108/81811208/942341, 08/942578, 08/858998, 08/859381, 09/056556, 09/072596, 09/072967, 09/073009, 09/073010, 09/223040, 091287849 09/597796; e nos pedidos de patente PCT PCT/US00/28095; PCT/US98/10407, PCT/US98/10514, PCT/US99/03265, PCT/US99/03268, PCT/US99/07717, WO97/09428 e WO97/09429, WO98/16645, WO98/16646.

Os antigénios aqui descritos incluem variantes polimórficas e variações modificadas de forma conservadora, assim como homólogos de *Mycobacterium* inter-estirpe e interespecíficos. Além disso, os antigénios aqui descritos incluem subsequências ou sequências truncadas. As proteínas de fusão podem conter

também polipéptidos adicionais, opcionalmente péptidos heterólogos provenientes de *Mycobacterium* ou de outras fontes. Estes antígenos podem ser modificados, por exemplo, adicionando sequências de elementos de ligação peptídicos como descrito abaixo. Estes elementos de ligação peptídicos podem ser inseridos entre um ou mais polipéptidos que constituem cada das proteínas de fusão.

DEFINIÇÕES

“Polipéptido de fusão” ou “proteína de fusão” refere-se a uma proteína tendo, pelo menos, dois polipéptidos heterólogos de *Mycobacterium* sp. ligados covalentemente, directamente ou através de um elemento de ligação aminoacídico. Os polipéptidos que formam a proteína de fusão estão tipicamente ligados do terminal C ao terminal N, embora estes possam também estar ligados do terminal C ao terminal C, do terminal N ao terminal N ou do terminal N ao terminal C. Os polipéptidos da proteína de fusão podem estar sob qualquer ordem. Este termo refere-se também a variantes modificadas de forma conservadora, variantes polimórficas, alelos, mutantes, subsequências, homólogos interespecíficos e fragmentos imunogénicos dos antígenos que constituem a proteína de fusão. Os antígenos de *Mycobacterium tuberculosis* são descritos em Cole et al., *Nature* **393**:537 (1998), que divulga o genoma de *Mycobacterium tuberculosis* completo. A sequência completa de *Mycobacterium tuberculosis* pode ser também encontrada em <http://www.sanger.ac.uk> e em <http://www.pasteur.fr/mycdb/> (MycDB). Os antígenos de outras espécies de *Mycobacterium* que correspondem a antígenos de *M. tuberculosis* podem ser identificados, e. g., utilizando algoritmos de comparação de sequências, como aqui descritos ou

outros métodos conhecidos do especialista na técnica, e. g., ensaios de hibridação e ensaios de ligação de anticorpos. As proteínas de fusão da invenção podem também compreender cópias adicionais de um antígeno componente ou seu fragmento imunogénico.

Em algumas formas de realização, os polipéptidos individuais da proteína de fusão estão ordenados dos grandes para os pequenos (do terminal N para o C). Os antígenos grandes têm aproximadamente 30 a 150 kD de tamanho, os antígenos médios têm aproximadamente 10 a 30 kD de tamanho e os antígenos pequenos têm aproximadamente menos de 10 kD de tamanho. A sequência codificando o polipéptido individual pode ser tão pequena quanto, e. g., um fragmento imunogénico tal como um epitopo de CTL individual codificando cerca de 8 a 9 aminoácidos, ou, e. g., um epitopo de HTL ou de célula B. O fragmento pode também incluir epitopos múltiplos. O fragmento imunogénico pode também representar uma parte maior da sequência do antígeno, e. g., cerca de 50% ou mais de MTB39 e MTB32A, e. g., as porções do terminal N e C de MTB32A. Os fragmentos imunogénicos preferidos de MTB32A incluem Ra12, Ra35 e Ra35 MutSA.

Um polipéptido de fusão da invenção liga-se especificamente a anticorpos produzidos contra, pelo menos, dois polipéptidos antigénicos, em que cada polipéptido antigénico é seleccionado do grupo consistindo em MTB39 ou uma sua porção imunogénica ou fragmento e MTB32A ou uma sua porção imunogénica. Os anticorpos podem ser policlonais ou monoclonais. Opcionalmente, o polipéptido de fusão liga-se especificamente a anticorpos produzidos contra a junção de fusão dos antígenos, em que os anticorpos não se ligam aos antígenos individualmente, i. e.,

quando estes não são parte de uma proteína de fusão. Os polipéptidos de fusão compreendem opcionalmente polipéptidos adicionais, e. g., três, quatro, cinco, seis ou mais polipéptidos, até cerca de 25 polipéptidos, opcionalmente polipéptidos heterólogos ou polipéptidos homólogos repetidos, fundidos com, pelo menos, dois antígenos heterólogos. Os polipéptidos adicionais da proteína de fusão são opcionalmente derivados de *Mycobacterium*, assim como de outras fontes, tais como outras fontes bacterianas, virais ou de invertebrado, de vertebrado ou de mamífero. Os polipéptidos individuais da proteína de fusão podem estar sob qualquer ordem. Como aqui descrito, a proteína de fusão pode estar também ligada a outras moléculas, incluindo polipéptidos adicionais. As composições da invenção podem também compreender polipéptidos adicionais que não estão ligados às proteínas de fusão da invenção. Estes polipéptidos adicionais podem ser polipéptidos heterólogos ou homólogos.

O termo “fundido” refere-se à ligação covalente entre dois polipéptidos numa proteína de fusão. Os polipéptidos estão tipicamente ligados através de uma ligação peptídica, directamente entre si ou através de um elemento de ligação aminoacídico. Opcionalmente, os péptidos podem ser ligados através de ligações covalentes não-peptídicas conhecidas do especialista na técnica.

“FL” refere-se a completo, i. e., um polipéptido que tem o mesmo comprimento que o polipéptido do tipo selvagem.

O termo “seu fragmento imunogénico” refere-se a um polipéptido compreendendo um epitopo que é reconhecido por linfócitos T citotóxicos, linfócitos T auxiliares ou células

B. Os fragmentos imunogénicos preferidos de, e. g., MTB32A, são RA35, Ra35MutSA ou Ra12.

O termo “espécie de *Mycobacterium* do complexo da tuberculose” inclui as espécies normalmente consideradas como causadoras da doença da tuberculose, assim como espécies ambientais e oportunistas de *Mycobacterium* que causam tuberculose e doenças pulmonares em doentes imunitariamente comprometidos, tal como doentes com SIDA, e. g., *M. tuberculosis*, *M. bovis* ou *M. africanum*, BCG, *M. avium*, *M. intracellulare*, *M. celatum*, *M. genavense*, *M. haemophilum*, *M. kansasii*, *M. simiae*, *M. vaccae*, *M. fortuitum* e *M. scrofulaceum* (ver e. g., *Harrison's Principles of Internal Medicine*, volume 1, pp 1004-1014 e 1019-1023 (14^a ed., Fauci et al., eds., 1998)).

Um adjuvante refere-se aos componentes numa vacina ou numa composição terapêutica que aumentam a resposta imunitária específica ao antigénio (ver e. g., Edelman, *AIDS Res. Hum Retroviruses* **8**:1409-1411 (1992)). Os adjuvantes induzem respostas imunitárias de respostas do tipo Th1 e do tipo Th-2. As citocinas do tipo Th1 (e. g., IFN- γ , IL-2 e IL-12) tendem a favorecer a indução da resposta imunitária mediada por células a um antigénio administrado, enquanto que as citocinas do tipo Th-2 (e. g., IL-4, IL-5, IL-6, IL-10 e TNF- β) tendem a favorecer a indução de respostas imunitárias humorais.

“Ácido nucleico” refere-se a desoxirribonucleótidos ou ribonucleótidos e seus polímeros na forma de cadeia simples ou dupla. O termo abrange ácidos nucleicos contendo análogos de nucleótido conhecidos ou resíduos ou ligações de estrutura base modificados, que sejam sintéticos, de ocorrência natural e de ocorrência não natural, que tenham propriedades de ligação

semelhantes ao ácido nucleico de referência e que sejam metabolizados de uma forma semelhante aos nucleótidos de referência. Os exemplos desses análogos incluem, sem limitação, fosforotioatos, fosforamidatos, metilo fosfonatos, metilfosfonatos quirais, 2-O-metil ribonucleótidos, ácidos nucleicos peptídicos (PNA).

A menos que indicado de outra forma, uma sequência de ácido nucleico particular abrange, também, implicitamente as suas variantes modificadas de forma conservadora (e. g., substituições de codão degeneradas) e sequências complementares, assim como a sequência explicitamente indicada. Especificamente, as substituições de codão degeneradas podem ser realizadas produzindo sequências nas quais a terceira posição de um ou mais codões seleccionados (ou de todos) está substituída com resíduos de base mista e/ou de desoxi-inosina (Batzner *et al.*, *Nucleic Acid Res.* **19**:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* **260**:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* **8**:91-98 (1994)). O termo ácido nucleico é utilizado, de uma forma alternada, com gene, ADNc, ARNm, oligonucleótido e polinucleótido.

Os termos "polipéptido", "péptido" e "proteína" são aqui utilizados indiscriminadamente para se referirem a um polímero de resíduos de aminoácidos. Os termos aplicam-se a polímeros de aminoácidos em que um ou mais resíduos de aminoácidos são um composto químico artificial mimético de um aminoácido de ocorrência natural correspondente, assim como a polímeros de aminoácidos de ocorrência natural e polímeros de aminoácidos de ocorrência não natural.

O termo "aminoácido" refere-se aminoácidos de ocorrência natural e sintéticos, assim como a análogos de aminoácido e miméticos de aminoácidos que funcionam de uma forma semelhante aos aminoácidos de ocorrência natural. Os aminoácidos de ocorrência natural são os codificados pelo código genético, assim como os aminoácidos que são posteriormente modificados, *e. g.*, hidroxiprolina, γ -carboxiglutamato e O-fosfoserina. Os análogos de aminoácido referem-se a compostos que têm a mesma estrutura química básica de um aminoácido de ocorrência natural, *i. e.*, um carbono α que está ligado a um hidrogénio, um grupo carboxilo, um grupo amina e um grupo R, *e. g.*, homosserina, norleucina, sulfóxido de metionina, metilsulfónio de metionina. Esses análogos têm grupos R modificados (*e. g.*, norleucina) ou estruturas base peptídicas modificadas, mas conservam a mesma estrutura química básica de um aminoácido de ocorrência natural. Miméticos de aminoácido referem-se a compostos químicos que têm uma estrutura que é diferente da estrutura química geral de um aminoácido, mas que funcionam de uma forma semelhante a um aminoácido de ocorrência natural.

Os aminoácidos podem ser aqui designados pelos seus símbolos de três letras normalmente conhecidos, ou pelos símbolos de uma letra recomendados pela Comissão de Nomenclatura Bioquímica IUPAC-IUB. Do mesmo modo, os nucleótidos podem ser designados pelos seus códigos de uma letra normalmente aceites.

"Variantes modificadas de forma conservadora" aplica-se a sequências de aminoácidos e de ácidos nucleicos. No que respeita a sequências de ácidos nucleicos particulares, as variantes modificadas de forma conservadora referem-se aos ácidos nucleicos que codificam sequências de aminoácidos idênticas ou essencialmente idênticas ou em que o ácido nucleico não codifica

uma sequência de aminoácidos, a sequências essencialmente idênticas. Devido à degenerescência do código genético, um grande número de ácidos nucleicos funcionalmente idênticos codifica uma determinada proteína. Por exemplo, os codões GCA, GCC, GCG e GCU codificam todos o aminoácido alanina. Deste modo, para cada posição em que é especificada uma alanina por um codão, o codão pode ser alterado para qualquer dos codões correspondentes descritos, sem alterar o polipéptido codificado. Essas variações de ácidos nucleicos são “variações silenciosas”, as quais são um tipo de variações modificadas de forma conservadora. Cada sequência de ácido nucleico aqui que codifica um polipéptido descreve também cada variação silenciosa possível do ácido nucleico. Um especialista na técnica irá reconhecer que cada codão num ácido nucleico (excepto o AUG, o qual é normalmente o único codão da metionina e o TGG, o qual é normalmente o único codão do triptofano) pode ser modificado para produzir uma molécula funcionalmente idêntica. Consequentemente, cada variação silenciosa de um ácido nucleico que codifica um polipéptido está implícita em cada sequência descrita.

No que se refere a sequências de aminoácidos, um especialista na técnica irá reconhecer que substituições, eliminações ou adições individuais num sequência de ácido nucleico, péptido, polipéptido ou proteína que alteram, adicionam ou eliminam um único aminoácido ou uma pequena percentagem de aminoácidos na sequência codificada é “uma variante modificada de forma conservadora”, em que a alteração resulta na substituição de um aminoácido por um aminoácido quimicamente semelhante. As tabelas de substituição conservadoras que proporcionam aminoácidos funcionalmente semelhantes são bem conhecidas na técnica. Essas variantes

modificadas de forma conservadora são adicionais e não excluem variantes polimórficas, homólogos interespecíficos e alelos da invenção.

Cada dos oito grupos seguintes contém aminoácidos que são substituições conservadoras entre si:

- 1) Alanina (A), Glicina (G);
- 2) Ácido aspártico (D), Ácido glutâmico (E);
- 3) Asparagina (N), Glutamina (Q);
- 4) Arginina (R), Lisina (K);
- 5) Isoleucina (I), Leucina (L), Metionina (M), Valina (V);
- 6) Fenilalanina (F), Tirosina (Y), Triptofano (W);
- 7) Serina (S), Treonina (T); e
- 8) Cisteína (C), Metionina (M)

(ver, e. g., Creighton, *Proteins* (1984)).

O termo “heterólogo” quando utilizado em referência a porções de um ácido nucleico indica que o ácido nucleico compreende duas ou mais subsequências que não são encontradas com a mesma relação entre si na natureza. Por exemplo, o ácido nucleico é tipicamente produzido recombinantemente, tendo duas ou mais sequências de genes não relacionados, arranjos para produzir um novo ácido nucleico funcional, e. g., um promotor de

uma fonte e uma região codificante de outra fonte. De um modo semelhante, uma proteína heteróloga indica que a proteína compreende duas ou mais subsequências que não são encontradas com a mesma relação entre si na natureza (e. g., uma proteína de fusão).

A frase “hibrida selectivamente (ou especificamente)” refere-se a ligação, formação de duplexes ou hibridação de uma molécula apenas com uma sequência de nucleótidos particular sob condições de hibridação restritivas quando essa sequência está presente numa mistura complexa (e. g., ADN ou ARN celular ou de biblioteca total).

A frase “condições de hibridação de severidade” refere-se a condições sob as quais uma sonda irá hibridar com a sua subsequência alvo, tipicamente numa mistura complexa de ácidos nucleicos, mas não com qualquer outra sequência. As condições restritivas são dependentes da sequência e irão ser diferentes em circunstâncias diferentes. Sequências mais longas hibridam especificamente a temperaturas mais elevadas. 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) encontra-se um guia abrangente da hibridação de ácidos nucleicos. Geralmente, são seleccionadas condições restritivas de forma a serem cerca de 5-10 °C inferiores ao ponto de fusão térmica (T_m) da sequência específica numa força iónica definida pH. T_m é a temperatura (sob uma força iónica, pH e concentração nucleica definidas) na qual 50% das sondas complementares ao alvo hibridam com a sequência alvo no equilíbrio (uma vez que as sequências alvo estão presentes em excesso, na T_m , 50% das sondas estão ocupadas no equilíbrio). As

condições restritivas irão ser aquelas em que a concentração de sal é inferior a cerca de 1,0 M de ião de sódio, tipicamente cerca de 0,01 até à concentração de ião de sódio 1,0 M (ou de outros sais) a pH 7,0 a 8,3 e a temperatura é, pelo menos, cerca de 30 °C para sondas curtas (e. g., 10 a 50 nucleótidos) e, pelo menos, cerca de 60 °C para sondas longas (e. g., mais de 50 nucleótidos). Podem ser também obtidas condições restritivas com a adição de agentes desestabilizantes tais como formamida. Para a hibridação selectiva ou específica, um sinal positivo é, pelo menos, duas vezes o fundo, opcionalmente 10 vezes o fundo da hibridação. Exemplos de condições de hibridação restritivas podem ser como se segue: formamida 50%, SSC 5 x e SDS a 1%, incubando a 42 °C ou, SSC 5 x, SDS a 1%, incubando a 65 °C, com lavagem em SSC 0,2 x e SDS a 0,1% a 65 °C.

Os ácidos nucleicos que não hibridam entre si sob condições restritivas são ainda substancialmente idênticos se os polipéptidos que estes codificam forem substancialmente idênticos. Isto ocorre, por exemplo, quando é criada uma cópia de um ácido nucleico utilizando a máxima degenerescência de codões permitida pelo código genético. Nesse casos, os ácidos nucleicos hibridam tipicamente sob condições de hibridação restritivas moderada. Os exemplos de "condições de hibridação moderadamente restritivas" incluem uma hibridação num tampão com formamida a 40%, NaCl 1 M, SDS a 1% a 37 °C e lavagem em SSC 1 X a 45 °C. Uma hibridação positiva é, pelo menos, duas vezes o fundo. O especialista na técnica irá reconhecer facilmente que podem ser utilizadas condições de hibridação e de lavagem alternativas para proporcionar condições restritivas semelhantes.

“Anticorpo” refere-se a um polipéptido compreendendo uma região estrutural de um gene de imunoglobulina ou os seus fragmentos que se ligam especificamente e reconhecem um antígeno. Os genes da imunoglobulina reconhecidos incluem os genes das regiões constantes kappa, lambda, alfa, gama, delta, epsilon e mu, assim como os vários genes de regiões variáveis da imunoglobulina. As cadeias leves são classificadas como kappa ou lambda. As cadeias pesadas são classificadas como gama, mu, alfa, delta ou epsilon, as quais, por sua vez, definem, respectivamente, as classes de imunoglobulina IgG, IgM, IgA, IgD e IgE.

Um exemplo de unidade estrutural de imunoglobulina (anticorpo) compreende um tetrâmero. Cada tetrâmero é composto por dois pares idênticos de cadeias polipeptídicas, tendo cada par uma cadeia “leve” (cerca de 25 kDa) e uma “pesada” (cerca de 50-70 kDa). O terminal N de cada cadeia define uma região variável com cerca de 100 a 110 ou mais aminoácidos principalmente responsáveis pelo reconhecimento do antígeno. Os termos cadeia leve variável (V_L) e cadeia pesada variável (V_H) referem-se, respectivamente a estas cadeias leves e pesadas.

Os anticorpos estão presentes, e. g., sob a forma de imunoglobulinas intactas ou sob a forma de vários fragmentos, bem caracterizados, produzidos por digestão com várias peptidases. Deste modo, por exemplo, a pepsina digere um anticorpo abaixo das ligações persulfureto na região de charneira para produzir $F(ab)'_2$, um dímero de Fab, o qual, em si, é uma cadeia leve ligada a V_H-C_{H1} por uma ligação persulfureto. O $F(ab)'_2$ pode ser reduzido sob condições suaves para quebrar a ligação persulfureto na região de charneira, convertendo deste modo o dímero $F(ab)'_2$ num monómero Fab'. O

monómero Fab' é essencialmente Fab com parte da região de charneira (ver *Fundamental Immunology* (Paul ed., 3ª ed. 1993)). Embora vários fragmentos de anticorpo sejam definidos em termos da digestão de um anticorpo intacto, um especialista na técnica irá ter em consideração que esse fragmento pode ser sintetizado de novo, quimicamente ou utilizando metodologia de ADN recombinante. Deste modo, o termo anticorpo, como aqui utilizado, inclui também fragmentos de anticorpo produzidos pela modificação de anticorpos completos ou os sintetizados de novo, utilizando metodologias de ADN recombinante (e. g., cadeia única Fv) ou os identificados utilizando bibliotecas de apresentação de fagos (ver e. g., McCafferty et al., *Nature* **348**:552-554 (1990)).

Pode ser utilizada qualquer técnica conhecida no campo técnico na preparação de anticorpos monoclonais ou policlonais, (ver e. g., Kohler e Milstein, *Nature* **256**:495-497 (1975); Kozbor et al., *Immunology Today* **4**: 72 (1983); Cole et al., páginas 77-96 em *Monoclonal Antibodies and Cancer Therapy* (1985)). As técnicas de produção de anticorpos de cadeia única (Patente U.S. 4946778) podem ser adaptadas para produzir anticorpos contra polipéptidos desta invenção. Do mesmo modo, podem ser utilizados organismos transgénicos ou outros organismos, tal como outros mamíferos, para expressar anticorpos humanizados. Alternativamente, pode ser utilizada a tecnologia de apresentação de fagos para identificar anticorpos e fragmentos Fab heteroméricos que se ligam especificamente a antígenos seleccionados (ver e. g., McCafferty et al., *Nature* **348**:552-554 (1990); Marks et al., *Biotechnology* **10**:779-783 (1992)).

A frase "liga-se especificamente (ou selectivamente)" a um anticorpo ou "especificamente (ou selectivamente) imunorreactivo

com", em referência a uma proteína ou péptido, refere-se a uma reacção de ligação que determina a presença da proteína numa população heterogénea de proteínas e outras entidades biológicas. Deste modo, sob as condições de imunoensaio indicadas, os anticorpos especificados ligam-se a uma proteína particular em, pelo menos, duas vezes o fundo e não se ligam substancialmente numa quantidade significativa a outras proteínas presentes na amostra. A ligação específica a um anticorpo nessas condições pode requerer um anticorpo que seja seleccionado pela sua especificidade para uma proteína particular. Por exemplo, podem ser seleccionados anticorpos policlonais produzidos contra proteínas de fusão para obter apenas os anticorpos policlonais que são especificamente imunorreactivos com a proteína de fusão e não com componentes individuais das proteínas de fusão. Esta selecção pode ser realizada subtraindo os anticorpos que apresentam reacção cruzada com os antigénios individuais. Podem ser utilizados vários formatos de imunoensaio para seleccionar anticorpos especificamente imunorreactivos com uma proteína particular. Por exemplo, são normalmente utilizados imunoensaios de ELISA em fase sólida para seleccionar anticorpos especificamente imunorreactivos com uma proteína (ver e. g., Harlow e Lane, *Antibodies, A Laboratory Manual* (1988), para uma descrição de formatos de imunoensaio e condições que podem ser utilizadas para determinar imunorreactividade específica). Tipicamente uma reacção específica ou selectiva será, pelo menos, duas vezes o sinal de fundo ou o ruído e mais tipicamente mais de 10 a 100 vezes o fundo.

Os polinucleótidos podem compreender uma sequência nativa (*i. e.*, uma sequência endógena codificando um antigénio individual ou uma sua porção) ou podem compreender uma variante

dessa sequência. As variantes de polinucleótidos podem conter uma ou mais substituições, adições, eliminações e/ou inserções de forma a que a actividade biológica do polipéptido de fusão codificado não seja diminuída em relação a um polipéptido de fusão compreendendo antigénios nativos. De um modo preferido, as variantes apresentam uma identidade de, pelo menos, cerca de 70%, de um modo mais preferido, pelo menos, uma identidade de cerca de 80% e, de um modo muito preferido, identidade de, pelo menos, cerca de 90% com uma sequência polinucleotídica codificando um polipéptido nativo ou uma sua porção.

Os termos “idênticos” ou percentagem de “identidade”, no contexto de duas ou mais sequências de ácidos nucleicos ou polipeptídicas, referem-se a duas ou mais sequências ou subsequências que são iguais ou têm uma percentagem especificada de resíduos de aminoácido ou nucleótidos que são iguais (*i. e.*, identidade de 70%, opcionalmente identidade de 75%, 80%, 85%, 90% ou 95% ao longo de uma região especificada), quando comparadas e alinhadas para uma correspondência máxima ao longo de uma janela de comparação ou região indicada como medida utilizando um dos algoritmos de comparação de sequências seguintes ou por alinhamento manual e inspecção visual. Essas sequências são então designadas como “substancialmente idênticas”. Esta definição refere-se também ao cumprimento de uma sequência de teste. Opcionalmente, a identidade ocorre ao longo de uma região que tem, pelo menos, cerca de 25 a cerca de 50 aminoácidos ou nucleótidos de comprimento ou, opcionalmente, ao longo de uma região que tem 75-100 aminoácidos ou nucleótidos de comprimento.

Tipicamente, para a comparação de sequências, uma sequência funciona como uma sequência de referência, com a qual as sequências a testar são comparadas. Quando é utilizado um algoritmo de comparação de sequências, as sequências de teste e de referência são introduzidas num computador, se necessário, são indicadas as coordenadas da subsequência e são indicados os parâmetros de programa de algoritmo da sequência. Podem ser utilizados parâmetros de programa por omissão ou podem ser indicados parâmetros alternativos. O algoritmo de comparação de sequência calcula então a identidade de sequência percentual das sequências de teste em relação à sequência de referência, baseada nos parâmetros do programa.

Uma “janela de comparação”, como aqui utilizado, inclui uma referência a um segmento de qualquer uma das várias posições contíguas seleccionadas do grupo consistindo de desde 25 a 500, normalmente cerca de 50 a cerca de 200, mais normalmente cerca de 100 a cerca de 150, em que uma sequência pode ser comparada com uma sequência de referência com o mesmo número de posições contíguas após as duas sequências terem sido alinhadas optimamente. Os métodos de alinhamento de sequências para comparação são bem conhecidos na técnica. O alinhamento óptimo das sequências para comparação pode ser realizado, e. g., através do algoritmo de homologia local de Smith e Waterman, *Adv. Appl. Math.* **2**:482 (1981), através do algoritmo de alinhamento de homologia de Needleman e Wunsch, *J. Mol. Biol.* **48**:443 (1970), através do método de pesquisa de semelhança de Pearson e Lipman, *Proc. Nat’l. Acad. Sci. USA* **85**:2444 (1988), através do formas de realização computadorizadas destes algoritmos (GAP, BESTFIT, FASTA e TFASTA no Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI) ou por alinhamento manual e inspecção visual (ver

e. g., *Current Protocols in Molecular Biology* (Ausubel et al., ed. Suplemento de 1995)).

Um exemplo de um algoritmo útil é o PILE UP. O PILE UP cria um alinhamento múltiplo de sequências a partir de um grupo de sequências relacionadas utilizando alinhamentos emparelhados progressivos para mostrar a relação e a identidade de sequência percentual. Este representa também graficamente uma árvore ou um dendograma mostrando as relações de agrupamento utilizadas para criar o alinhamento. O PILE UP utiliza uma simplificação do método de alinhamento progressivo de Feng e Doolittle, *J. Mol. Evol.* **35**:351-360 (1987). O método utilizado é semelhante ao método descrito por Higgins e Sharp, *CABIOS* **5**:151-153 (1989). O programa pode alinhar até 300 sequências, cada com um comprimento máximo de 5000 nucleótidos ou aminoácidos. O processo de alinhamento múltiplo começa com o alinhamento emparelhado das duas sequências mais semelhantes, produzindo um grupo de duas sequências alinhadas. Este grupo é então alinhado com a sequência seguinte ou com o grupo de sequências alinhadas mais relacionado. São alinhados dois grupos de sequências por uma extensão simples do alinhamento emparelhado de duas sequências individuais. O alinhamento final é obtido por uma série de alinhamentos emparelhados progressivos. O programa é utilizado indicando sequências específicas e as suas coordenadas de aminoácidos ou nucleótidos de regiões de comparação de sequência e indicando os parâmetros do programa. É comparada uma sequência de referência com outras sequências de teste para determinar a relação de identidade de sequência percentual utilizando o PILE UP, utilizando os parâmetros seguintes: ponderação de lacuna por omissão (3,00), ponderação da extensão do lacuna por omissão (0,10) e lacunas de extremidade ponderados. O PILE UP pode ser obtido a partir do pacote de

software de análise de sequências GCG, e. g., versão 7.0 (Devereaux *et al.*, *Nuc. Acids Res.* **12**:387-395 (1984).

Outro exemplo de um algoritmo que é adequado para determinar a identidade de sequência percentual e a semelhança de sequência são os algoritmos BLAST e BLAST 2.0, os quais são descritos), respectivamente, em Altschul *et al.*, *Nuc. Acids Res.* **25**:3389-3402 (1977) e Altschul *et al.*, *J. Mol. Biol.* **215**:403-410 (1990). O programa informático para realizar análises BLAST está publicamente disponível através National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo envolve inicialmente a identificação de pares de sequência com pontuação elevada (HSP) identificando palavras de pequeno comprimento W na sequência da questão, as quais correspondem ou satisfazem algum limiar de pontuação estimado positivamente T quando alinhadas com uma palavra do mesmo comprimento numa sequência de uma base de dados. T é designado como o limiar de pontuação de palavras da vizinhança (Altschul *et al.*, *supra*). Estas vizinhanças de resultados positivos de palavras funcionam como fundamentos para iniciar pesquisas para encontrar HSP mais longas que as contenham. Os resultados positivos para palavras são estendidos em ambas as direcções ao longo de cada sequência até que a pontuação cumulativa do alinhamento possa ser aumentada. As pontuações cumulativas são calculadas utilizando, para sequências nucleotídicas, os parâmetros M (pontuação de recompensa para um par de resíduos correspondentes; sempre > 0) e N (pontuação de penalização para resíduos não correspondentes; sempre < 0). Para sequências de aminoácidos, é utilizada uma matriz de pontuação para calcular a pontuação cumulativa. A extensão dos resultados positivos de palavras em cada direcção é terminada quando: a pontuação de alinhamento cumulativa diminui na quantidade X a partir do seu

valor máximo obtido; a pontuação cumulativa vai até zero ou abaixo, devido à acumulação de um ou mais alinhamentos de resíduos com pontuação negativa; ou ser atingida a extremidade de qualquer uma das sequências. Os parâmetros W, T e X do algoritmo BLAST determinam a sensibilidade e a velocidade do alinhamento. O programa BLASTN (para sequências nucleotídicas) utiliza por omissão um comprimento de palavra (W) de 11, uma expectativa (E) de 10, $M = 5$, $N = -4$ e uma comparação entre ambas as cadeias. Para sequências de aminoácidos, o programa BLASTP utiliza por omissão um comprimento de palavra de 3 e expectativa (E) de 10 e a matriz de pontuação BLOSUM62 (ver Henikoff e Henikoff, *Proc. Natl. Acad. Sci. USA* **89**:10915 (1989)) alinhamentos (B) de 50, expectativa (E) de 10, $M = 5$, $N = -4$ e uma comparação entre ambas as cadeias.

O algoritmo BLAST realiza também uma análise estatística da semelhança entre duas sequências (ver e. g., Karlin e Altschul, *Proc. Nat'l. Acad. Sci. USA* **90**:5873-5787 (1993)). Uma medida da semelhança proporcionada pelo algoritmo BLAST é a probabilidade da menor soma ($P(N)$), o que proporciona uma indicação da probabilidade pela qual uma correspondência entre dois nucleótido ou sequências de aminoácidos irá ocorrer por acaso. Por exemplo, um ácido nucleico é considerado semelhante a uma sequência de referência, se a probabilidade da menor soma numa comparação do ácido nucleico de teste com o ácido nucleico de referência for inferior a cerca de 0,2, de um modo mais preferido, menos de cerca de 0,01 e, de um modo muito preferido, menos de cerca de 0,001.

COMPOSIÇÕES POLINUCLEOTÍDICAS

Como aqui utilizado, os termos “segmento de ADN” e “polinucleótido” referem-se a uma molécula de ADN que foi isolada isenta de ADN genómico total de uma espécie particular. Consequentemente, um segmento de ADN codificando um polipéptido refere-se a um segmento de ADN que contém uma ou mais sequências codificantes mas que é substancialmente isolado a partir de ou isento por purificação de, ADN genómico total da espécie do qual o segmento de ADN é obtido. Dentro dos termos “segmento de ADN” e “polinucleótido” são incluídos segmentos de ADN e fragmentos mais pequenos desses segmentos e também vectores recombinantes, incluindo, por exemplo, plasmídeos, cosmídeos, fagemídeos, fagos, vírus e semelhantes.

Como será entendido pelo especialista na técnica, os segmentos de ADN desta invenção podem incluir sequências genómicas, sequências extra-genómicas e codificadas por plasmídeos e segmentos génicos modificados mais pequenos que expressam ou podem ser adaptados para expressar proteínas, polipéptidos, péptidos e semelhantes. Esses segmentos podem ser isolados naturalmente ou serem modificados sinteticamente por manipulação humana.

Consequentemente, os termos “isolado”, “purificado” ou “biologicamente puro” referem-se a material que está substancialmente ou essencialmente isento de componentes que normalmente o acompanham quando encontrado no seu estado nativo. Obviamente, isto refere-se ao segmento de ADN como originalmente isolado e não exclui outras proteínas, genes ou regiões codificantes isoladas adicionadas posteriormente à composição por manipulação humana. A pureza e a homogeneidade são

determinadas, tipicamente, utilizando técnicas de química analítica, tais como electroforese em gel de poliacrilamida ou cromatografia líquida de elevado desempenho. Uma proteína que é a espécie predominante presente numa preparação está substancialmente purificada. Um ácido nucleico isolado está separado de outras grelhas de leitura aberta que flanqueiam o gene e que codificam outras proteínas além do gene.

Como será reconhecido pelo especialista na técnica, os polinucleótidos podem ser de cadeia simples (codificante ou anti-sentido) ou de cadeia dupla e podem ser moléculas de ADN (genómico, ADNc ou sintético) ou de ARN. As moléculas de ARN incluem moléculas de ARNHn, as quais contêm intrões e correspondem a uma molécula de ADN de uma forma individual e moléculas de ARNm que não contêm intrões. Podem, mas não é necessário, estar presentes sequências codificantes ou não-codificantes adicionais dentro de um polinucleótido da presente invenção e um polinucleótido, pode, mas não é necessário, estar ligado a outras moléculas e/ou materiais de suporte.

Os polinucleótidos podem compreender uma sequência nativa (*i. e.*, uma sequência endógena codificando um antigénio de *Mycobacterium* ou uma sua porção) ou podem compreender uma variante ou um equivalente funcional biológico ou antigénico dessa sequência. As variantes de polinucleótidos podem conter uma ou mais substituições, adições, eliminações e/ou inserções, como adicionalmente descrito abaixo, de um modo preferido, de forma a que a imunogenicidade do polipéptido codificado não seja diminuída em relação a uma proteína nativa. O efeito sobre a imunogenicidade do polipéptido codificado pode ser, geralmente,

avaliado como aqui descrito. O termo “variantes” abrange também genes homólogos da origem xenogénica.

Em formas de realização adicionais, a presente invenção proporciona polinucleótidos isolados e polipéptidos compreendendo vários comprimentos de prolongamentos contíguos da sequência idêntica ou complementar a uma ou mais das sequências aqui divulgadas. Por exemplo, são proporcionados por esta invenção polinucleótidos compreendendo, pelo menos, cerca de 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 ou 1000 ou mais nucleótidos contíguos de uma ou mais das sequências aqui divulgadas, assim como todos os comprimentos intermédios aí compreendidos. Será facilmente entendido que “comprimentos intermédios”, neste contexto, significam qualquer comprimento entre os valores referidos, tais como 16, 17, 18, 19, etc.; 21, 22, 23, etc.; 30, 31, 32, etc.; 50, 51, 52, 53, etc.; 100, 101, 102, 103, etc.; 150, 151, 152, 153, etc.; incluindo todos os números inteiros de 200-500; 500-1000 e semelhantes.

Os polinucleótidos da presente invenção ou os seus fragmentos, independentemente do comprimento da própria sequência codificante, podem ser combinados com outras sequências de ADN, tais como promotores, sinais de poliadenilação, locais de enzima de restrição adicionais, locais de clonagem múltipla, outros segmentos codificantes e semelhantes, de forma a que o seu comprimento total possa variar consideravelmente. Consequentemente, é contemplado que possa ser utilizado um fragmento de ácido nucleico de quase qualquer comprimento, sendo o comprimento total de um modo preferido, limitado pela facilidade de preparação e utilização no protocolo de ADN recombinante pretendido. Por exemplo, os segmentos de ADN ilustrativos com comprimentos totais de cerca de 10000, cerca de

5000, cerca de 3000, cerca de 2000, cerca de 1000, cerca de 500, cerca de 200, cerca de 100, cerca de 50 pares de bases de comprimento e semelhantes, (incluindo todos os comprimentos intermédios) são contemplados como sendo úteis em muitas formas de realização desta invenção.

Além disso, será tido em consideração pelo especialista na técnica que, em consequência da degenerescência do código genético, existem muitas sequências nucleotídicas codificando um polipéptido, como aqui descrito. Alguns destes polinucleótidos possuem uma homologia mínima com sequências de nucleótidos de qualquer gene nativo. No entanto, são contemplados especificamente polinucleótidos que variam devido a diferenças na utilização de codões pela presente invenção, por exemplo, polinucleótidos que são otimizados para a selecção de codões de primata e/ou de humano. Além disso, os alelos dos genes compreendendo as sequências polinucleotídicas aqui proporcionadas estão dentro do âmbito da presente invenção. Os alelos são genes endógenos que são alterados em consequência de uma ou mais mutações, tais como eliminações, adições e/ou substituições de nucleótidos. O ARNm e proteína resultantes podem, mas não é necessário, ter uma estrutura ou função alterada. Podem ser identificados alelos utilizando técnicas padrão (tais como hibridação, amplificação e/ou comparação de sequências de bases de dados).

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE POLINUCLEÓTIDOS

Os polinucleótidos podem ser identificados, preparados e/ou manipulados utilizando qualquer das várias técnicas bem estabelecidas.

Os polinucleótidos podem ser amplificados a partir de ADNc preparado a partir de células que expressam as proteínas aqui descritas, tal como células de *M. tuberculosis*. Esses polinucleótidos podem ser amplificados através da reacção em cadeia da polimerase (PCR). Nesta abordagem, podem ser concebidos iniciadores específicos para uma sequência com base nas sequências aqui proporcionadas e podem ser adquiridos ou sintetizados.

Pode ser utilizada uma porção amplificada de um polinucleótido da presente invenção para isolar um gene completo a partir de uma biblioteca adequada (e. g., uma biblioteca de ADNc de *M. tuberculosis*) utilizando técnicas bem conhecidas. Dentro dessas técnicas, é rastreada uma biblioteca (ADNc ou genómico) utilizando uma ou mais sondas polinucleotídicas ou iniciadores adequados para a amplificação. De um modo preferido, uma biblioteca é seleccionada por tamanho para incluir moléculas maiores. Podem ser também preferidas bibliotecas iniciadas aleatoriamente para identificar regiões 5' e a montante de genes. São preferidas bibliotecas genómicas para obter intrões e prolongar sequências 5'.

Para técnicas de hibridação, pode ser marcada uma sequência parcial (e. g., por tradução por corte ou marcador da extremidade com ³²P) utilizando técnicas bem conhecidas. É então geralmente, rastreada uma biblioteca bacteriana ou a biblioteca de bacteriófago por hibridação de filtros contendo colónias bacterianas desnaturadas (ou plaqueamentos contendo placas fágicas) com a sonda marcada (ver Sambrook et al., *Molecular Cloning: A Laboratory Manual* (1989)). São seleccionadas colónias ou placas que hibridam e são expandidas e o ADN é isolado para uma nova análise. Os clones de ADNc podem ser analisados para

determinar a quantidade de sequência adicional, por exemplo, por PCR utilizando um iniciador para a sequência parcial e um iniciador para o vector. Podem ser produzidos mapas de restrição e sequências parciais para identificar um ou mais clones em sobreposição. A sequência completa pode ser então determinada utilizando técnicas padrão, as quais podem envolver a produção de uma série de clones de eliminação. As sequências sobrepostas resultantes podem ser então montadas numa única sequência contígua. Pode ser produzida uma molécula ADNc completa por ligação de fragmentos adequados, utilizando técnicas bem conhecidas.

Alternativamente, existem numerosas técnicas de amplificação para obter uma sequência codificante completa a partir de uma sequência de ADNc parcial. Dentro dessas técnicas, a amplificação é, geralmente, realizada através de PCR. Pode ser utilizado qualquer de vários kits comercialmente disponíveis para realizar o passo de amplificação. Os iniciadores podem ser concebidos utilizando, por exemplo, software bem conhecido na técnica. Os iniciadores tem de um modo preferido, 22-30 nucleótidos de comprimento, têm um conteúdo de GC de, pelo menos, 50% e emparelham com a sequência alvo a temperaturas de cerca de 68 °C a 72 °C. A região amplificada pode ser sequenciada como descrito acima e as sequências sobrepostas podem ser montadas numa sequência contígua.

Uma dessas técnicas de amplificação é a PCR inversa (ver Triglia et al., *Nucl. Acids Res.* **16**:8186 (1988)), a qual utiliza enzimas de restrição para produzir um fragmento na região conhecida do gene. O fragmento é então circularizado por ligação intramolecular e é utilizado como um molde para PCR com iniciadores divergentes derivados da região conhecida. Dentro de

uma abordagem alternativa, podem ser obtidas sequências adjacentes a uma sequência parcial por amplificação com um iniciador para uma sequência de um elemento de ligação e um iniciador específico para uma região conhecida. As sequências amplificadas são tipicamente submetidas a um segundo ciclo de amplificação com o mesmo iniciador para o elemento de ligação e um segundo iniciador específico para região conhecida. No documento WO 96/38591 é descrita uma variação deste processo que utiliza dois iniciadores que iniciam extensão em direcções opostas da sequência conhecida. Uma outra dessas técnicas é conhecida como “amplificação rápida de extremidades de ADNc” ou RACE. Esta técnica envolve a utilização de um iniciador interno e um iniciador externo que hibrida com uma região poliA ou sequência de vector, para identificar sequências que se situam 5' e 3' numa sequência conhecida. As técnicas adicionais incluem a captura por PCR (Lagerstrom et al., *PCR Methods Applic.* **1**:111-19 (1991)) e PCR andante (Parker et al., *Nucl. Acids. Res.* **19**:3055-60 (1991)). Podem ser também utilizados outros métodos utilizando amplificação para obter uma sequência de ADNc completa.

Em determinados casos, é possível obter uma sequência de ADNc completa por análise de sequências proporcionadas numa base de dados de marcadores de sequências expressas (EST), tal como a disponível do GenBank. As pesquisas de EST sobrepostas podem ser, geralmente, realizadas utilizando programas bem conhecidos (e.g., pesquisas BLAST do NCBI) e essas EST podem ser utilizadas para produzir uma sequência completa contígua. Podem ser também obtidas sequências de ADN completas por análise de fragmentos genómicos.

EXPRESSÃO DE POLINUCLEÓTIDO EM CÉLULAS HOSPEDEIRAS

Em outras formas de realização da invenção, podem ser utilizadas sequências polinucleotídicas ou os seus fragmentos que codificam polipéptidos da invenção ou proteínas de fusão ou os seus equivalentes funcionais, em moléculas de ADN recombinantes para controlar a expressão de um polipéptido em células hospedeiras adequadas. Devido à degenerescência inerente do código genético, podem ser produzidas outras sequências de ADN codificando substancialmente a mesma ou uma sequência de aminoácidos funcionalmente equivalente e estas sequências podem ser utilizadas para clonar e expressar um determinado polipéptido.

Como será entendido pelo especialista na técnica, pode ser vantajoso em alguns casos produzir sequências nucleotídicas que codificam polipéptidos possuindo códons de ocorrência não natural. Por exemplo, podem ser seleccionados os códons preferidos por um hospedeiro procariota ou eucariota particular, para aumentar a taxa de expressão da proteína ou para produzir um transcrito de ARN recombinante tendo propriedades pretendidas, tal como uma meia-vida que seja mais longa do que a de um transcrito produzido a partir da sequência de ocorrência natural.

Além disso, as sequências polinucleotídicas da presente invenção podem ser concebidas utilizando métodos, geralmente, conhecidos na técnica de forma a alterar as sequências codificantes do polipéptido por vários motivos, incluindo mas não limitados a alterações que modificam a clonagem, o processamento e/ou a expressão do produto génico. Por exemplo, pode ser utilizado rearranjo do ADN por fragmentação aleatória e

remontagem de fragmentos gênicos e oligonucleótidos sintéticos por PCR para modificar as sequências nucleotídicas. Além disso, pode ser utilizada mutagênese dirigida para inserir novos locais de restrição, alterar padrões de glicosilação, modificar a preferência de códons, produzir variantes de processamento ou introduzir mutações e semelhantes.

As sequências de ácidos nucleicos naturais, modificadas ou recombinantes podem ser ligadas a uma sequência heteróloga para codificar uma proteína de fusão.

Uma proteína de fusão pode ser também modificada para conter um local de degradação localizado entre a sequência codificando o polipéptido e a sequência da proteína heteróloga, de forma a que o polipéptido possa ser quebrado e purificado a partir da parte heteróloga.

Podem ser sintetizadas sequências codificando um polipéptido pretendido, na totalidade ou em parte, utilizando métodos químicos bem conhecidos na técnica (ver Caruthers, M. H. *et al.*, *Nucl. Acids Res. Symp. Ser.* pp. 215-223 (1980), Horn *et al.*, *Nucl. Acids Res. Symp. Ser.* pp. 225-232 (1980)). Alternativamente, a própria proteína pode ser produzida utilizando métodos químicos para sintetizar a sequência de aminoácidos de um polipéptido ou uma sua porção. Por exemplo, pode ser realizada síntese de péptidos utilizando várias técnicas em fase sólida (Roberge *et al.*, *Science* **269**:202-204 (1995)) e pode ser realizada síntese automatizada utilizando, por exemplo, o ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

Um péptido sintetizado de novo pode ser substancialmente purificado por cromatografia líquida preparativa de elevado desempenho (e. g., Creighton, *Proteins, Structures and Molecular Principles* (1983)) ou outras técnicas comparáveis disponíveis na técnica. A composição dos péptidos sintéticos pode ser confirmada por análise de aminoácidos ou sequenciação (e. g., processo de degradação de Edman). Adicionalmente, a sequência de aminoácidos de um polipéptido ou de qualquer sua parte, pode ser alterada durante a síntese directa e/ou ser combinada utilizando, métodos químicos com sequências de outras proteínas ou uma sua qualquer parte, para produzir um polipéptido variante.

As sequências nucleotídicas codificando o polipéptido ou equivalentes funcionais, podem ser inseridas no vector de expressão adequado, i. e., um vector que contenha os elementos necessários para a transcrição e a tradução da sequência codificante inserida, para expressar um polipéptido pretendido. Podem ser utilizados métodos que são bem conhecidos do especialista na técnica para construir vectores de expressão contendo sequências codificando um polipéptido de interesse e elementos de controlo da transcrição e da tradução adequados. Estes métodos incluem técnicas de ADN recombinante *in vitro*, técnicas de síntese e recombinação genética *in vivo*. Essas técnicas são descritas em Sambrook et al., *Molecular Cloning, A Laboratory Manual* (1989) e Ausubel et al., *Current Protocols in Molecular Biology* (1989).

Podem ser utilizados vários sistemas de vector de expressão/hospedeiro para conter e expressar sequências polinucleotídicas. Estas incluem, mas não estão limitadas a microrganismos, tais como bactérias transformadas com vectores

de expressão de ADN de bacteriófago recombinante, plasmídeo ou cosmídeo; leveduras transformadas com vectores de expressão de levedura; sistemas de células de insecto infectadas com vectores de expressão de vírus (e. g., baculovírus); sistemas de células vegetais transformadas com vectores de expressão de vírus (e. g., vírus do mosaico da couve-flor, CaMV; vírus do mosaico do tabaco, TMV) ou com vectores de expressão bacterianos (e. g., plasmídeos Ti ou pBR322); ou sistemas de células animais.

Os “elementos de controlo” ou “sequências reguladoras” presentes num vector de expressão são regiões não-traduzidas do vector - intensificadores, promotores, regiões 5' e 3' não traduzidas - que interagem com as proteínas celulares do hospedeiro para realizar a transcrição e a tradução. Esses elementos podem variar na sua força e especificidade. Dependendo do sistema de vector e hospedeiro utilizado, pode ser utilizado qualquer número elementos de transcrição e de tradução adequados, incluindo promotores constitutivos e induzíveis. Por exemplo, para clonagem em sistemas bacterianos podem ser utilizados promotores induzíveis, tais como o promotor híbrido lacZ do fagemídeo PBLUESCRIPT (Stratagene, La Jolla, Calif.) ou o plasmídeo PSPORT1 (Gibco BRL, Gaithersburg, MD) e semelhantes. Em sistemas de célula de mamífero, são, geralmente, preferidos promotores de genes de mamífero ou de vírus de mamífero. Se for necessário produzir uma linha celular que contenha múltiplas cópias da sequência codificando um polipéptido, podem ser vantajosamente utilizados os vectores baseados em SV40 ou EBV com um marcador seleccionável adequado.

Em sistemas bacterianos podem ser seleccionados vários vectores de expressão dependendo da utilização pretendida para o polipéptido expresso. Por exemplo, quando são necessárias

grandes quantidades, por exemplo, para a indução de anticorpos, podem ser utilizados vectores que regulam a expressão em níveis elevados de proteínas de fusão que são facilmente purificadas. Esses vectores incluem, mas não estão limitados aos vectores de clonagem e de expressão multifuncionais de *E. coli*, tal como BLUESCRIPT (Stratagene), nos quais a sequência codificando o polipéptido de interesse pode ser ligada ao vector na grelha de leitura correcta com sequências para o terminal amina Met e os 7 resíduos subsequentes da β -galactosidase para que seja produzida uma proteína híbrida; vectores pIN (Van Heeke & Schuster, *J. Biol. Chem.* **264**:5503-5509 (1989)); e semelhantes. Podem ser também utilizados vectores pGEX (Promega, Madison, Wis.) para expressar polipéptidos exógenos como proteínas de fusão com S-transferase da glutathiona (GST). Em geral, essas proteínas de fusão são solúveis e podem ser facilmente purificadas a partir de células lisadas por adsorção a esferas de glutathiona-agarose seguida por eluição na presença de glutathiona livre. As proteínas preparadas nesses sistemas podem ser concebidas de forma a incluir locais de degradação da heparina, trombina ou da protease factor XA para que o polipéptido clonado de interesse possa ser libertado da parte GST sem restrições.

Podem ser utilizados vários vectores contendo promotores constitutivos ou induzíveis, tais como factor alfa, álcool oxidase e PGH na levedura *Saccharomyces cerevisiae*. Para revisões, ver Ausubel *et al.* (*supra*) e Grant *et al.*, *Methods Enzymol.* **153**:516-544 (1987).

Nos casos em que são utilizados vectores de expressão vegetais, a expressão de sequências codificando polipéptidos pode ser dirigida por qualquer de vários promotores. Por

exemplo, podem ser utilizados promotores virais, tais como os promotores 35S e 19S do CaMV, isoladamente ou em combinação, com a sequência líder ómega do TMV (Takamatsu, *EMBO J.* **6**:307-311 (1987)). Alternativamente, podem ser utilizados promotores vegetais, tais como os promotores da subunidade pequena da RUBISCO ou de choque térmico (Coruzzi et al., *EMBO J.* **3**:1671-1680 (1984); Broglie et al., *Science* **224**:838-843 (1984); e Winter et al., *Results Probl. Cell Differ.* **17**:85-105 (1991)). Estas construções podem ser introduzidas em células vegetais por transformação de ADN directa ou transfecção mediada por um agente patogénico. Essas técnicas são descritas, de um modo geral, em várias revisões disponíveis (ver e. g., Hobbs em *McGraw Hill Yearbook of Science and Technology* 191-196 (1992)).

Pode ser também utilizado um sistema de insecto para expressar um polipéptido de interesse. Por exemplo, num desses sistemas, é utilizado o vírus da poli-hedrose nuclear (AcNPV) de *Autographa californica* como um vector para expressar genes exógenos em células de *Spodoptera frugiperda* ou em *Trichoplusia larvae*. As sequências codificando o polipéptido podem ser clonadas numa região não essencial do vírus, tal como o gene poli-hedrina e ser colocadas sob o controlo do promotor da poli-hedrina. A inserção com sucesso da sequência codificando polipéptido irá tornar o gene da poli-hedrina inactivo e irá produzir o vírus recombinante isento de proteína do invólucro. Os vírus recombinantes podem ser então utilizados para infectar, por exemplo, células de *S. frugiperda* ou *Trichoplusia larvae* nas quais o polipéptido de interesse pode ser expresso (Engelhard et al., *Proc. Natl. Acad. Sci. U.S.A.* **91**:3224-3227 (1994)).

De um modo geral, estão disponíveis vários sistemas de expressão viral em células hospedeiras de mamífero. Por exemplo,

nos casos em que é utilizado um adenovírus como vector de expressão, as sequências codificando um polipéptido de interesse podem ser ligadas num complexo de transcrição/tradução de adenovírus consistindo no promotor tardio e na sequência de líder tripartida. A inserção numa região E1 ou E3 não essencial do genoma viral pode ser utilizada para obter um vírus viável, o qual é capaz de expressar o polipéptido em células hospedeiras infectadas (Logan e Shenk, *Proc. Natl. Acad. Sci. U.S.A.* **81**:3655-3659 (1984)). Além disso, podem ser utilizados intensificadores da transcrição, tal como o intensificador do vírus do sarcoma de Rous (RSV), para aumentar a expressão em células hospedeiras de mamífero.

Podem ser também utilizados sinais de iniciação específicos para realizar a tradução mais eficiente de sequências codificando um polipéptido de interesse. Esses sinais incluem o codão de iniciação ATG e sequências adjacentes. Nos casos em que são inseridas no vector de expressão adequado, sequências codificando o polipéptido, o seu codão de iniciação e a as sequências a montante, pode não ser necessário qualquer sinal de controlo da transcrição ou da tradução adicional. No entanto, nos casos em que apenas é inserida a sequência codificante ou uma sua porção, devem ser proporcionados sinais de controlo de tradução exógenos, incluindo o codão de iniciação ATG. Além disso, o codão de iniciação deve estar na grelha de leitura correcta para assegurar a tradução da totalidade da inserção. Os elementos de tradução exógenos e os codões de iniciação podem ter várias origens, naturais e de síntese. A eficiência da expressão pode ser intensificada pela inclusão de intensificadores que sejam adequados para o sistema de células particular que é utilizado, tal como os descritos na literatura (Scharf. *et al.*, *Results Probl. Cell Differ.* **20**:125-162 (1994)).

Além disso, pode ser seleccionada uma estirpe de células hospedeiras pela sua capacidade para modular a expressão das sequências inseridas ou para processar a proteína expressa de uma forma pretendida. Essas modificações do polipéptido incluem, mas não estão limitadas a acetilação, carboxilação, glicosilação, fosforilação, lipidação e acilação. Pode ser também utilizado processamento pós-tradução que quebra uma forma "pré-pro" da proteína para facilitar a inserção correcta, enrolamento e/ou função. Podem ser seleccionadas células hospedeiras diferentes, tais como CHO, HeLa, MDCK, HEK293 e WI38, que têm maquinaria celular específica e mecanismos característicos para essas actividades pós-tradução, para assegurar a modificação e o processamento correctos da proteína exógena.

Para a produção de elevado rendimento a longo prazo de proteínas recombinantes, é, geralmente, preferida a expressão estável. Por exemplo, as linhas celulares que expressam um polinucleótido de interesse de forma estável podem ser transformadas utilizando vectores de expressão que podem conter origens de replicação virais e/ou elementos de expressão endógenos e um gene marcador seleccionável no mesmo vector ou num vector separado. Após a introdução do vector, pode permitir-se o crescimento das células durante 1-2 dias em meios enriquecidos antes destas serem transferidas para meios selectivos. O objectivo do marcador seleccionável é conferir resistência à selecção e a sua presença permite o crescimento e a recuperação de células que expressam as sequências introduzidas com sucesso. Os clones resistentes de células transformadas de forma estável podem ser propagados utilizando técnicas de cultura de tecidos adequadas ao tipo celular.

Pode ser utilizado qualquer de vários sistemas de selecção para recuperar linhas celulares transformadas. Estes incluem, mas não estão limitados aos genes da cinase da timidina do vírus herpes simplex (Wigler et al., *Cell* **11**:223-32 (1977)) e da fosforribosiltransferase da adenina (Lowy et al., *Cell* **22**:817-23 (1990)) os quais podem ser utilizados em, respectivamente células tk.sup.-ou aprt.sup.-. Pode ser também utilizada resistência a antimetabolitos, antibióticos ou herbicidas como base para a selecção; por exemplo, dhfr que confere resistência ao metotrexato (Wigler et al., *Proc. Natl. Acad. Sci. U.S.A.* **77**:3567-70 (1980)); npt que confere a resistência aos aminoglicósidos, neomicina e G-418 (Colbere-Garapin et al., *J. Mol. Biol.* **150**:1-14 (1981)); e als ou pat, que conferem resistência, respectivamente, ao clorsulfurão e à acetiltransferase da fosfinotricina (Murry, *supra*). Foram descritos genes seleccionáveis adicionais, por exemplo, trpB, que permite às células utilizarem indole em vez de triptofano ou hisD, que permite às células utilizarem histinol em vez de histidina (Hartman e Mulligan, *Proc. Natl. Acad. Sci. U.S.A.* **85**:8047-51 (1988)). Recentemente, a utilização de marcadores visíveis ganhou popularidade sendo esses marcadores antocianinas, a β -glucuronidase e o seu substrato GUS e a luciferase e o seu substrato luciferina, sendo amplamente utilizado não apenas para identificar transformantes, mas também para quantificar a quantidade da expressão proteica transiente ou estável atribuível a um sistema de vector específico (Rhodes et al., *Methods Mol. Biol.* **55**:121-131 (1995)).

Embora a presença/ausência da expressão de um gene marcador sugira que o gene de interesse esteja também presente, pode ser necessário confirmar a sua presença e expressão. Por exemplo, se

a sequência codificando um polipéptido é inserida dentro de uma sequência de um gene marcador, as células recombinantes contendo sequências, podem ser identificadas pela ausência da função do gene marcador. Alternativamente, pode ser colocado um gene marcador em tandem com uma sequência codificando um polipéptido sob o controlo de um único promotor. A expressão do gene marcador em resposta à indução ou à selecção indica normalmente a expressão também do gene em tandem.

Alternativamente, as células hospedeiras que contêm e expressam uma sequência polinucleotídica pretendida podem ser identificadas por vários processos conhecidos do especialista na técnica. Estes processos incluem, mas não estão limitados a hibridações ADN-ADN ou ARN-ADN e técnicas de bioensaio ou imunoensaio de proteínas, as quais incluem tecnologias baseadas em membranas, soluções ou placas para a detecção e/ou quantificação de ácido nucleicos ou proteínas.

São conhecidos na técnica vários protocolos para detectar e medir a expressão de produtos codificados por polinucleótidos, utilizando anticorpos policlonais ou monoclonais específicos para o produto. Os exemplos incluem ensaio imunoabsorvente ligado a enzima (ELISA), radioimunoensaio (RIA) e separação celular activada por fluorescência (FACS). Para algumas aplicações pode ser preferido, um imunoensaio baseado em monoclonal de dois locais, utilizando anticorpos monoclonais reactivos com dois epitopos não-interferentes num determinado polipéptido, mas pode ser também utilizado um ensaio de ligação competitivo. Estes e outros ensaios são descritos em Hampton *et al.*, *Serological Methods, a Laboratory Manual* (1990) e Maddox *et al.*, *J. Exp. Med.* **158**:1211-1216 (1983), entre outros locais.

É conhecida do especialista na técnica uma ampla variedade de marcadores e técnicas de conjugação e podem ser utilizadas em vários ensaios de ácido nucleicos e de aminoácidos. Os meios para produzir sondas de hibridação ou de PCR marcadas para detectar sequências relacionadas com polinucleótidos incluem a marcador de oligonucleótidos, tradução por corte, marcador de extremidades ou amplificação por PCR utilizando um nucleótido marcado. Alternativamente, as sequências ou qualquer sua porção podem ser clonadas num vector para a produção de uma sonda de ARNm. Esses vectores são conhecidos na técnica, estão comercialmente disponíveis e podem ser utilizados para sintetizar sondas de ARN *in vitro* pela adição de uma polimerase de ARN adequada, tais como de T7, T3 ou SP6 e nucleótidos marcados. Estes processos podem ser realizados utilizando vários kits comercialmente disponíveis. As moléculas repórter ou marcadores adequados que pode ser utilizadas incluem radionuclídeos, enzimas, agentes fluorescentes, quimioluminescentes ou cromogénicos, assim como substratos, cofactores, inibidores, partículas magnéticas e semelhantes.

As células hospedeiras transformadas com uma sequência polinucleotídica de interesse podem ser cultivadas sob condições adequadas para a expressão e a recuperação da proteína a partir da cultura celular. A proteína produzida por uma célula recombinante pode ser segregada ou ser contida intracelularmente dependendo da sequência e/ou do vector utilizado. Como será entendido pelo especialista na técnica, os vectores de expressão contendo os polinucleótidos da invenção podem ser concebidos para conter sequências sinal que direccionam a secreção do polipéptido codificado através de uma membrana celular procariota ou eucariota. Podem ser utilizadas outras construções recombinantes para ligar sequências codificando um polipéptido

de interesse à sequência de nucleótidos codificando um domínio polipeptídico que irá facilitar a purificação de proteínas solúveis. Esses domínios facilitadores da purificação incluem, mas não estão limitados a péptidos quelantes de metais, tal como módulos histidina-triptofano que permitem a purificação em metais imobilizados, domínios de proteína A que permitem a purificação em imunoglobulina imobilizada e o domínio utilizado no sistema de purificação de extensão/afinidade FLAGS (Immunex Corporation, Seattle, Wash.). Pode ser utilizada a inclusão de sequências de ligação quebráveis, tais como as específicas para o Factor XA ou a enterocinase (Invitrogen, San Diego, Calif.) entre o domínio de purificação e o polipéptido codificado para facilitar a purificação. Um desses vectores de expressão proporciona a expressão de uma proteína de fusão contendo um polipéptido de interesse e um ácido nucleico codificando 6 resíduos de histidina precedendo um local de degradação de tiorredoxina ou de enterocinase. Os resíduos de histidina facilitam a purificação em IMIAC (cromatografia de afinidade de ião metálico imobilizado) como descrito em Porath *et al.*, *Prot. Exp. Purif.* **3**:263-281 (1992) enquanto que o local de degradação da enterocinase proporciona um meio para purificar o polipéptido pretendido a partir da proteína de fusão. Em Kroll *et al.*, *DNA Cell Biol.* **12**:441-453 (1993)) é proporcionada uma discussão de vectores que contêm proteínas de fusão.

Os polipéptidos da invenção e seus fragmentos, podem ser produzidos por síntese peptídica directa utilizando técnicas em fase sólida (Merrifield, *J. Am. Chem. Soc.* **85**:2149-2154 (1963)), além de através de métodos de produção recombinantes. A síntese proteica pode ser realizada utilizando técnicas manuais ou por automação. A síntese automatizada pode ser realizada, por exemplo, utilizando Applied Biosystems 431A Peptide Synthesizer

(Perkin Elmer). Alternativamente, podem ser sintetizados quimicamente separadamente vários fragmentos e serem combinados utilizando métodos químicos para produzir a molécula completa.

TÉCNICAS DE TRANSFERÊNCIA DE POLINUCLEÓTIDOS *IN VIVO*

Em formas de realização adicionais, são introduzidas construções genéticas compreendendo um ou mais dos polinucleótidos da invenção em células *in vivo*. Isto pode ser realizado utilizando qualquer de várias ou bem conhecidas abordagens, algumas delas são delineadas abaixo para objectivos de ilustração.

1. ADENOVÍRUS

Um dos métodos preferidos para a transferência *in vivo* de uma ou mais sequências de ácidos nucleicos envolve a utilização de um vector de expressão de adenovírus. "Vector de expressão de adenovírus" pretende incluir as construções contendo sequências de adenovírus suficientes para (a) suportar o empacotamento da construção e (b) expressar um polinucleótido que foi aí clonado com uma orientação sentido ou anti-sentido. Obviamente, no contexto de uma construção anti-sentido, a expressão não requer que o produto génico seja sintetizado.

O vector de expressão compreende uma forma geneticamente modificada de um adenovírus. O conhecimento da organização genética do adenovírus, um vírus de ADN de cadeia dupla, linear, de 36 kb, permite a substituição de grandes partes do ADN adenoviral por sequências exógenas até 7 kb (Grunhaus e Horwitz,

1992). Ao contrário dos retrovírus, a infecção adenoviral de células hospedeiras não resulta na integração cromossômica porque o ADN adenoviral pode duplicar de uma forma epissômica sem potencial genotoxicidade. Os adenovírus são também estruturalmente estáveis e não tem sido detectado qualquer rearranjo do genoma após amplificação extensa. O adenovírus pode infectar praticamente todas as células epiteliais independentemente da sua etapa do ciclo celular. Até agora, a infecção adenoviral parece estar apenas ligada a doença benigna, tal como doença respiratória aguda em humanos.

O adenovírus é particularmente adequado para utilização como um vector de transferência génica devido ao seu genoma de tamanho médio, facilidade de manipulação, título elevado, ampla gama de células alvo e infecciosidade elevada. Ambas as extremidades do genoma viral contêm repetições invertidas de 100-200 pares de bases (ITR) que são elementos em cis necessários para a replicação e o empacotamento do ADN viral. As regiões precoce (E) e tardia (L) do genoma contêm unidades de transcrição diferentes que são divididas pelo estabelecimento da replicação de ADN viral. A região E1 (E1A e E1B) codifica proteínas responsáveis pela regulação da transcrição do genoma viral e alguns genes celulares. A expressão da região E2 (E2A e E2B) resulta na síntese das proteínas de replicação do ADN viral. Estas proteínas estão envolvidas na replicação do ADN, expressão de genes tardios e inactivação da célula hospedeira (Renan, 1990). Os produtos dos genes tardios, incluindo a maioria das proteínas da cápside viral, são expressos apenas após o processamento significativo de um único transcrito primário emitido pelo promotor tardio principal (MLP). O MLP, (localizado nas 16,8 u. m.) é particularmente eficiente durante a fase tardia da infecção e todos os ARNm emitido a partir deste

promotor possui uma sequência líder tripartida (TPL) 5' a qual os torna os ARNm preferidos para a tradução.

Num sistema actual, o adenovírus recombinante é produzido a partir da recombinação homóloga entre vector transportador e vector de proviral. Devido à possível recombinação entre dois vectores provirais, pode ser produzido adenovírus de tipo selvagem a partir deste processo. Consequentemente, é crucial isolar um clone único do vírus a partir de uma placa individual e examinar a sua estrutura genómica.

A produção e a propagação dos vectores adenovirais actuais, os quais são deficientes para a replicação, dependem de uma única linha de células auxiliadoras, designada 293, a qual foi transformada a partir de células embrionárias de rim humano com fragmentos de ADN Ad5 e expressa constitutivamente proteínas E1 (Graham *et al.*, 1977). Uma vez que a região E3 é dispensável do genoma do adenovírus (Jones e Shenk, 1978), os vectores adenovirais actuais, com o auxílio de células 293, contém ADN exógeno quer em E1, D3 ou em ambas as regiões (Graham e Prevec, 1991). Na natureza, o adenovírus pode empacotar aproximadamente 105% do genoma de tipo selvagem (Ghosh-Choudhury *et al.*, 1987), proporcionando capacidade de cerca de 2 kb extra de ADN. Combinado com os aproximadamente 5,5 kb do ADN que são substituíveis nas regiões E1 e E3, a capacidade máxima do vector de adenovírus actual é inferior a 7,5 kb ou cerca de 15% do comprimento total do vector. Mais de 80% do genoma viral do adenovírus permanece na estrutura base do vector e é a fonte de citotoxicidade contida no vector. A deficiência para a replicação do vírus eliminado de E1 também é incompleta. Por exemplo, foi observada uma fuga de expressão génica viral com os

vectores actualmente disponíveis em multiplicidades elevadas de infecção (MOI) (Mulligan, 1993).

As linhas celulares auxiliaadoras podem ser derivadas de células humanas, tais como células embrionárias de rim humano, células musculares, células hematopoiéticas ou outras células embrionárias humanas mesenquimatosas ou epiteliais. Alternativamente, as células auxiliaadoras podem ser derivadas de células de outras espécies de mamífero que são permissivas ao adenovírus humano. Essas células incluem, *e. g.*, células Vero ou outras células embrionárias mesenquimatosas ou epiteliais de macaco. Como mencionado acima, a linha celular auxiliadora actualmente preferida é a 293.

Recentemente, Racher *et al.* (1995) divulgou métodos melhorados para cultivar células 293 e propagar adenovírus. Num formato, os agregados celulares naturais são cultivados inoculando células individuais em frascos de centrífuga siliconizados de 1 litro (Techne, Cambridge, UK) contendo 100-200 mL de meio. Após agitação a 40 rpm, a viabilidade celular é estimada com azul tripiano. Num outro formato, são utilizados microsuportes Fibra-Cel (Bibby Sterlin, Stone, UK) (5 g/L) como se segue. É adicionado um inóculo celular, ressuspenso em 5 mL de meio ao suporte (50 mL) num frasco Erlenmeyer de 250 mL e é deixado a repousar, com agitação ocasional, durante 1 a 4 h. O meio é então substituído por 50 mL de meio fresco e é iniciada a agitação. Para a produção do vírus, é permitido o crescimento celular até uma confluência de cerca de 80%, depois desse tempo o meio é substituído (até 25% do volume final) e é adicionado adenovírus num MOI de 0,05. As culturas são deixadas estacionárias durante a noite, após o que

o volume é aumentado para 100% e é iniciada agitação durante mais 72 h.

Além da exigência de que o vector de adenovírus seja deficiente para a replicação ou, pelo menos, condicionalmente deficiente, não se considera que a natureza do vector de adenovírus seja crucial para a realização, com sucesso, da invenção. O adenovírus pode ser de qualquer dos 42 serotipos ou subgrupos A-F diferentes conhecidos. O tipo 5 de adenovírus do subgrupo C é o material inicial preferido, para obter um vector de adenovírus deficiente para a replicação condicional para utilização na presente invenção, uma vez que o tipo 5 do Adenovírus é um adenovírus humano acerca do qual é conhecida muita informação bioquímica e genética e tem sido historicamente utilizado para a maioria das construções utilizando adenovírus como um vector.

Como mencionado acima, o vector típico de acordo com a presente invenção é deficiente para a replicação e não irá ter uma região E1 do adenovírus. Deste modo, será mais conveniente introduzir o polinucleótido codificando o gene de interesse na posição a partir da qual as sequências codificando E1 foram removidas. No entanto, a posição da inserção da construção dentro das sequências do adenovírus não é crucial para a invenção. O polinucleótido codificando o gene de interesse pode ser também inserido em vez da região E3 eliminada em vectores de substituição de E3 como descrito por Karlsson *et al.* (1986) ou na região E4 em que uma linha celular auxiliadora ou um vírus auxiliador completam a deficiência de E4.

O adenovírus é fácil de cultivar e de manipular e apresenta um ampla gama de hospedeiros *in vitro* e *in vivo*. Este grupo de

vírus pode ser obtido em títulos elevados, e. g., 10^9 - 10^{11} unidades formadoras de placa por mL e estes são altamente infecciosos. O ciclo de vida do adenovírus não requer a integração no genoma da célula hospedeira. Os genes exógenos distribuídos por vectores adenovirais são epissómicos e, conseqüentemente, têm uma baixa genotoxicidade para as células hospedeiras. Não foi descrito qualquer efeito secundário em estudos da vacinação com o adenovírus de tipo selvagem (Couch *et al.*, 1963; Top *et al.*, 1971), demonstrando o seu potencial de segurança e terapêutico como vectores de transferência génica *in vivo*.

Têm sido utilizados vectores adenovirais em expressão génica eucariota (Levrero *et al.*, 1991; Gomez-Foix *et al.*, 1992) e no desenvolvimento de vacinas (Grunhaus e Horwitz, 1992; Graham e Prevec, 1992). Recentemente, estudos em animais sugeriram que possa ser utilizado adenovírus recombinante em terapia genética (Stratford-Perricaudet e Perricaudet, 1991; Stratford-Perricaudet *et al.*, 1990; Rich *et al.*, 1993). Os estudos de administração de adenovírus recombinante a diferentes tecidos incluem a instilação pela traqueia (Rosenfeld *et al.*, 1991; Rosenfeld *et al.*, 1992), injeção muscular (Ragot *et al.*, 1993), injeções intravenosas periféricas (Herz e Gerard, 1993) e inoculação estereotáctica no cérebro (Le Gall de La Salle *et al.*, 1993).

2. RETROVÍRUS

Os retrovírus são um grupo de vírus de ARN de cadeia simples caracterizados por uma capacidade para converter o seu ARN em ADN de cadeia dupla em células infectadas por um processo

da transcrição inversa (Coffin, 1990). O ADN resultante integra-se então de forma estável nos cromossomas celulares como um pró-vírus e controla a síntese de proteínas virais. A integração resulta na retenção das sequências génicas virais na célula receptora e nos seus descendentes. O genoma retroviral contém três genes, gag, pol e env codificando, respectivamente, para proteínas da cápside, enzima polimerase e componentes do invólucro. Uma sequência encontrada a montante do gene gag contém um sinal de empacotamento do genoma em viriões. Estão presentes duas sequências de repetições terminais longas (LTR) nas extremidades 5' e 3' do genoma viral. Estas contêm um promotor forte e sequências intensificadoras e são também necessárias para a integração no genoma da célula hospedeira (Coffin, 1990).

Para construir um vector retroviral, é inserido um ácido nucleico codificando uma ou mais sequências oligonucleotídicas ou polinucleotídicas de interesse no genoma viral em vez de determinadas sequências virais para produzir um vírus que é deficiente para a replicação. Para produzir viriões, é construída uma linha celular de empacotamento contendo os genes gag, pol e env mas sem a LTR e os componentes de empacotamento (Mann *et al.*, 1983). Quando um plasmídeo recombinante contendo ADNc, em conjunto com a LTR retroviral e sequências de empacotamento é introduzido nesta linha celular (por exemplo por precipitação com fosfato de cálcio), a sequência de empacotamento permite que o transcrito de ARN do plasmídeo recombinante seja empacotado em partículas virais, que são então segregadas para os meios de cultura (Nicolas e Rubenstein, 1988; Temin, 1986; Mann *et al.*, 1983). Os meios contendo o retrovírus recombinante são então recolhidos, são opcionalmente concentrados e são utilizados na transferência genética. Os

vectores retrovirais são capazes de infectar uma ampla variedade de tipos celulares. No entanto, a integração e a expressão estável requerem a divisão de células hospedeiras (Paskind *et al.*, 1975).

Foi recentemente desenvolvida uma nova abordagem concebida para permitir visar especificamente vectores retrovirais baseada na modificação química de um retrovírus pela adição química de resíduos de lactose ao invólucro viral. Esta modificação pode permitir a infecção específica dos hepatócitos através receptores sialoglicoproteicos.

Foi concebida uma abordagem diferente para visar retrovírus recombinantes em que foram utilizados anticorpos biotinilados contra uma proteína do invólucro retroviral e contra um receptor celular específico. Os anticorpos foram ligados via componentes de biotina utilizando estreptavidina (Roux *et al.*, 1989). Estes demonstraram a infecção de várias células humanas que continham aqueles antígenos de superfície com um vírus ecotrópico *in vitro* (Roux *et al.*, 1989), utilizando anticorpos contra antígenos do complexo de histocompatibilidade principal da classe I e classe II.

3. VÍRUS ADENO-ASSOCIADOS

O AAV (Ridgeway, 1988; Hermonat e Muzycska, 1984) é um parovírus, descoberto como uma contaminação de armazenamentos de adenovírus. Este é um vírus ubíquo (estão presentes anticorpos em 85% da população humana dos Estados Unidos) que não foi relacionado com qualquer doença. É também classificado como um dependovírus, uma vez que a sua replicação está dependente da

presença de um vírus auxiliador, tal como um adenovírus. Foram isolados cinco serotipos, dos quais o AAV-2 é o melhor caracterizado. O AAV tem um ADN linear de cadeia simples que está encapsidado nas proteínas da cápside VP1, VP2 e VP3 para formar um virião icosaédrico com 20 a 24 nm de diâmetro (Muzyczka e McLaughlin, 1988).

O ADN do AAV tem aproximadamente 4,7 kilobases de comprimento. Este contém duas grelhas de leitura abertas e está flanqueado por duas ITR. Existem dois genes principais no genoma do AAV: *rep* e *cap*. O gene *rep* codifica para proteínas responsáveis pela replicação viral, enquanto que o *cap* codifica para a proteína da cápside VP1-3. Cada ITR forma uma estrutura em gancho com forma de T. Estas repetições terminais são os únicos componentes em *cis* essenciais do AAV para a integração cromossómica. Consequentemente, o AAV pode ser utilizado como um vector com todas as sequências codificante virais removidas e substituídas pela cassette de genes para transferência. Foram identificados três promotores virais e foram denominados p5, p19 e p40, de acordo com a sua posição no mapa. A transcrição de p5 e p19 resulta na produção de proteínas *rep* e a transcrição de p40 produz as proteínas da cápside (Hermonat e Muzyczka, 1984).

Existem vários factores que induziram os investigadores a estudar a possibilidade da utilização de rAAV como um vector de expressão. Um destes é o facto as exigências para transferir um gene de forma a integrar-se no cromossoma do hospedeiro serem surpreendentemente reduzidas. É necessário estar presente a ITR 145-bp, a qual é apenas 6% do genoma do AAV. Isto deixa espaço no vector para a montagem de uma inserção de ADN de 4,5 kb. Embora esta capacidade de contenção possa impedir o AAV de

transferir genes grandes, é significativamente adequada para a transferência de construções anti-sentido da presente invenção.

O AAV é também uma boa selecção como veículo de transferência devido à sua segurança. Existe um mecanismo de resgate relativamente complicado: são necessários não apenas o adenovírus de tipo selvagem mas também os genes AAV para mobilizar o rAAV. Do mesmo modo, o AAV não é patogénico e não está associado a nenhuma doença. A remoção de sequências codificante virais minimiza as reacções imunitárias à expressão génica viral e, conseqüentemente, o rAAV não provoca uma resposta inflamatória.

4. OUTROS VECTORES VIRAIS COMO CONSTRUÇÕES DE EXPRESSÃO

Podem ser utilizados outros vectores virais como construções de expressão na presente invenção para a transferência de sequências oligonucleotídicas ou polinucleotídicas a uma célula hospedeira. Podem ser utilizados vectores derivados de vírus, tais como vírus vaccinia (Ridgeway, 1988; Coupar *et al.*, 1988), lentivírus, vírus da poliomielite e vírus do herpes. Estes oferecem várias características atractivas para várias células de mamífero (Friedmann, 1989; Ridgeway, 1988; Coupar *et al.*, 1988; Horwich *et al.*, 1990).

Com o reconhecimento recente do vírus da hepatite B deficiente, foi obtida uma nova percepção na relação entre a função e a estrutura de sequências virais diferentes. Os estudos *in vitro* mostraram que o vírus pode conservar a capacidade de empacotamento e de transcrição reversa dependente de auxílio apesar da eliminação de até 80% do seu genoma (Horwich *et al.*,

1990). Isto sugeriu que possam ser substituídas grandes porções do genoma por material genético exógeno. O hepatotropismo e a persistência (integração) foram propriedades particularmente atractivas para a transferência génica direccionada para o fígado. Chang et al. (1991) introduziram o gene da acetiltransferase do cloranfenicol (CAT) no genoma do vírus da hepatite B de pato em vez das sequências codificantes da polimerase, superfície e pré-superfície. Este foi cotransfectado com o vírus do tipo selvagem numa linha celular de hepatoma aviário. Foram utilizados meios de cultura contendo títulos elevados do vírus recombinante para infectar hepatócitos primários de pato recém-nascido. Foi detectada a expressão estável do gene da CAT durante, pelo menos, 24 dias após a transfecção (Chang et al., 1991).

5. VECTORES NÃO-VIRAIS

Para efectuar a expressão das sequências oligonucleotídicas ou polinucleotídicas da presente invenção, a construção de expressão deve ser transferida numa célula. Esta transferência pode ser realizada *in vitro*, como em processos laboratoriais para transformar linhas celulares ou *in vivo* ou *ex vivo*, como no tratamento de determinados estados de doença. Como descrito acima, um mecanismo preferido, para a transferência é via infecção viral, em que a construção de expressão é encapsulada numa partícula viral infecciosa.

Uma vez tendo sido transferida a construção de expressão no interior da célula o ácido nucleico codificando as sequências oligonucleotídicas ou polinucleotídicas pretendidas pode ser posicionado e ser expresso em locais diferentes. Em determinadas

formas de realização, o ácido nucleico codificando construção pode estar integrado de forma estável no genoma celular. Esta integração pode estar na localização e orientação específica via recombinação homóloga (substituição génica) ou pode estar integrada numa posição aleatória, não-específica (aumento génico). Em ainda novas formas de realização adicionais, o ácido nucleico pode ser mantido de forma estável na célula como um segmento de ADN epissómico separado. Esses segmentos de ácidos nucleicos ou “epissomas” codificam sequências suficientes para permitir a manutenção e a replicação independente ou em sincronização com o ciclo da célula hospedeira. A forma como as construções de expressão são transferidas a uma célula e em que localização o ácido nucleico permanece na célula está dependente do tipo de construções de expressão utilizadas.

Em determinadas formas de realização da invenção, a construção de expressão compreendendo uma ou mais sequências oligonucleotídicas ou polinucleotídicas pode consistir apenas em ADN recombinante ou plasmídeos nus. A transferência da construção pode ser realizada por qualquer dos métodos acima referidos os quais permeabilizam física ou quimicamente a membrana celular. Isto é particularmente aplicável para a transferência *in vitro* mas pode ser também aplicado à utilização *in vivo*. Dubensky et al. (1984) injectou com sucesso ADN de poliomavírus sob a forma de precipitado em fosfato de cálcio no fígado e no baço de murganhos adultos e recém-nascidos apresentando replicação viral activa e infecção aguda. Benvenisty e Reshef (1986) demonstraram também que a injeção intraperitoneal directa de plasmídeos precipitados com fosfato de cálcio resulta na expressão dos genes transfectados. É perspectivado que possa ser também transferido ADN codificando

um gene de interesse de uma forma semelhante *in vivo* e que expresse o produto génico.

Uma outra forma de realização da invenção para transferir uma construção de expressão de ADN nu para células pode envolver bombardeamento de partículas. Este método depende da capacidade de acelerar microprojecteis revestidos por ADN a uma velocidade elevada permitindo que seta perfurem as membranas celulares e se introduzam nas células sem as matar (Klein *et al.*, 1987). Foram desenvolvidos vários dispositivos para acelerar pequenas partículas. Um desses dispositivos baseia-se numa descarga de voltagem elevada para produzir uma corrente eléctrica que, por sua vez, proporciona a força aceleradora (Yang *et al.*, 1990). Os microprojecteis utilizados têm consistido em substâncias biologicamente inertes, tais como esferas de tungsténio ou de ouro.

Foram bombardeados órgãos seleccionados de ratos e murganhos incluindo fígado, pele e tecido muscular *in vivo* (Yang *et al.*, 1990; Zelenin *et al.*, 1991). Isto pode requerer a exposição cirúrgica do tecido ou das células, para eliminar qualquer tecido interferente entre o dispositivo e o órgão alvo, *i. e.*, tratamento *ex vivo*. Novamente, pode ser distribuído ADN codificando um gene particular *via* este método e ainda ser incorporado pela presente invenção.

COMPOSIÇÕES POLIPEPTÍDICAS

Em outros aspectos, a presente invenção proporciona composições polipeptídicas.

Podem ser, geralmente, identificadas porções imunogénicas utilizando técnicas bem conhecidas, tais como as resumidas em Paul, *Fundamental Immunology*, 3^a ed., 243-247 (1993) e referências aí contidas. Essas técnicas incluem o rastreio de polipéptidos com capacidade para reagir com anticorpos específicos para antigénio, antissoros e/ou linhas ou clones de células T. Como aqui utilizado, os antissoros e os anticorpos são “específicos para um antigénio” se estes se ligarem especificamente a um antigénio (*i. e.*, estes reagirem com a proteína num ELISA ou em outro imunoensaio e não reagirem de um modo detectável com proteínas não relacionadas). Esses antissoros e anticorpos podem ser preparados como aqui descrito e utilizando técnicas bem conhecidas. Uma porção imunogénica de uma proteína de *Mycobacterium* sp. é uma porção que reage com esses antissoros e/ou células T num nível que não é substancialmente inferior à reactividade do polipéptido completo (*e. g.*, num ELISA e/ou num ensaio de reactividade de células T). Essas porções imunogénicas podem reagir dentro desses ensaios num nível que é semelhante ou superior à reactividade do polipéptido completo. Esses rastreios podem ser, geralmente, realizados utilizando métodos bem conhecidos do especialista na técnica, tal como os descritos em Harlow e Lane, *Antibodies: A Laboratory Manual* (1988). Por exemplo, um polipéptido pode ser imobilizado num suporte sólido e ser feito contactar com soros de doentes para permitir a ligação de anticorpos dentro dos soros ao polipéptido imobilizado. Os soros não ligados podem ser então removidos e os anticorpos ligados serem detectados utilizando, por exemplo, Proteína A marcada com ¹²⁵I.

Os polipéptidos podem ser preparados utilizando qualquer de várias técnicas bem conhecidas. Os polipéptidos recombinantes codificados por sequências de ADN como descritos acima podem ser

facilmente preparados a partir das sequências de ADN utilizando qualquer de vários vectores de expressão conhecidos do especialista na técnica. A expressão pode ser realizada em qualquer célula hospedeira adequada que tenha sido transformada ou transfectada com um vector de expressão contendo uma molécula de ADN codificando um polipéptido recombinante. As células hospedeiras adequadas incluem procariotas, leveduras e células de eucariotas superiores, tais como células de mamífero e células vegetais. De um modo preferido, as células hospedeiras utilizadas são *E. coli*, leveduras ou uma linha celular de mamífero tal como COS ou CHO. Os sobrenadantes provenientes de sistemas de hospedeiro/vector adequados que segregam a proteína recombinante ou o polipéptido em meios de cultura podem ser inicialmente concentrados utilizando um filtro comercialmente disponível. Após concentração, o concentrado pode ser aplicado a uma matriz de purificação adequada, tais como uma matriz de afinidade ou uma resina de troca iónica. Finalmente, podem ser utilizados um ou mais passos de HPLC em fases reversa para purificar adicionalmente um polipéptido recombinante.

Os polipéptidos da invenção, seus fragmentos imunogénicos e outras variantes tendo menos de cerca de 100 aminoácidos e, geralmente, menos de cerca de 50 aminoácidos, podem ser também produzidos por meios sintéticos, utilizando técnicas bem conhecidas do especialista na técnica. Por exemplo, esses polipéptidos podem ser sintetizados utilizando qualquer das técnicas em fase sólida comercialmente disponíveis, tal como o método de síntese em fase sólida Merrifield, em que os aminoácidos são adicionados sequencialmente a uma cadeia de aminoácidos crescente. Ver Merrifield, *J. Am. Chem. Soc.* **85**:2149-2146 (1963). O equipamento da síntese automatizada de polipéptidos está comercialmente disponível de fornecedores tal

como Perkin Elmer/Applied BioSystems Division (Foster City, CA) e pode ser operado de acordo com as instruções do fabricante.

Dentro de algumas formas de realização específicas, um polipéptido pode ser uma proteína de fusão como descrita aqui ou compreender, pelo menos, uma proteína de fusão como aqui descrita e uma sequência não relacionada.

Por exemplo, um parceiro de fusão pode auxiliar no fornecimento de epitopos T auxiliares (um parceiro de fusão imunológico), de um modo preferido, epitopos T auxiliares reconhecidos por humanos ou pode auxiliar na expressão da proteína (um intensificador da expressão) com maiores rendimentos do que a proteína recombinante nativa. Alguns parceiros de fusão preferidos são parceiros de fusão imunológicos e intensificadores da expressão. Podem ser seleccionados outros parceiros de fusão para aumentar a solubilidade da proteína ou permitir à proteína ser direccionada para compartimentos intracelulares pretendidos. Os parceiros de fusão adicionais incluem ainda marcadores de afinidade, que facilitam a purificação da proteína.

As proteínas de fusão podem ser, geralmente, preparadas utilizando técnicas padrão, incluindo conjugação química. De um modo preferido, uma proteína de fusão é expressa como uma proteína recombinante, permitindo a produção de níveis aumentados, em relação a uma proteína não-fundida, num sistema de expressão. Resumidamente, as sequências de ADN codificando os componentes polipeptídicos, podem ser montadas separadamente e ser ligadas num vector de expressão adequado. A extremidade 3' da sequência de ADN codificando um componente polipeptídico é ligada, com ou sem um elemento de ligação peptídico, à

extremidade 5' de uma sequência de ADN codificando o segundo componente polipeptídico para que as grelhas de leitura das sequências estejam em fase. Isto permite a tradução numa única proteína de fusão que conserva a actividade biológica de ambos os polipéptidos componentes.

Pode ser utilizada uma sequência de um elemento de ligação peptídico para separar os primeiros e segundos componentes polipeptídicos por uma distância suficiente para assegurar que cada polipéptido se enrola nas suas estruturas secundárias e terciárias. Essa sequência de um elemento de ligação peptídico é incorporada na proteína de fusão utilizando técnicas padrão bem conhecidas no campo técnico. Podem ser seleccionadas sequências adequadas de elementos de ligação peptídicos com base nos factores seguintes: (1) a sua capacidade para adoptar uma conformação estendida flexível; (2) a sua incapacidade para adoptar uma estrutura secundária que possa interagir com epitopos funcionais nos primeiros e segundos polipéptidos; e (3) a ausência de resíduos hidrofóbicos ou carregados que possam reagir com os epitopos funcionais do polipéptido. As sequências preferidas do elemento de ligação peptídico contêm resíduos de Gly, Asn e Ser. Podem ser também utilizados outros aminoácidos quase neutros, tais como Thr e Ala na sequência do elemento de ligação. As sequências de aminoácidos que podem ser utilizadas de um modo útil como elementos de ligação incluem as divulgadas 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°. 4935233 e Patente U.S. N° 4751180. A sequência do elemento de ligação pode ter, geralmente, de 1 a cerca de 50 aminoácidos de comprimento. As sequências do elemento de ligação não são necessárias quando os primeiros e segundos polipéptidos têm regiões de aminoácidos não essenciais no terminal N que podem

ser utilizadas para separar os domínios funcionais e prevenir a interferência estérica.

As sequências de ADN ligadas estão operativamente ligadas a elementos reguladores da transcrição ou da tradução adequados. Os elementos reguladores, responsáveis pela expressão do ADN, estão apenas localizados em 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 em 3' da sequência de ADN codificando o segundo polipéptido.

São também proporcionadas proteínas de fusão. Essas proteínas compreendem um polipéptido como aqui descrito em conjunto com uma proteína imunogénica não relacionada. De um modo preferido, a proteína imunogénica é capaz de promover uma resposta de memória. Os exemplos dessas proteínas incluem proteínas de tétano, tuberculose e hepatite (ver e. g., Stoute *et al.*, *New Engl. J. Med.* 336:86-91 (1997)).

Dentro das formas de realização preferidas, um parceiro de fusão imunológico é derivado da proteína D, uma proteína da superfície da bactéria gram negativa *Haemophilus influenza B* (documento WO 91/18926). De um modo preferido, um derivado da proteína D compreende aproximadamente o primeiro terço da proteína (e. g., os primeiros 100-110 aminoácidos do terminal N) e um derivado da proteína D pode ser lipidado. Dentro de algumas formas de realização preferidas, são incluídos no terminal N os 109 primeiros resíduos de um parceiro de fusão lipoproteína D para proporcionar ao polipéptido epitopos de células T exógenos adicionais e aumentar o nível de expressão em *E. coli* (funcionando deste modo como um intensificador da expressão). A

cauda lipídica assegura a apresentação óptima do antígeno às células apresentadoras do antígeno. Outros parceiros de fusão incluem a proteína não-estrutural do vírus influenza, NS1 (hemaglutinina). Tipicamente, são utilizados os 81 aminoácidos do terminal N, embora possam ser utilizados fragmentos diferentes que incluam epitopos T auxiliares.

Numa outra forma de realização, o parceiro de fusão imunológico é a proteína conhecida como LYTA ou uma sua porção (de um modo preferido, uma porção do terminal C). A LYTA é derivada de *Streptococcus pneumoniae*, o qual sintetiza uma amidase de N-acetil-L-alanina conhecida como amidase LYTA (codificada pelo gene *Lyta*; *Gene* **43**:265-292 (1986)). A LYTA é uma autolisina que degrada especificamente determinadas ligações na estrutura base do peptidoglicano. O domínio do terminal C da proteína LYTA é responsável pela afinidade com a colina ou com alguns análogos da colina, tal como DEAE. Esta propriedade foi explorada para o desenvolvimento de plasmídeos expressando C-LYTA em *E. coli* úteis para a expressão de proteínas de fusão. Foi descrita a purificação de proteínas híbridas contendo o fragmento C-LYTA no terminal amino (ver *Biotechnology* **10**:795-798 (1992)). Dentro de uma forma de realização preferida, pode ser incorporada uma porção de repetição de LYTA numa proteína de fusão. Na região do terminal C que inicia no resíduo 178 encontra-se uma porção de repetição. Uma porção de repetição particularmente preferida incorpora os resíduos 188-305.

Em geral, são isolados polipéptidos (incluindo as proteínas de fusão) e polinucleótidos como aqui descritos. Um polipéptido ou polinucleótido "isolado" é aquele que é removido do seu ambiente original. Por exemplo, uma proteína de ocorrência natural está isolada se for separada de parte ou da totalidade

dos materiais coexistentes no sistema natural. De um modo preferido, esses polipéptidos estão, pelo menos, cerca de 90% puros, de um modo mais preferido, pelo menos, cerca de 95% puros e de um modo muito preferido, pelo menos, cerca de 99% puros. Considera-se que um polinucleótido está isolado se, por exemplo, for clonado num vector que não é uma parte do ambiente natural.

CÉLULAS T

As células T podem ser isoladas a partir de medula óssea, sangue periférico ou uma fracção de medula óssea ou sangue periférico de um doente, utilizando um sistema de separação celular comercialmente disponível, tal como o Isolex™ System, disponível de Nexell Therapeutics, Inc. (Irvine, CA; *ver também* Patente U.S. N° 5240856; Patente U.S. N° 5215926; documento WO 89/06280; documento WO 91/16116 e documento WO 92/07243). Alternativamente, as células T podem ser derivadas de linhas ou culturas celulares humanas, de mamífero não-humano, relacionadas ou não relacionadas.

As células T podem ser estimuladas com um polipéptido da invenção, um polinucleótido codificando esse polipéptido e/ou uma célula apresentadora de antigénio (APC) que expressa esse polipéptido. Esse estímulo é realizado sob condições e durante um tempo suficiente para permitir a produção de células T que sejam específicas para o polipéptido. De um modo preferido, o polipéptido ou o polinucleótido está presente dentro de um veículo de transferência, tal como uma microesfera, para facilitar a produção de células T específicas.

As células T são consideradas específicas para um polipéptido da invenção, se as células T especificamente proliferarem, segregarem citocinas ou matarem células alvo revestidas com o polipéptido ou expressando um gene codificando o polipéptido. A especificidade das células T pode ser avaliada utilizando qualquer de várias técnicas padrão. Por exemplo, num ensaio de libertação de crómio ou num ensaio de proliferação, um índice de estímulo superior a um aumento de duas vezes na lise e/ou na proliferação, em comparação com controlos negativos, indica especificidade das célula T. Esses ensaios podem ser realizados, por exemplo, como descrito em Chen *et al.*, *Cancer Res.* **54**:1065-1070 (1994)). Alternativamente, a detecção da proliferação de células T pode ser realizada através de várias técnicas conhecidas. Por exemplo, a proliferação de células T pode ser detectada pela medição de uma taxa aumentada de síntese de ADN (e. g., por marcador de culturas de células T com pulsos de timidina tritiada e medição da quantidade de timidina tritiada incorporada no ADN). O contacto com um polipéptido da invenção (100 ng/mL - 100 µg/mL, de um modo preferido, 200 ng/mL - 25 µg/mL) durante 3-7 dias deve resultar em, pelo menos, um aumento de duas vezes na proliferação das células T. O contacto como descrito acima durante 2-3 horas deve resultar na activação das células T, como medida utilizando ensaios de citocina padrão em que um aumento de duas vezes no nível de libertação de citocina (e. g., TNF ou IFN-γ) é indicativa de activação das células T (ver Coligan *et al.*, *Current Protocols in Immunology*, vol. 1 (1998)). As células T que foram activadas em resposta a um polipéptido, um polinucleótido ou a uma APC expressando um polipéptido podem ser CD4⁺ e/ou CD8⁺. As células T específicas para uma proteína podem ser expandidas utilizando técnicas padrão.

As células T CD4⁺ ou CD8⁺ que proliferam em resposta a um polipéptido, um polinucleótido ou uma APC podem ser expandidas em número *in vitro* ou *in vivo*. A proliferação dessas células T *in vitro* pode ser realizada de vários modos. Por exemplo, as células T podem ser re-expostas a um polipéptido ou a um péptido curto, correspondente a uma porção imunogénica desse polipéptido, com ou sem a adição de factores de crescimento de células T, tais como interleucina 2 e/ou células estimuladoras que sintetizam o polipéptido. Alternativamente, podem ser expandidas em número uma ou mais células T que proliferam na presença da proteína por clonagem. Os métodos de clonagem de células são bem conhecidos na técnica e incluem diluição limitante.

COMPOSIÇÕES FARMACÊUTICAS

Em formas de realização adicionais, a presente invenção refere-se à formulação de um ou mais polinucleótidos ou polipéptidos aqui divulgados em soluções farmacêuticamente aceitáveis ou fisiologicamente aceitáveis para administração a uma célula ou a um animal, isoladamente ou em combinação com uma ou mais de outras modalidades de terapia. Essas composições são também úteis para utilizações em diagnóstico.

Será também entendido que, se pretendido, podem ser também administradas composições de segmento de ácido nucleico, ARN, ADN ou APN que expressam um polipéptido como divulgado aqui em combinação com outros agentes, tais como, *e. g.*, outras proteínas ou polipéptidos ou vários agentes farmacêuticamente activos. De facto, não existe praticamente qualquer limite para outros componentes que possam ser também incluídos, desde que os

agentes adicionais não causem um efeito adverso significativo após o contacto com as células alvo ou com os tecidos hospedeiros. As composições podem ser, deste modo, transferidas juntamente com vários outros agentes, de acordo com a necessidade de cada caso particular. Essas composições podem ser purificadas a partir de células hospedeiras ou de outras fontes biológicas ou, alternativamente, serem sintetizadas quimicamente como aqui descrito. Do mesmo modo, essas composições podem ainda compreender ARN substituído ou derivatizado ou composições de ADN.

A formulação de excipientes e soluções veículo farmacêuticamente aceitáveis é bem conhecida do especialista na técnica, assim como é o desenvolvimento de regimes de tratamento e dosagem adequados para utilização nas composições particulares aqui descritas em vários regimes de tratamento, incluindo e. g., administração e formulação oral, parentérica, intravenosa, intranasal e intramuscular.

1. DISTRIBUIÇÃO ORAL

Em determinadas aplicações, as composições farmacêuticas aqui divulgadas podem ser distribuídas a um animal via administração oral. Como tal, estas composições podem ser formuladas com um diluente inerte ou com um veículo comestível assimilável ou estas podem ser contidas em cápsulas de gelatina de revestimento duro ou mole ou estas podem ser prensadas em comprimidos ou estas podem ser incorporadas directamente com o alimento da dieta.

Os compostos activos podem ser mesmo incorporados com excipientes e serem utilizados sob a forma de comprimidos ingeríveis, comprimidos bucais, pastilhas, cápsulas, elixires, suspensões, xaropes, hóstias e semelhantes (Mathiowitz *et al.*, 1997; Hwang *et al.*, 1998; Patente U.S. 5641515; Patente U.S. 5580579 e Patente U.S. 5792451.

Os comprimidos, pastilhas, pílulas, cápsulas e semelhantes podem também conter o seguinte: um ligante, como goma de tragacanto, acácia, amido de milho ou gelatina; excipientes, tal como fosfato dicálcico; um agente desintegrante, tais como amido de milho, amido de batata, ácido algínico e semelhantes; pode ser adicionado um lubrificante, tal como estearato de magnésio; e um agente adoçante, tal como sacarose, lactose ou sacarina ou um agente aromatizante, tal como aroma de hortelã-pimenta, óleo de gaultéria ou cereja. Quando a forma de unidade de dosagem é uma cápsula; esta pode conter, além de materiais do tipo acima, um veículo líquido. Podem estar presentes vários outros materiais como revestimentos ou para de qualquer outro modo modificar a forma física da unidade de dosagem. Por exemplo, os comprimidos, as pílulas ou as cápsulas podem ser revestidas com goma-laca, açúcar ou com ambos. Um xarope de elixir pode conter o composto activo sacarose como agente adoçante, metil e propilparabenos como conservantes, um corante e aromatizante, tais como aroma de cereja ou laranja. Obviamente, qualquer material utilizado na preparação de qualquer forma de unidade de dosagem deve estar farmacêuticamente puro e ser substancialmente não-tóxico nas quantidades utilizadas. Além disso, os compostos activos podem ser incorporados em preparações e formulações de libertação prolongada.

Tipicamente, estas formulações podem conter, pelo menos, cerca de 0,1% do composto activo ou mais, embora a percentagem do(s) ingrediente(s) activo(s) possa ser, obviamente, variada e pode situar-se convenientemente entre cerca de 1 ou 2% e cerca de 60% ou 70% ou mais do peso ou volume da formulação total. Obviamente, a quantidade do(s) composto(s) activo(s) em cada composição terapeuticamente útil pode ser preparada de forma a ser obtida uma dosagem adequada em qualquer dose unitária determinada do composto. Factores, tais como solubilidade, a biodisponibilidade, meia-vida biológica, via de administração, tempo de armazenamento de produto, assim como outras considerações farmacológicas irão ser contemplados por um especialista na técnica da preparação dessas formulações farmacêuticas e como tal, podem ser pretendidas várias dosagens e regimes de tratamento.

Para a administração oral as composições da presente invenção podem ser alternativamente incorporadas com um ou mais excipientes na forma de elixir bucal, dentífrico, comprimido bucal, vaporização oral ou formulação sublingual administrada oralmente. Por exemplo, pode ser preparado um elixir bucal incorporando o ingrediente activo na quantidade necessária num solvente adequado, tal como uma solução de borato de sódio (Solução de Dobell). Alternativamente, o ingrediente activo pode ser incorporado numa solução oral, tais como uma contendo borato de sódio, glicerina e bicarbonato de potássio ou ser disperso num dentífrico ou ser adicionado numa quantidade terapeuticamente eficaz a uma composição que pode incluir água, ligantes, abrasivos, agentes aromatizantes, agentes espumantes e humidificadores. Alternativamente as composições podem ser preparadas numa forma em comprimido ou em solução que pode ser colocada sob a língua ou de outra forma dissolvida na boca.

2. DISTRIBUIÇÃO INJECTÁVEL

Em determinadas circunstâncias será desejável transferir parentericamente, intravenosamente, intramuscularmente ou mesmo intraperitonealmente as composições farmacêuticas aqui divulgadas como descrito na Patente U.S. 5543158; Patente U.S. 5641515 e Patente U.S. 5399363.

As soluções dos compostos activos como bases livres ou sais farmacologicamente aceitáveis podem ser preparadas em água adequadamente misturada com um tensioactivo, tal como hidroxipropilcelulose. Podem ser também preparadas dispersões em glicerol, polietilenoglicóis líquidos e suas misturas e em óleos. Sob condições normais de armazenamento e utilização, estas preparações contêm um conservante para prevenir o crescimento de microrganismos.

As formas farmacêuticas adequadas para a utilização injectável incluem soluções aquosas estéreis ou dispersões e pó estéril para a preparação extemporânea de soluções injectáveis ou dispersões estéreis (Patente U.S. 5466468).

Em todos os casos, a forma deve ser estéril e deve ser fluida na medida de uma utilização fácil em seringa. Deve ser estável nas condições de preparação e armazenamento e deve ser conservada contra a acção contaminante de microrganismos, tais como bactérias e fungos. O veículo pode ser um solvente ou meio de dispersão contendo, por exemplo, água, etanol, poliol (e. g., glicerol, propilenoglicol e polietilenoglicol líquido e semelhantes), suas misturas adequadas e/ou óleos vegetais. A fluidez adequada pode ser mantida, por exemplo, pela utilização de um revestimento, tal como lecitina, pela manutenção do

tamanho de partícula necessário, no caso da dispersão e pela utilização de tensioactivos. A prevenção da acção de microrganismos pode ser facilitada por vários agentes antibacterianos e antifúngicos, por exemplo, parabenos, clorobutanol, fenol, ácido sórbico, timerosal e semelhantes. Em muitos casos, será preferível, incluir agentes isotónicos, por exemplo, açúcar ou cloreto de sódio. A absorção prolongada das composições injectáveis pode ser proporcionada pela utilização de agentes de atraso da absorção, por exemplo, monoestearato de alumínio e gelatina nas composições.

Por exemplo, para a administração parentérica numa solução aquosa, a solução deve ser tamponada adequadamente se necessário e o diluente líquido ser previamente tornado isotónico com soro fisiológico ou glicose suficiente. Estas soluções aquosas particulares são especialmente adequadas para a administração intravenosa, intramuscular, subcutânea e intraperitoneal. No que a isto se refere, um meio aquoso estéril que possa ser utilizado será conhecido do especialista na técnica de acordo com a presente divulgação. Por exemplo, uma dosagem pode ser dissolvida em 1 mL de solução isotónica de NaCl e ser quer adicionada a 1000 mL de fluido de hipodermoclise quer injectada no local de infusão proposto (ver e. g., *Remington's Pharmaceutical Sciences*, 15ª Edição, pp 1035-1038 e 1570-1580). Irá necessariamente ocorrer alguma variação na dosagem dependendo do estado do indivíduo a ser tratado. A pessoa responsável pela administração irá determinar, em qualquer situação, a dose adequada para o indivíduo individual. Além disso, para a administração humana, as preparações devem cumprir os padrões de esterilidade, pirogenicidade e segurança geral e pureza como requerido pelo FDA Office of Biologics standards.

São preparadas soluções injectáveis estéreis incorporando os compostos activos na quantidade necessária no solvente adequado com vários dos outros ingredientes listados acima, como necessário, seguido de esterilização por filtração. Geralmente, as dispersões são preparadas incorporando vários ingredientes activos esterilizados num veículo estéril, o qual contém o meio de dispersão básico e outros ingredientes necessários dentro dos listados acima. No caso de pós estéreis para a preparação de soluções injectáveis estéreis, os métodos de preparação preferidos consistem em técnicas de secagem sob vácuo e liofilização que produzem um pó do ingrediente activo adicionado de qualquer ingrediente pretendido adicional a partir de uma sua solução previamente esterilizada por filtração.

As composições aqui divulgadas podem ser formuladas sob uma forma neutra ou sob a forma de sal. Os sais farmacêuticamente aceitáveis, incluem os sais de adição ácida (formados com os grupos amina livres da proteína) e que são formados com ácidos inorgânicos tais como, por exemplo, ácidos clorídricos ou fosfóricos ou ácidos orgânicos, tais como acético, oxálico, tartárico, mandélico e semelhantes. Os sais formados com os grupos carboxilo livres podem ser também derivados de bases inorgânicas tais como, por exemplo, sódio, potássio, amónio, cálcio ou hidróxidos férricos e bases orgânicas, tais como isopropilamina, trimetilamina, histidina, procaína e semelhantes. Após formulação, as soluções serão administradas de uma forma compatível com a formulação de dosagem e numa quantidade que seja terapêuticamente eficaz. As formulações são facilmente administradas em várias formas de dosagem, tais como soluções injectáveis, cápsulas de libertação de fármacos e semelhantes.

Como aqui utilizado, "veículo" inclui todos e quaisquer solventes, meios de dispersão, veículos, revestimentos, diluentes, agentes antibacterianos e antifúngicos, agentes isotónicos e de retardamento da absorção, tampões, soluções veículo, suspensões, colóides e semelhantes. A utilização desses meios e agentes com substâncias activas farmacêuticas é bem conhecida na técnica. Excepto no caso em que algum meio ou agente convencional seja incompatível com o ingrediente activo, a sua utilização nas composições terapêuticas é contemplada. Os ingredientes activos suplementares também podem ser incorporados nas composições.

A frase "farmaceuticamente aceitável" refere-se a entidades moleculares e composições que não produzem uma reacção alérgica ou semelhante adversa quando administradas a um humano. A preparação de uma composição aquosa que contém uma proteína como um ingrediente activo é bem entendida na técnica. Tipicamente, essas composições são preparadas como injectáveis, como soluções líquidas ou como suspensões; também podem ser preparadas formas sólidas adequadas para solução ou suspensão em líquido antes da injeção. A preparação também pode ser emulsionada.

3. TRANSFERÊNCIA NASAL

Em determinadas formas de realização, as composições farmacêuticas podem ser transferidas por vaporizações intranasais, inalação e/ou outros veículos de transferência de aerossóis. Foram descritos métodos para transferência de composições de genes, ácidos nucleicos e péptidos directamente aos pulmões via sprays nasais de aerossol, e. g., Patente U. S. 5756353 e Patente U.S. 5804212. Do mesmo modo, é

também bem conhecida nas técnicas farmacêuticas a transferência de fármacos utilizando resinas de micropartículas intranasais (Takenaga *et al.*, 1998) e compostos de lisofosfatidil-glicerol (Patente U.S. 5725871). Do mesmo modo, na Patente U. S. 5780045 é descrita a transferência transmucosal de fármacos sob a forma de uma matriz de suporte de politetrafluoroetileno.

4. TRANSFERÊNCIA MEDIADA POR LIPOSSOMAS, NANOCÁPSULAS E MICROPARTÍCULAS

Em determinadas formas de realização, os requerentes contemplam a utilização de lipossomas, nanocápsulas, micropartículas, microesferas, partículas lipídicas, vesículas e semelhantes, para a introdução das composições da presente invenção em células hospedeiras adequadas. Em particular as composições da presente invenção podem ser formuladas para a transferência encapsuladas numa partícula lipídica, num lipossoma, numa vesícula, numa nanosfera ou numa nanopartícula ou semelhantes.

Essas formulações podem ser preferidas para a introdução de formulações farmaceuticamente aceitáveis dos ácidos nucleicos ou construções aqui divulgados. A formação e a utilização de lipossomas são, geralmente, conhecidas do especialista na técnica (ver por exemplo, Couvreur *et al.*, 1977; Couvreur, 1988; Lasic, 1998; que descrevem a utilização de lipossomas e nanocápsulas em terapia com antibiótico direccionada para infecções e doenças bacterianas intracelulares). Recentemente, foram desenvolvidos lipossomas com estabilidade e tempos de meia vida em circulação no soro melhorados (Gabizon e

Papahadjopoulos, 1988; Allen e Choun, 1987; Patente U. S. 5741516).

Adicionalmente, têm sido revistos vários métodos de preparação de lipossomas e semelhantes a lipossomas como potenciais veículos de fármacos (Takakura, 1998; Chandran *et al.*, 1997; Margalit, 1995; Patente U. S. 5567434; Patente U. S. 5552157; Patente U. S. 5565213; Patente U. S. 5738868 e Patente U.S. 5795587).

Têm sido utilizados com sucesso lipossomas com vários tipos celulares que são normalmente resistentes à transfecção por outros processos incluindo suspensões de célula T, culturas primárias de hepatócitos e células PC 12 (Renneisen *et al.*, 1990; Muller *et al.*, 1990). Além disso, os lipossomas estão isentos do constrangimentos de comprimento do ADN que é típico dos sistemas de transferência virais. Os lipossomas têm sido utilizados eficazmente para introduzir genes, fármacos (Heath e Martin, 1986; Heath *et al.*, 1986; Balazsovits *et al.*, 1989; Fresta e Puglisi, 1996), agentes radioterapêuticos (Pikul *et al.*, 1987), enzimas (Imaizumi *et al.*, 1990a; Imaizumi *et al.*, 1990b), vírus (Faller e Baltimore, 1984), factores de transcrição e efectores alostéricos (Nicolau e Gersonde, 1979) em várias linhas celulares cultivadas e em animais. Além disso, foram concluídos vários, com sucesso, ensaios clínicos examinando a eficácia da transferência de fármacos mediada por lipossomas (Lopez-Berestein *et al.*, 1985a; 1985b; Coune, 1988; Sculier *et al.*, 1988). Além disso, vários estudos sugerem que a utilização de lipossomas não está associada a respostas autoimunitárias, toxicidade ou localização gonadal após transferência sistémica (Mori e Fukatsu, 1992).

Os lipossomas são formados a partir de fosfolípidos que são dispersos num meio aquoso e formam espontaneamente vesículas em bicamada concêntricas multilamelares (também designadas vesículas multilamelares (MLV)). As MLV têm geralmente diâmetros de 25 nm a 4 μ m. A sonicação das MLV resulta na formação de pequenas vesículas unilamelares (SUV) com diâmetros na gama de 200 a 500 Å, contendo uma solução aquosa no núcleo.

Os lipossomas apresentam semelhança com as membranas celulares e são contemplados para utilização associada à presente invenção como veículos para composições peptídicas. Estes são amplamente adequados dado que podem ser aprisionadas substâncias solúveis em água e lipídicas, *i. e.*, respectivamente, nos espaços aquosos e no interior da própria bicamada. É possível que os lipossomas contendo fármacos possam ser mesmo utilizados na transferência específica para um local de agentes activos por modificação selectiva da formulação lipossómica.

Além dos ensinamentos de Couvreur *et al.* (1977; 1988), pode ser utilizada a informação seguinte na produção de formulações lipossómicas. Os fosfolípidos podem formar várias estruturas além de lipossomas quando dispersos na água, dependendo da proporção molar entre o lípido e à água. Em proporções baixas o lipossoma é a estrutura preferida. As características físicas dos lipossomas dependem do pH, da força iónica e da presença de catiões divalentes. Os lipossomas podem apresentar permeabilidade baixa a substâncias iónicas e polares, mas sofrer uma transição de fase a temperaturas elevadas que altera significativamente a sua permeabilidade. A transição de fase envolve uma alteração de uma estrutura fortemente empacotada, ordenada, conhecida como o estado de gel, para uma estrutura

pouco empacotada, menos ordenada, conhecida como o estado de fluido. Isto ocorre a uma temperatura de transição de fase característica e resulta num aumento da permeabilidade a iões, açúcares e fármacos.

Além da temperatura, a exposição a proteínas pode alterar a permeabilidade dos lipossomas. Determinadas proteínas solúveis, tal como o citocromo c, ligam-se, deformam e penetram a bicamada, causando deste modo modificações na permeabilidade. O colesterol inibe esta penetração de proteínas, aparentemente empacotando os fosfolípidos mais fortemente. É contemplado que as formações de lipossomas mais úteis para a transferência de antibióticos e inibidores irão conter colesterol.

A capacidade para aprisionar solutos varia entre tipos diferentes de lipossomas. Por exemplo, as MLV são moderadamente eficientes no aprisionamento de solutos, mas as SUV são extremamente ineficientes. No entanto as SUV oferecem a vantagem de homogeneidade e da reproduzibilidade na transferência de tamanho e é oferecido um compromisso entre tamanho e eficiência de aprisionamento por grandes vesículas unilamelares (LUV). Estas são preparadas por evaporação de éter e são três a quatro vezes mais eficientes no aprisionamento de solutos do que as MLV.

Além das características do lipossoma, um determinante importante no aprisionamento de compostos consiste nas propriedades fisico-químicas do próprio composto. Os compostos polares são aprisionados nos espaços aquosos e os compostos não-polares ligam-se à bicamada lipídica da vesícula. Os compostos polares são libertados através de permeação ou quando a bicamada é interrompida, mas os compostos não-polares

permanecem associados à bicamada a menos que esta seja interrompida por temperatura ou exposição a lipoproteínas. Ambos os tipos apresentam taxas máximas de efluxo à temperatura de transição de fase.

Os lipossomas interagem com células via quatro mecanismos diferentes: endocitose por células fagocíticas do sistema reticuloendotelial, tais como macrófagos e neutrófilos; adsorção à superfície celular, por forças não-específicas fracas hidrofóbicas ou electrostáticas ou por interações específicas com componentes da superfície celular; fusão com a membrana plasmática celular por inserção da bicamada lipídica do lipossoma na membrana plasmática, com a libertação simultânea do conteúdo lipossómico no citoplasm; e por transferência dos lípidos lipossómicos para membranas celulares ou subcelulares ou vice-versa, sem qualquer associação dos conteúdos do lipossoma. Frequentemente é difícil determinar qual o mecanismo que é funcional e pode funcionar mais de um simultaneamente.

O destino e a disposição dos lipossomas injectados intravenosamente dependem das suas propriedades físicas, tais como tamanho, fluidez e carga superficial. Estes podem persistir em tecidos durante h ou dias, dependendo da sua composição e as meias vidas no sangue variam desde min e várias h. Os lipossomas maiores, tais como MLV e LUV, são incorporados rapidamente pelas células fagocíticas do sistema reticuloendotelial, mas a fisiologia do sistema circulatório restringe a saída dessas espécies grandes na maioria das localizações. Estas podem sair apenas em localizações em que existam grandes aberturas ou poros no endotélio capilar, tais como os sinusóides do fígado ou do baço. Deste modo, estes órgãos são o local predominante da incorporação. Por outro lado, as SUV apresentam uma maior

transferência tecidual mas são ainda altamente sequestradas no fígado e no baço. Em geral, este comportamento *in vivo* limita o direccionamento potencial dos lipossomas apenas para os órgãos e tecidos acessíveis ao seu grande tamanho. Estes incluem sangue, fígado, baço, medula óssea e órgãos linfóides.

O direccionamento não é, geralmente, uma limitação para a presente invenção. No entanto, se for pretendido um direccionamento específico, estão disponíveis métodos para este ser realizado. Podem ser utilizados anticorpos para ligação à superfície do lipossoma e direccionar o anticorpo e os seus conteúdos de fármaco para receptores antigénicos específicos localizados numa superfície de um tipo celular particular. Podem ser também utilizados determinantes hidratos de carbono (componentes glicoproteicos ou glicolipídicos da superfície celular que desempenham um papel no reconhecimento, interacção e adesão intercelular) como locais de reconhecimento uma vez que estes têm potencial no direccionamento de lipossomas para tipos celulares particulares. Principalmente, é contemplado que seja utilizada a injeção intravenosa de preparações lipossómicas, mas são também concebíveis outras vias de administração.

Alternativamente, a invenção proporciona formulações de nanocápsulas farmacologicamente aceitáveis das composições do presente invenção. As nanocápsulas podem, geralmente, aprisionar compostos de um modo estável e reprodutível (Henry-Michelland *et al.*, 1987; *et al.* Quintanar-Guerrero., 1998; Douglas *et al.*, 1987). Essas partículas ultrafinas (com um tamanho de aproximadamente 0,1 μm) deve ser concebidas utilizando polímeros capazes de serem degradados *in vivo* para evitar efeitos secundários devidos ao sobrecarregamento polimérico intracelular. São contemplados para utilização na

presente invenção nanopartículas biodegradáveis de polialquilcianoacrilato que cumprem estes requisitos. Essas partículas podem ser facilmente preparadas como descrito (Couvreur *et al.*, 1980; 1988; zur Muhlen *et al.*, 1998; Zambaux *et al.* 1998; Pinto-Alphandry *et al.*, 1995 e Patente U.S. 5145684).

VACINAS

São proporcionadas vacinas em algumas formas de realização preferidas da presente invenção. As vacinas irão compreender, geralmente, uma ou mais composições farmacêuticas, tais como as discutidas acima, em combinação com um imunoestimulante. Um imunoestimulante pode ser qualquer substância que intensifique ou potencie uma resposta imunitária (de anticorpo e/ou mediada por célula) a um antígeno exógeno. Os exemplos de imunoestimulantes incluem adjuvantes, microesferas biodegradáveis (e. g., galactido poliláctico) e lipossomas (no quais o composto é incorporado; ver e. g., Fullerton, Patente U.S. N° 4235877). A preparação de vacinas está de um modo geral descrita, por exemplo, em Powell e Newman, ed., *Vaccine Design (the subunit and adjuvant approach)* (1995). As composições farmacêuticas e as vacinas dentro do âmbito da presente invenção podem também conter outros compostos, o quais podem ser biologicamente activos ou inactivos. Por exemplo, podem estar presentes uma ou mais porções imunogénicas de outros antígenos tumorais, incorporados num polipéptido de fusão ou como um composto separado, dentro da composição ou vacina.

As vacinas ilustrativas podem conter ADN codificando um ou mais dos polipéptidos como descritos acima, de forma a que o

polipéptido seja produzido *in situ*. Como observado acima, o ADN pode estar presente dentro de qualquer de vários sistemas de transferência conhecidos do especialista na técnica, incluindo sistemas de expressão de ácidos nucleicos, bactérias e sistemas de expressão virais. São bem conhecidas na técnica numerosas técnicas de transferência génica, tais como as descritas por Rolland, *Crit. Rev. Therap. Drug Carrier Systems* **15**:143-198 (1998) e referências aí contidas. Os sistemas de expressão de ácidos nucleicos adequados contêm as sequências de ADN necessárias para a expressão no doente (tais como um promotor e um sinal de terminação adequados). Os sistemas de transferência bacterianos envolvem a administração de uma bactéria (tal como *Bacillus-Calmette-Guerrin*) que expressa uma porção imunogénica do polipéptido na sua superfície celular ou segrega esse epitopo. Numa forma de realização preferida, o ADN pode ser introduzido utilizando um sistema de expressão viral (e. g., vacinia ou outro poxvírus, retrovírus ou adenovírus), o que pode envolver a utilização de um vírus não-patogénico deficiente, competente para a replicação. Os sistemas adequados são divulgados, por exemplo, em Fisher-Hoch *et al.*, *Proc. Natl. Acad. Sci. USA* **86**:317-321 (1989); Flexner *et al.*, *Ann. N. Y. Acad. Sci.* **569**:86-103 (1989); Flexner *et al.*, *Vaccine* **8**:17-21 (1990); Patentes U.S. N° 4603112, 4769330 e 5017487; documento WO 89/01973; Patente U.S. N° 4777127; documento GB 2200651; documento EP 0345242; documento WO 91/02805; Berkner, *Biotechniques* **6**:616-627 (1988); Rosenfeld *et al.*, *Science* **252**:431-434 (1991); Kolls *et al.*, *Proc. Natl. Acad. Sci. USA* **91**:215-219 (1994); Kass-Eisler *et al.*, *Proc. Natl. Acad. Sci. USA* **90**:11498-11502 (1993); Guzman *et al.*, *Circulation* **88**:2838-2848 (1993); e Guzman *et al.*, *Cir. Res.* **73**:1202-1207 (1993). As técnicas para incorporar ADN nesses sistemas de expressão são bem conhecidas do especialista na técnica. O ADN pode ser também

“nu”, como descrito, por exemplo, em Ulmer *et al.*, *Science* **259**:1745-1749 (1993) e revisto por Cohen, *Science* **259**:1691-1692 (1993). A incorporação de ADN nu pode ser aumentada revestindo o ADN em esferas biodegradáveis, as quais são eficientemente transportadas para o interior das células. Será evidente que uma vacina pode compreender uma componente polinucleotídica e polipeptídica. Essas vacinas podem prover uma resposta imunitária intensificada.

Será evidente que uma vacina pode conter sais farmacologicamente aceitáveis dos polinucleótidos e dos polipéptidos aqui proporcionados. Esses sais podem ser preparados a partir de bases não-tóxicas farmacologicamente aceitáveis, incluindo bases orgânicas (e. g., sais de aminas primárias, secundárias e terciárias e aminoácidos básicos) e bases inorgânicas (e. g., sais de sódio, potássio, lítio, amônio, cálcio e magnésio).

Embora possa ser utilizado qualquer veículo adequado conhecido do especialista na técnica, nas composições de vacina desta invenção, o tipo do veículo irá variar dependendo do modo de administração. As composições da presente invenção podem ser formuladas para qualquer forma adequada de administração, incluindo, por exemplo, administração tópica, oral, nasal, intravenosa, intracraniana, intraperitoneal, subcutânea ou intramuscular. Para administração parentérica, tal como injeção subcutânea, o veículo compreende, de um modo preferido, água, soro fisiológico, álcool, uma gordura, uma cera ou um tampão. Para a administração oral, pode ser utilizado qualquer dos veículos acima ou um veículo sólido, tais como manitol, lactose, amido, estearato de magnésio, sacarina de sódio, talco, celulose, glicose, sacarose e carbonato de magnésio. Podem ser

também utilizadas microesferas biodegradáveis (e. g., poliglicolato de polilactato) como veículos das composições farmacêuticas desta invenção. As microesferas biodegradáveis adequadas são divulgadas, por exemplo, nas Patentes U.S. N° 4897268; 5075109; 5928647; 5811128; 5820883; 5853763; 5814344 e 5942252. Pode-se também utilizar um veículo compreendendo os complexos partícula-proteína descritos na Patente U.S. N° 5928647, os quais são capazes de induzir uma resposta de linfócito T citotóxico restrita à classe I num hospedeiro.

Essas composições também podem compreender tampões (e. g., soro fisiológico tamponado neutro ou soro fisiológico tamponado com fosfato), hidratos de carbono (e. g., glicose, manose, sacarose ou dextranos), manitol, proteínas, polipéptidos ou aminoácidos, tal como glicina, antioxidantes, bacteriostáticos, agentes quelantes, tais como EDTA ou glutathione, adjuvantes (e. g., hidróxido de alumínio), solutos que tornam a formulação isotônica, hipotônica ou fracamente hipertônica com o sangue de um receptor, agentes de suspensão, agentes espessantes e/ou conservantes. Alternativamente, as composições da presente invenção podem ser formuladas sob a forma de um liofilizado. Os compostos podem ser também encapsulados dentro de lipossomas utilizando tecnologia bem conhecida.

Pode ser utilizado qualquer de vários imunoestimulantes nas vacinas desta invenção. Por exemplo, pode ser incluído um adjuvante. A maioria dos adjuvantes contém uma substância concebida para proteger o antigénio do catabolismo rápido, tais como o hidróxido de alumínio ou o óleo mineral e um estimulante de respostas imunitárias, tais como lípido A, proteínas obtidas de *Bordetella pertussis* ou de espécies de *Mycobacterium* ou de *Mycobacterium*. Por exemplo, pode ser utilizado *M. vaccae*

("pVac") deslipidado, desglicolipidado. Numa outra forma de realização, é utilizado o BCG como um adjuvante. Além disso, a vacina pode ser administrada a um indivíduo anteriormente exposto ao BCG. Estão comercialmente disponíveis adjuvantes adequados como, por exemplo, o Adjuvante Incompleto e o Adjuvante Completo de Freund (Difco Laboratories, Detroit, MI); Adjuvante Merck 65 (Merck and Company, Inc., Rahway, NJ); AS-2 e seus derivados (SmithKline Beecham, Filadélfia, PA); CWS, TDM, Leif, sais de alumínio, tais como gel de hidróxido de alumínio (alumina) ou fosfato de alumínio; sais de cálcio, ferro ou zinco; uma suspensão insolúvel de tirosina acilada; açúcares acilados; polissacáridos derivatizados cationicamente ou anionicamente; polifosfazenos; microesferas biodegradáveis; monofosforil lípido A e quil A. Podem ser também utilizadas citocinas, tais como GM-CSF ou interleucina 2,-7, ou-12, como adjuvantes.

Dentro das vacinas aqui proporcionadas, a composição adjuvante é, de um modo preferido, concebida para induzir uma resposta imunitária predominantemente do tipo Th1. Níveis elevados de citocinas do tipo Th1 (e. g., IFN- γ , TNF α , IL-2 e IL-12) tendem a favorecer a indução de respostas imunitárias mediadas por células contra um antigénio administrado. Pelo contrário níveis elevados de citocinas do tipo Th2 (e. g., IL-4, IL-5, IL-6 e IL-10) tendem a favorecer a indução de respostas imunitárias humorais. Após a aplicação de uma vacina como aqui proporcionada, um doente irá suportar uma resposta imunitária que inclui respostas do tipo Th1 e Th2. Dentro de uma forma de realização preferida, em que uma resposta é predominantemente do tipo Th1, o nível de citocinas do tipo Th1 irá aumentar numa maior extensão do que o nível de citocinas do tipo Th2. Os níveis destas citocinas podem ser facilmente avaliados

utilizando ensaios padrão. Para uma revisão das famílias de citocinas, ver Mosmann e Coffman, *Ann. Rev. Immunol.* **7**:145-173 (1989).

Os adjuvantes preferidos para utilização na promoção de uma resposta predominantemente do tipo Th1 incluem, por exemplo, uma combinação do monofosforil lípido A, de um modo preferido, monofosforil lípido A 3-des-O-acilado (3D-MPL), em conjunto com um sal de alumínio. Os adjuvantes MPL estão disponíveis de Corixa Corporation (Seattle, WA; ver as Patentes US N°. 4436727; 4877611; 4866034 e 4912094). Os oligonucleótidos contendo CpG (em que o dinucleótido CpG é não metilado) induzem também uma resposta predominantemente Th1. Esses oligonucleótidos são bem conhecidos e estão descritos, por exemplo, nos documentos WO 96/02555, WO 99/33488 e nas Patente U.S. N° 6008200 e 5856462. São também descritas sequências de ADN imunoestimulantes, por exemplo, por Sato *et al.*, *Science* **273**:352 (1996). Outro adjuvante preferido, compreende uma saponina, tais como Quil A ou os seus derivados, incluindo QS21 e QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escina; Digitonina; ou saponinas de *Gypsophila* ou de *Chenopodium quinoa*. Outras formulações preferidas incluem mais de uma saponina nas combinações de adjuvante da presente invenção, por exemplo combinações de, pelo menos, dois do grupo seguinte compreendendo QS21, QS7, Quil A, β -escina ou digitonina.

Alternativamente as formulações de saponina podem ser combinadas com veículos de vacina compostos por quitosano ou outros polímeros policatiônicos, partículas de polilactido e polilactido-co-glicólido, matriz de polímero baseada em poli-N-acetilglucosamina, partículas compostas por polissacáridos ou polissacárido modificados quimicamente,

lipossomas e partículas à base de lípidos, partículas compostas por monoésteres do glicerol, etc. As saponinas podem ser também formuladas na presença de colesterol para formar estruturas particuladas, tais como lipossomas ou ISCOM. Além disso, as saponinas podem ser formuladas em conjunto com um éter ou éster de polioxietileno, numa solução ou suspensão não-particulada ou numa estrutura particulada, tais como um lipossoma paucilamelar ou ISCOMS. As saponinas podem ser também formuladas com excipientes tal como Carbopol^R para aumentar a viscosidade ou serem formuladas sob a forma de pó seco com um excipiente em pó tal como a lactose.

Numa forma de realização preferida, o sistema de adjuvante inclui a combinação de um monofosforil lípido A e de um derivado de saponina, tal como a combinação de QS21 e adjuvante 3D-MPL[®], como descrito no documento WO 94/00153 ou uma composição menos reactogénica em que a QS21 é atenuada com colesterol, como descrito no documento WO 96/33739. Outras formulações preferidas compreendem uma emulsão de óleo-em-água e tocoferol. No documento WO 95/17210 é descrita uma outra formulação de adjuvante particularmente preferida utilizando QS21, adjuvante 3D-MPL[®] e tocoferol numa emulsão de óleo-em-água.

Outro sistema de adjuvante melhorado envolve a combinação de um oligonucleótido contendo CpG- e um derivado de saponina particularmente a combinação de CpG e QS21 como divulgado no documento WO 00/09159. De um modo preferido, a formulação compreende adicionalmente uma emulsão de óleo em água e tocoferol.

Outros adjuvantes preferidos incluem Montanide ISA 720 (Seppic, França), SAF (Chiron, Califórnia, Estados Unidos),

ISCOMS (CSL), MF-59 (Chiron), série SBAS de adjuvantes (e. g., SBAS-2, AS2', AS2," SBAS-4 ou SBAS6, disponível de SmithKline Beecham, Rixensart, Bélgica), Detox (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) e outros 4-fosfatos de aminoalquilglucosaminida (AGP), tais como os descritos nos Pedidos de Patente U.S. pendentes com os N°. de Série 08/853826 e 09/074720, e adjuvantes de éter de polioxietileno, tal como os descritos no documento WO. 99/52549A1.

Outros adjuvantes preferidos incluem moléculas de adjuvante da fórmula geral (I): $\text{HO}(\text{CH}_2\text{CH}_2\text{O})_n\text{-A-R}$, em que, n é 1-50, A é uma ligação ou $\text{C}(\text{O})-$, R é alquilo C_{1-50} ou Fenilalquilo C_{1-50} .

Uma forma de realização da presente invenção consiste numa formulação de vacina compreendendo um éter de polioxietileno de fórmula geral (I), em que n se situa entre 1 e 50, de um modo preferido, 4-24, de um modo muito preferido, 9; o componente R é C_{1-50} , de um modo preferido, alquilo $\text{C}_4\text{-C}_{20}$ e de um modo muito preferido, alquilo C_{12} e A é uma ligação. A concentração dos éteres de polioxietileno deve-se situar na gama de 0,1-20%, de um modo preferido, de 0,1-10% e, de um modo muito preferido, na gama de 0,1-1%. Os éteres de polioxietileno preferidos são seleccionados do grupo seguinte: éter polioxietileno-9-laurílico, éter polioxietileno-9-esteorílico, éter polioxietileno-8-esteorílico, éter polioxietileno-4-laurílico, éter polioxietileno-35-laurílico e éter polioxietileno-23-laurílico. Os éteres de polioxietileno tal como o éter polioxietilenolaurílico são descritos no índice Merck (12ª edição: entrada 7717). Estas moléculas adjuvantes são descritas no documento WO 99/52549.

O éter de polioxietileno de acordo com a fórmula geral (I) acima pode, se pretendido, ser combinado com outro adjuvante. Por exemplo, uma combinação de adjuvante preferida é, de um modo preferido, com CpG como descrito no pedido de patente UK pendente GB 9820956.2.

Qualquer vacina aqui proporcionada pode ser preparada utilizando métodos bem conhecidos que resultam numa combinação de antigénio, resposta imunitária intensificadora e um veículo ou excipiente adequado. As composições aqui descritas podem ser administradas como parte de uma formulação de libertação prolongada (*i. e.*, uma formulação, tais como uma cápsula, esponja ou gel (por exemplo, constituída por polissacáridos) que efectua uma libertação lenta do composto após a administração). Essas formulações podem ser, geralmente, preparadas utilizando tecnologia bem conhecida (ver, *e. g.*, Coombes *et al.*, *Vaccine* **14**:1429-1438 (1996)) e serem administradas, por exemplo, por implantação oral, rectal ou subcutânea ou por implantação no local alvo pretendido. As formulações de libertação prolongada podem conter um polipéptido, polinucleótido ou anticorpo disperso numa matriz veículo e/ou contido dentro de um reservatório rodeado de uma membrana de controlo da taxa.

Os veículos para utilização dentro dessas formulações são biocompatíveis e podem ser também biodegradáveis; de um modo preferido, a formulação proporciona um nível relativamente constante de libertação do componente activo. Esses veículos incluem micropartículas de poli(lactido-co-glicólido), poliacrilato, látex, amido, celulose, dextrano e semelhantes. Outros veículos de libertação retardada incluem biovectores supramoleculares que compreendem um núcleo hidrofílico não-líquido (*e. g.*, um polissacárido ou oligossacárido

reticulado) e, opcionalmente, uma camada externa compreendendo um composto por anfifílico, tal como um fosfolípido (ver e. g., Patente U.S. N° 5151254 e pedidos PCT WO 94/20078, WO/94/23701 e WO 96/06638). A quantidade de composto activo contida dentro de uma formulação de libertação prolongada depende do local de implantação, da taxa e da duração esperada da libertação e da natureza do estado a ser tratado ou prevenido.

Pode ser utilizado qualquer um de vários veículos de transferência dentro das composições farmacêuticas e vacinas para facilitar a produção de uma resposta imunitária específica de antigénio que visa células tumorais. Os veículos de transferência incluem células apresentadoras de antigénio (APC), tais como células dendríticas, macrófagos, células B, monócitos e outras células que possam ser modificadas para serem APC eficientes. Essas células podem ser, mas não é necessário, geneticamente modificadas para aumentar a capacidade para apresentar o antigénio, melhorar a activação e/ou a manutenção da resposta de células T, para ter efeitos anti-tumorais *per se* e/ou serem imunologicamente compatíveis com o receptor (*i. e.*, haplotipo HLA correspondente). As APC podem ser de um modo geral isoladas a partir de qualquer de vários fluidos biológicos e órgãos, incluindo tecidos tumorais e peritumorais e podem ser células autólogas, alogénicas, singenésicas ou xenogénicas.

Algumas formas de realização preferidas da presente invenção utilizam células dendríticas ou seus progenitores como células apresentadoras de antigénio. As células dendríticas são APC altamente potentes (Banchereau e Steinman, *Nature* **392**:245-251 (1998)) e demonstraram ser eficazes como um adjuvante fisiológico para promover a imunidade antitumoral profiláctica ou terapêutica (ver Timmerman e Levy, *Ann. Rev. Med.* **50**:507-529

(1999)). Em geral, as células dendríticas podem ser identificadas com base na sua forma típica (radiada *in situ*, com processos citoplasmáticos significativos (dendrites) *in vitro* visíveis), na sua capacidade para incorporar, processar e apresentar antígenos com uma eficiência elevada e na sua capacidade para activar respostas de células T ingénuas. As células dendríticas podem ser, obviamente, concebidas para expressar receptores celulares superficiais específicos ou ligandos que não são normalmente encontrados em células dendríticas *in vivo* ou *ex vivo* e essas células dendríticas modificadas são contempladas pela presente invenção. Podem ser utilizadas vesículas carregadas de antígeno segregadas por células dendríticas (designadas exossomas) dentro de uma vacina como uma alternativa às células dendríticas, (ver Zitvogel *et al.*, *Nature Med.* **4**:594-600 (1998)).

Podem ser obtidas células dendríticas e progenitores a partir de sangue periférico, medula óssea, células infiltrantes de tumor, células infiltrantes de tecidos peritumorais, nódulos linfáticos, baço, pele, sangue do cordão umbilical ou qualquer outro tecido ou fluido adequado. Por exemplo, as células dendríticas podem ser diferenciadas *ex vivo* adicionando uma combinação de citocinas, tais como GM-CSF, IL-4, IL-13 e/ou TNF α a culturas de monócitos recolhidos de sangue periférico. Alternativamente, podem ser diferenciadas células CD34 positivas recolhidas do sangue periférico, medula óssea ou sangue do cordão umbilical em células dendríticas por adição ao meio de cultura de combinações de GM-CSF, IL-3, TNF α , ligando CD40, LPS, ligando flt3 e/ou outro(s) composto(s) que induza(m) a diferenciação, a maturação e a proliferação de células dendríticas.

As células dendríticas são convenientemente categorizadas como células “imaturas” e “maduras”, o que proporciona um modo simples de discriminar entre dois fenótipos bem caracterizados. No entanto, esta nomenclatura não deve ser interpretada como excluindo todas as etapas intermediárias possíveis da diferenciação. As células dendríticas imaturas são caracterizadas como APC com uma elevada capacidade de incorporação e processamento de antigénio, o que se correlaciona com a expressão elevada do receptor Fcγ e do receptor da manose. O fenótipo maduro é tipicamente caracterizado por uma expressão inferior destes marcadores, mas por uma expressão elevada de moléculas da superfície celular responsáveis pela activação de célula T, tais como MHC da classe I e classe II, moléculas de adesão (e. g., CD54 e CD11) e moléculas co-estimulantes (e. g., CD40, CD80, CD86 e 4-1BB).

As APC podem ser, geralmente, transfectadas com um polinucleótido codificando uma proteína (ou porção ou sua outra variante) de forma a que o polipéptido ou uma sua porção imunogénica, sejam expressos na superfície celular. Essa transfecção pode realizar-se *ex vivo* e pode ser, então, utilizada uma composição ou uma vacina compreendendo essas células transfectada para objectivos terapêuticos, como aqui descrito. Alternativamente, pode ser administrado um veículo de transferência génica que visa uma célula dendrítica ou outras apresentadoras de antigénio a um doente, resultando na transfecção que ocorre *in vivo*. A transfecção *in vivo* e *ex vivo* de células dendríticas pode ser, por exemplo, geralmente, realizada utilizando qualquer método conhecido na técnica, tais como os descritos no documento WO 97/24447 ou a abordagem do disparador génico descrita por Mahvi *et al.*, *Immunology and Cell Biology* **75**:456-460 (1997). O carregamento das células

dendríticas com antígeno pode ser realizado incubando células dendríticas ou células progenitoras com o polipéptido, ADN (nu ou dentro de um vector plasmídico) ou ARN; ou com bactéria ou vírus recombinante expressando o antígeno (e. g., vectores de vacinia, de varíola aviária, adenovirais ou lentivirais). Antes do carregamento, o polipéptido pode ser conjugado covalentemente com um parceiro imunológico que proporciona o auxílio de células T (e. g., uma molécula veículo). Alternativamente, pode ser pulsada uma célula dendrítica com um parceiro imunológico não-conjugado, separadamente ou na presença do polipéptido.

As vacinas e as composições farmacêuticas podem ser apresentadas em recipientes de dose unitária ou multi-dose, tais como ampolas seladas ou frasquinhos. Esses recipientes são de um modo preferido, hermeticamente selados para conservar a esterilidade da formulação até à utilização. Em geral, as formulações podem ser armazenadas sob a forma de suspensões, soluções ou emulsões em veículos oleosos ou aquosos. Alternativamente, uma vacina ou composição farmacêutica pode ser armazenada num estado liofilizado requerendo apenas a adição de um veículo líquido estéril imediatamente antes da utilização.

EXEMPLOS

Os exemplos seguintes são proporcionados apenas a título de ilustração e não a título de limitação. O especialista na técnica irá reconhecer facilmente vários parâmetros não-críticos que podem ser alterados ou modificados para produzir resultados essencialmente semelhantes.

Exemplo Antecedente 1:

Vacinação de cobaios com proteína de fusão MTB72F e composições com antígenos individuais

Os cobaios foram imunizados com apenas adjuvante (SBAS1, SBAS2 ou ASAS7 mais $\text{Al}(\text{OH})_3$), proteína de fusão MTB72F em adjuvante ou TbH9 mais a composição de antígeno Ra35.

Métodos:

- Grupos:
- 1) SBAS1
 - 2) SBAS2
 - 3) SBAS7 + $\text{Al}(\text{OH})_3$
 - 4) TbH9+Ra35 + SBAS 1
 - 5) TbH9 + Ra35 + SBAS2
 - 6) TbH9 + Ra35 + SBAS7 ($\text{Al}(\text{OH})_3$)
 - 7) MTB72F em SBAS 1
 - 8) MTB72F em SBAS2
 - 9) MTB72F em SBAS7 + $\text{Al}(\text{OH})_3$
 - 10) PBS
 - 11) BCG

Dosagem: 4 μg de cada um de TbH9 e Ra35
8 μg MTB72F

Protocolo: 1ª imunização, 2ª imunização, aproximadamente, 3 semanas mais tarde, 3ª imunização aproximadamente duas semanas e meia mais tarde.

Pré-estimulação: DTH (hipersensibilidade do tipo retardado, utilizada para determinar a antigenicidade; 10 μg de antígeno)

Estimulação:	Aerossol com ~30 cfu de estirpe Erdman
Monitorização após estimulação:	Perda de peso
	Morte (~6 meses após estimulação)

Resultados:

1. DTH

Reacção positiva aos antigénios imunizantes. As reacções a antigénios individuais ou à proteína de fusão foram comparáveis. A reactividade do teste dérmico a PPD foi apenas verificada com grupos imunizados com BCG

2. Protecção: Os cobaios vacinados com a proteína de fusão MTB72F obtiveram protecção em comparação com os imunizados com uma mistura de antigénios (ver a Figura 1).

Exemplo Antecedente 2:

Vacinação de murganhos com proteína de fusão MTB72F e composições com antigénios individuais

Como descrito acima, os murganhos foram imunizados com apenas adjuvante (SBAS2, SBAS2', SBAS2'' ou SBAS6), proteína de fusão de MTB72F em adjuvante, ADN de MTB72F, proteína de fusão MTB59F em adjuvante ou composição de antigénio TbH9, Ra35 e Ra12.

Métodos:

- Grupos:
- 1) MTB72F + SBAS2
 - 2) MTB72F + SBAS2'
 - 3) MTB72F + SBAS2''
 - 4) MTB72F + SBAS6
 - 5) Ra12 + TbH9 + Ra35 em SEAS2
 - 6) MTB59F em SBAS2
 - 7) SBAS2
 - 8) MTB72F + *M.vaccae* deslipidado, desglicolipidado
 - 9) ADN de MTB72F
 - 10) MTB72F + IFA
 - 11) MTB72F + BCG
 - 12) *M.vaccae* deslipidado, desglicolipidado
 - 13) BCG
 - 14) Soro fisiológico
 - 15) MTB72F +SBAS2 (formulação proprietária)

8 animais por grupo

Calendário de imunização: primeira imunização, segunda imunização aproximadamente 3 semanas mais tarde; terceira imunização aproximadamente três semanas mais tarde.

Estimulação com aerossol aproximadamente três meses após a primeira dose

Foram isoladas e cultivadas células de baço ou de pulmão; contagem de CFU das culturas aproximadamente três semanas após o plaqueamento.

Dose: 8 µg MTB72F, 6,56 µg MTB59F ou, respectivamente, 1,52, 4,3 e 2,24 µg, de Ra12, TbH9 e Ra35, misturadas.

Resultados:

Do adjuvantes AS, AS2''' + MTB72F proporcionaram a melhor protecção tanto no baço, como nos pulmões neste conjunto de experiências (ver as Figuras 2A e 2B). O MTB72F proporcionou ~1 log de melhor protecção do que MTB59F tanto no baço como nos pulmões neste conjunto de experiências, indicando que Ra12 proporciona um benefício adicional. A mistura de 12/H9/35 + AS2 proporcionou uma melhor protecção do que MTB72F nesta experiência. O ADN de MTB72F proporcionou a melhor protecção nesta experiência, particularmente no baço (>2 log). A protecção nos pulmões foi comparável à observada com a proteína MTB72F + AS2'', nesta experiência.

Exemplo Antecedente 3:

Vacinação de cobaios com a proteína de fusão MTB72F e composições com antigénios individuais

Como descrito acima, os cobaios foram imunizados com apenas adjuvante (SBAS2, SBAS2', SBAS2'' ou SBAS6), proteína de fusão MTB72F em adjuvante, ADN de MTB72F, proteína de fusão MTB59F em adjuvante ou TbH9, composição de antigénios Ra35 e Ra12.

Métodos:

- Grupos:
- 1) MTB72F + SBAS2
 - 2) MTB72F + SBAS2'
 - 3) MTB72F + SBAS2''
 - 4) MTB72F + SBAS6
 - 5) Ra12 + TbH9 + Ra35 em SBAS2
 - 6) MTB59F em SBAS2
 - 7) SBAS2
 - 8) MTB72F + pvac
 - 9) ADN de MTB72F
 - 10) MTB72F + IFA
 - 11) MTB72F + BCG
 - 12) BCG
 - 13) Soro fisiológico
 - 14) *M.vaccae* deslipidado, desglicolipidado

Antigénios:

Os antigénios foram formulados num equivalente molar

5 animais por grupo

O volume de injeção por dose é de 250 µL (IM) contendo

MTB72F	20 µg
Ra12, TbH9, Ra35	3,8, 10,8 e 5,6 µg
MTB59F	16,4 µg

Calendarização:

1ª imunização, 2ª imunização aproximadamente três semanas mais tarde, 3ª imunização aproximadamente três semanas mais tarde.

Estimulação: ~um mês e meio após a primeira imunização.

Resultados:

~38 Smn após a estimulação

Grupos	Vivos	Estado
G1. MTB72F + AS2	1/5	[perda de peso]
G2. MTH72F + AS2'	2/5	[sem ganho de peso]
G3. MTB72F + AS2''	3/5	[aspecto normal, mas sem ganho de peso]
G4. MTB72F + AS6	2/5	[ambos com ganho de peso]
G5. MTBRa12 + H9 + Ra35 +AS2	4/5	[um possivelmente um pouco no máximo, mas dois com ganho]
G6. MTB59F + AS2	2/5	[ambos com um pouco de perda]
Grupo de Sete. AS2	2/5	[ambos com perda]
G8. MTB72F + pVac	1/5	[aspecto pouco normal]
G9. ADN de MTB72F	3/5	[totalidade mantendo-se estável]
G10. MTB72F + IFA	2/5	[normais]
G11. MTB72F + BCG	5/5	[comendo muito bem]
G12 BCG	4/5	[normais]
Soro fisiológico de G13	todos mortos	
G14 pVac	2/5	[sem ganho de peso]

Às 50 semanas após a estimulação, enquanto 80% (4/5) dos cobaias imunizadas com BCG + Mtb72F estavam ainda vivas, apenas 20% (1/5) das imunizados apenas com BCG estavam vivos. Às 85 semanas, 4/5 dos cobaias imunizados com BCG + Mtb72F estavam ainda vivos e saudáveis (ver a Figura 7).

Exemplo Antecedente 4:

Protecção a longo prazo

Como descrito acima, os cobaias foram imunizados com apenas adjuvante (AS2 ou AS2''), proteína de fusão MTB72F em adjuvante, e composição de antigénio TbH9, Ra35 Ra12 ou vários antigénios individuais em adjuvante.

Métodos

GRUPOS	DOSE DE ANTIGÉNIO
1. AS2'' + MTB39 (TbH9)	20 ug/250 uL (IM)
2. AS2'' + MTB8.4 (DPV)	20 ug
3. AS2'' + MTB9.9 (MTI)	20 ug
4. AS2'' + MTB41 (MTCC#2)	20 ug
5. AS2'' + MTB40 (HTCC#1)	20 ug
6. AS2'' + MTB9.8 (MSL)	20 ug
7. AS2'' + MTB72F	20 ug
8. AS2'' + Ra12+TbH9 + Ra35 (equivalente molar)	3,8 µg +10,8 µg +5,6 µg
9. AS2'' + MTB71F + MTB72F + HTCC#1	20 µg +20 µg +10 µg
10. AS2'' + Ra12	20 µg
11. BCG	

(continuação)

GRUPOS	DOSE DE ANTIGÉNIO
--------	-------------------

12. AS2''	
-----------	--

13. AS2 + MTB72F	
------------------	--

14. AS2 + Ra12 + TbH9 + Ra35	
------------------------------	--

15. AS2	
---------	--

Exemplo Antecedente 5:

Vacinação de macacos com proteína de fusão MTB72F e composições com antígenos individuais

Como descrito acima, os macacos foram imunizados com a proteína de fusão MTB72F em adjuvante SBAS2 ou com composição de antígeno MTB8.4 em adjuvante ou com uma mistura de MTB72F e MTB8.4.

Métodos:

Grupos

1. Soro fisiológico

2. BCG

3. MTB8.4/AS2

4. MTB72F/AS2

5. MTB72F/AS2 (um braço) + MTB8.4/AS2 (outro braço)

40 µg de cada antígeno

Resultados:

Às 8 semanas após a estimulação, os macacos imunizados com BCG estão a mostrar sintomas de infecção

Os dados actuais para 16 semanas de estimulação revelam a tendência seguinte:

Os grupos imunizados com MTB72F (4 e 5) estão a manter os seus pesos e têm valores baixos de ESR em comparação com o grupo 3 (imunização com MTB8.4) (Tabelas 1 e 2).

Tabela 1

Estudo de Vacina Profiláctica em Macacos *Cynomolgus* com MTB8.4 e MTB72F formulados em AS2 às 20 Semanas Após a Estimulação

Grupos	ID	Alteração do peso	Raio x do tórax		Estado
		líquido (kg)	(estabelecimento)		
AS2	1398K	-24%	Pn, bil, prog (smn 8)		Vivo
	4437B	-33%	Pn, bil, prog (smn 4)		Morto
	2959G	-8,30%	Pn, bil, prog (smn 4)		Vivo
	605AE	-14,00%	Pn, rt, estável (smn 8)		Vivo
BCG	3436A	-15,00%	Neg		Vivo
	3642G	Mais 4,5%	Pn, rt, prog (smn 8)		Vivo
	1190H	0%	Neg		Vivo
	10511	-30%	Pn, rt, prog (smn 8)		Morto
MTB8.4	3665C	-25%	Pn, rt, prog (smn 8)		Morto
	2200F	-18,00%	Pn, rt, estável (smn 8)		Vivo
	1654J	-33,00%	Pn, bil, prog (smn 4)		Morto
	4141C	-33%	Pn, bil, prog (smn 4)		Morto

(continuação)

Grupos	ID	Alteração do peso líquido (kg)	Raio x do tórax (estabelecimento)	Estado
	3061C*	Morto após estimulação IT		
MTB72F	1228G	Mais 3,6%	Bron, bil, estável durante 3 meses (smn 8)	Vivo
	3462E	-2,20%	Neg	Vivo
	4254C	Mais 1,21	Pn, rt, estável durante 3 meses (smn 4)	Vivo
	4496A	Mais 7%	Pn, rt, estável durante 1 meses (smn 8)	Vivo
MTB8.4/	4422C	-39,00%	Pn, bil, prog (smn 4)	Morto
MTB72F	4416A	Mais 11%	Pn, rt, estável durante 2 meses (smn 12)	Vivo
	2734E	Mais 12,5%	Susp infil rt, estável durante 3 meses (smn 8)	Vivo

Tabela 2

Estudo de Vacina Profiláctica em Macacos *Cynomolgus* com
MTB8.4 e MTB72F formulados em AS2

Smns Após Estimulação						
ESR						
Grupos	ID	4	8	12	16	Raio x do Tórax 16 smns
AS2	1398K	3	3	10	19	Pn, bil, progrsv
	4437B	10	20	3		Morreu
	2959G	6	3	3	0	Pn, rt, progrsv
	605AE	1	4	7	3	Pn, rt, estável
	3436A	0	8	7	15	Neg
BCG	3642G	0	0	0	0	Pn, rt, progrsv
	1190o	1	0	2	0	Neg
	10511	0	8	22	7	Pn, bil, w/furt progrsv

(continuação)

Smns Após Estimulação						
Grupos	ID	ESR				Raio x do Tórax 16 smns
		4	8	12	16	
						Morreu
	3665C	12	30	19		Morreu
<i>MTB8.4</i>	2200F	1	7	2	0	Pn, rt, progrsv
	1654J	20	8	21	7	Pn, bil, w/fur progrsv
	4141C	13	8	2	15	Pn, bil, w/fur progrsv
	3061C*	Morreu após estimulação IT				
<i>MTB72F</i>	1228G	0	1	20	0	Agora estável
	3462E	0	0	0	0	Neg
	4254C	13	0	0	0	Pn, agora estável
	4496A	5	1	0	5	Pn, rt, w/furt progrsv
<i>MTB8.4/</i>	4422C	10	3	0		Morreu
<i>MTB72F</i>	4416A	6	0	1	0	Pn, agora estável
	2734E	0	0	0	0	Susp infil, agora estável

Exemplo Antecedente 6:

Ensaio de iniciação de BCG em macacos

Foram imunizados com BCG 5 animais por grupo com quatro grupos e seguidamente repousaram, seguidamente foram imunizados como descrito acima e foram estimulados. Será utilizado o protocolo seguinte:

Grupos	Nº de animais	Imunização com Antígeno	Dose de Antígeno
1. Nada	5	AS2	
2. BCG	5	AS2	
3. BCG	5	MTB72F	40 ug

(continuação)

Grupos	Nº de animais	Imunização com Antígeno	Dose de Antígeno
4. BCG	4	Ra12 + TbH9 + Ra3	Equiv molar de antígenos na dose de MTB72F
5. BCG	4	MTB72F + MTB71F + MTB40	40 ug de MTB72F 40 ug de MTB72F 20 ug de MTB40

Todos os antígenos foram formulados em AS2

Os grupos 4 e 5 têm quatro animais cada um. Dois dos macacos imunizados com BCG morreram

Grupos	Nº de animais	Imunização com Antígeno	Antígenos para proliferação de células T e ensaios de produção de citocina
1. Nada	5	AS2	PHA, PPD, MTB72F, MTB71F, HTCC#1, DPV, MTCC#2, Ra12, TbH9, Ra35, MSL, MTI
2. BCG	5	AS2	PHA, PPD, MTB72F, MTB71F, HTCC#1, DPV, MTCC#2, Ra12, TbH9, Era35, MSL, MTI
3. BCG	5	MTB72F	PHA, PPD, MTB72F, Ra12, TbH9, Ra35
4. BCG	4	Ra12 + TbH9 + Ra35	PHA, PPD, MTB72F, Ra12, TbH9, Ra35
5. BCG	4	MTB72F + MTB71F + MTB40	PHA, PPD, MTB72F, MTB71F, HTCC#1, DPV, MTCC-2, Ra12, TbH9, Ra35, MSL, MTI

Exemplo 7: Construção de Ra35MutSA e MTB72FMutSA

A expressão de Mtb72f resulta tipicamente em alguns produtos de degradação. Além disso, a expressão das sequências completas da forma madura ou completa de Ra35 (Mtb32A) em *E. coli* foi difícil. O produto expresso foi apenas visível depois de imunotransferência com um Ab policlonal de coelho anti-Ra35 indicando níveis baixos de expressão da proteína. Mesmo assim, foram detectadas múltiplas espécies específicas (bandas) indicativas de degradação auto-catalítica (degradação) do antigénio recombinante. Foi considerado que isto era devido à expressão de Ra35FL em *E. coli* numa forma biologicamente activa.

Foi anteriormente demonstrado que era possível expressar Ra35FL como duas metades sobrepostas compreendendo o terminal N (Ra35N-term, designado Ra35) e metades C-term (Ra35C-term designado Ra12). Foi introduzida uma mutação pontual única num dos resíduos dentro da tríade do local activo (substituição de Ser por Ala; ver a Figura 6) para intensificar e estabilizar a expressão da molécula Ra35 completa. Esta forma mutageneizada de Mtb32A pode ser agora facilmente expressa em níveis elevados de uma forma estável. Além disso, foi incorporada uma substituição de um único nucleótido (T em G, resultando numa modificação de Ser em Ala na posição 710 do polipéptido de fusão) na sequência de Mtb72F na posição de nucleótido 2128 (ver a Figura 5) para estabilizar a expressão de Mtb72F.

Esta estabilização é também facilmente realizada por mutageneização de qualquer um, de quaisquer dois ou da totalidade dos três dos três resíduos compreendendo a tríade do local activo em Ra35FL, Ra35 ou Mtb72F ou em outras proteínas de fusão compreendendo Ra35 (His, Asp ou Ser). A mutagénese pode

ser realizada utilizando qualquer técnica conhecida de um especialista na técnica.

Exemplo 8: Imunização de murganhos com Ra35FLMutSA-TbH9 e MTB72FMutSA

Foram imunizados oito murganhos por grupo com as composições listadas abaixo, as quais incluem o adjuvante AS2A. Os murganhos foram então estimulados com *Mycobacterium tuberculosis* e foi medida a sobrevivência dos murganhos.

Grupo	Concentração de proteína ou de ADN
1. Proteína Mtb72f	1,5 mg/mL
2. ADN de Mtb72f	1,2 mg/mL
3. Proteína Mtb72f-85b	0,6 mg/mL
4. ADN de Mtb72f-85b	1,1 mg/mL
5. Proteína Mtb72f-MTI	1,3 mg/mL
6. ADN de Mtb72f-MTI	1,1 mg/mL
7. Proteína Mtb72f MutSA	1,7 mg/mL
8. Proteína MTB3AMutSA-TbH9	2,4 mg/mL
9. BCG	
10. AS2	
11. apenas vector	1,5 mg/mL
12. soro fisiológico	

LISTAGEM DE SEQUÊNCIAS

(2) INFORMAÇÃO PARA SEQ ID N°: 1: MTB32A (Ra35 FL)

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 1872 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADEIA: simples

(D) TOPOLOGIA: linear

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

GACTACGTTG	GTGTAGAAAA	ATCCTGCCGC	CCGGACCCTT	AAGGCTGGGA	CAATTTCTGA	60
TAGCTACCCC	GACACAGGAG	GTTACGGGAT	GAGCAATTCT	CGCCGCCGCT	CACTCAGGTG	120
GTCATGGTTG	CTGAGCGTGC	TGGCTGCCGT	CGGGCTGGGC	CTGGCCACGG	CGCCGGCCCA	180
GGCGGCCCG	CCGGCCTTGT	CGCAGGACCG	GTTCGCCGAC	TTCCCCGCGC	TGCCCCCTGA	240
CCCGTCCGCG	ATGGTCGCCC	AAGTGGCGCC	ACAGGTGGTC	AACATCAACA	CCAAACTGGG	300
CTACAACAAC	GCCGTGGGCG	CCGGGACCGG	CATCGTCATC	GATCCCAACG	GTGTCTGTCT	360
GACCAACAAC	CACGTGATCG	CGGGCGCCAC	CGACATCAAT	GCGTTCAGCG	TCGGCTCCGG	420
CCAAACCTAC	GGCGTCGATG	TGGTCGGGTA	TGACCGCACC	CAGGATGTCT	CGGTGCTGCA	480
GCTGCGCGGT	GCCGTGGGCC	TGCCGTGGGC	GGCGATCGGT	GGCGGCGTCG	CGGTTGGTGA	540
GCCCGTCGTC	GCGATGGGCA	ACAGCGGTGG	GCAGGGCGGA	ACGCCCCGTG	CGGTGCTTGG	600
CAGGGTGGTC	GCGCTCGGCC	AAACCGTGCA	GGCGTCGGAT	TCGCTGACCG	GTGCCGAAGA	660
GACATTGAAC	GGGTTGATCC	AGTTCGATGC	CGCAATCCAG	CCCGGTGATT	CGGGCGGGCC	720
CGTCGTCAAC	GGCCTAGGAC	AGGTGGTCGG	TATGAACACG	GCCGCGTCCG	ATAACTTCCA	780
GCTGTCCCAG	GGTGGGCAGG	GATTTCGCCAT	TCCGATCGGG	CAGGCGATGG	CGATCGCGGG	840
CCAAATCCGA	TCGGGTGGGG	GGTCACCCAC	CGTTCATATC	GGGCCTACCG	CCTTCCTCGG	900
CTTGGGTGTT	GTCGACAACA	ACGGCAACGG	CGCACGAGTC	CAACGCGTGG	TCGGAAGCGC	960
TCCGGCGGCA	AGTCTCGGCA	TCTCCACCGG	CGACGTGATC	ACCGCGGTCT	ACGGCGCTCC	1020
GATCAACTCG	GCCACCGCGA	TGGCGGACGC	GCTTAACGGG	CATCATCCCG	GTGACGTCAT	1080
CTCGGTGAAC	TGGCAAACCA	AGTCGGGCGG	CACGCGTACA	GGGAACGTGA	CATTGGCCGA	1140
GGGACCCCCG	GCCTGATTTC	TCGCGGATAC	CACCCGCCGG	CCGGCCAATT	GGATTGGCGC	1200
CAGCCGTGAT	TGCCGCGTGA	GCCCCGAGT	TCCGTCTCCC	GTGCGCGTGG	CATTGTGGAA	1260
GCAATGAACG	AGGCAGAACA	CAGCGTTGAG	CACCTTCCCG	TGCAGGGCAG	TTACGTCGAA	1320
GGCGGTGTGG	TCGAGCATCC	GGATGCCAAG	GACTTCGGCA	GCGCCGCCGC	CCTGCCCGCC	1380
GATCCGACCT	GGTTTAAGCA	CGCCGTCTTC	TACGAGGTGC	TGGTCCGGGC	GTTCCTTCGAC	1440
GCCAGCGCGG	ACGGTTCCGN	CGATCTGCGT	GGACTCATCG	ATCGCCTCGA	CTACCTGCAG	1500
TGGCTTGGCA	TCGACTGCAT	CTGTTGCCGC	CGTTCCTACG	ACTCACCGCT	GCGCGACGGC	1560
GGTTACGACA	TTCCGCACTT	CTACAAGGTG	CTGCCCGAAT	TCGGCACCCT	CGACGATTTC	1620
GTCGCCCTGG	TCGACACCGC	TCACCGGCGA	GGTATCCGCA	TCATCACCGA	CCTGGGTGATG	1680
AATCACACCT	CGGAGTCGCA	CCCCTGGTTC	CAGGAGTCCC	GCCGCGACCC	AGACGGACCG	1740
TACGGTGACT	ATTACGTGTG	GAGCGACACC	AGCGAGCGCT	ACACCGACGC	CCGGATCATC	1800
TTCGTGACGA	CCGAAGAGTC	GAACGTGTCA	TTCGATCCTG	TCCGCCGACA	GTTNCTACTG	1860
GCACCGATTG	TT					1872

(2) INFORMAÇÃO PARA SEQ ID N°: 2:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 355 aminoácidos

(B) TIPO: aminoácido

(C) TIPO DE CADEIA: simples

(D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 2: MTB32A (Ra35FL)

Met 1	Ser	Asn	Ser	Arg 5	Arg	Arg	Ser	Leu	Arg 10	Trp	Ser	Trp	Leu	Leu 15	Ser
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	Ala	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 85	Pro	Asn	Gly	Val	Val 90	Leu	Thr	Asn	Asn	His 95	Val
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	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	Asn	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													

<212> ADN

<213> Ra35 maduro

<400> SEQ ID N°: 3

```

catatgcatac accatcacca tcacgccccg ccggccttgt cgcaggaccg gttcgccgac 60
ttccccgcgc tgcccctcga cccgtccgcg atggtcgccc aagtggggcc acagggtggtc 120
aacatcaaca ccaaactggg ctacaacaac gccgtgggcg ccgggaccgg catcgtcatac 180
gatecccaacg gtgtcgtgct gaccaacaac cacgtgatcg cgggcgccac cgacatcaat 240
gcggttcagcg tcgggtcccg ccaaacctac ggcgtcgatg tggtcgggta tgaccgcacc 300
caggatgtcg cgggtgctgca gctgcgcggt gccggtggcc tgccgtcggc ggcgatcggg 360
ggcggcgctcg cgggttggtga gcccgctcgtc gcgatgggca acagcgggtg gcagggcgga 420
acgccccgtg cgggtgcctgg cagggtggtc gcgctcggcc aaaccgtgca ggcgtcggat 480
tcgctgaccg gtgcccgaaga gacattgaac gggttgatcc agttcgatgc cgcgatccag 540
cccggtgagg cgggcggggc cgtcgtcaac ggccctaggac aggtgggtcgg tatgaacacg 600
gccgcgtccg ataacttcca gctgtcccag ggtgggcagg gattcgccat tccgatcggg 660
caggcgatgg cgatcgcggg ccagatccga tcgggtgggg ggtcaccac cgttcatac 720
gggcctaccg ccttcctcgg cttgggtggt gtgcacaaca acggcaacgg cgcacgagtc 780
caacgcgtgg tcgggagcgc tccggcgga agtctcggca tctccaccgg cgacgtgatc 840
accgcggtcg acggcgctcc gatcaactcg gccaccgcga tggcggacgc gcttaacggg 900
catcatcccg gtgacgtcat ctcggtgacc tggcaaacca agtcgggcgg cacgcgtaca 960
gggaacgtga cattggccga gggacccccg gcctgagaat tc 1002

```

<212> PRT

<213> Ra35 maduro

<400> SEQ ID N°: 4

Met	His	His	His	His	His	His	Ala	Pro	Pro	Ala	Leu	Ser	Gln	Asp	Arg
				5					10					15	
Phe	Ala	Asp	Phe	Pro	Ala	Leu	Pro	Leu	Asp	Pro	Ser	Ala	Met	Val	Ala
			20					25					30		
Gln	Val	Gly	Pro	Gln	Val	Val	Asn	Ile	Asn	Thr	Lys	Leu	Gly	Tyr	Asn
		35					40					45			
Asn	Ala	Val	Gly	Ala	Gly	Thr	Gly	Ile	Val	Ile	Asp	Pro	Asn	Gly	Val
	50					55					60				
Val	Leu	Thr	Asn	Asn	His	Val	Ile	Ala	Gly	Ala	Thr	Asp	Ile	Asn	Ala
	65				70				75					80	
Phe	Ser	Val	Gly	Ser	Gly	Gln	Thr	Tyr	Gly	Val	Asp	Val	Val	Gly	Tyr
				85					90					95	

Asp	Arg	Thr	Gln	Asp	Val	Ala	Val	Leu	Gln	Leu	Arg	Gly	Ala	Gly	Gly			100	105	110
Leu	Pro	Ser	Ala	Ala	Ile	Gly	Gly	Gly	Val	Ala	Val	Gly	Glu	Pro	Val			115	120	125
Val	Ala	Met	Gly	Asn	Ser	Gly	Gly	Gln	Gly	Gly	Thr	Pro	Arg	Ala	Val			130	135	140
Pro	Gly	Arg	Val	Val	Ala	Leu	Gly	Gln	Thr	Val	Gln	Ala	Ser	Asp	Ser			145	150	155
Leu	Thr	Gly	Ala	Glu	Glu	Thr	Leu	Asn	Gly	Leu	Ile	Gln	Phe	Asp	Ala			165	170	175
Ala	Ile	Gln	Pro	Gly	Asp	Ser	Gly	Gly	Pro	Val	Val	Asn	Gly	Leu	Gly			180	185	190
Gln	Val	Val	Gly	Met	Asn	Thr	Ala	Ala	Ser	Asp	Asn	Phe	Gln	Leu	Ser			195	200	205
Gln	Gly	Gly	Gln	Gly	Phe	Ala	Ile	Pro	Ile	Gly	Gln	Ala	Met	Ala	Ile			210	215	220
Ala	Gly	Gln	Ile	Arg	Ser	Gly	Gly	Gly	Ser	Pro	Thr	Val	His	Ile	Gly			225	230	235
Pro	Thr	Ala	Phe	Leu	Gly	Leu	Gly	Val	Val	Asp	Asn	Asn	Gly	Asn	Gly			245	250	255
Ala	Arg	Val	Gln	Arg	Val	Val	Gly	Ser	Ala	Pro	Ala	Ala	Ser	Leu	Gly			260	265	270
Ile	Ser	Thr	Gly	Asp	Val	Ile	Thr	Ala	Val	Asp	Gly	Ala	Pro	Ile	Asn			275	280	285
Ser	Ala	Thr	Ala	Met	Ala	Asp	Ala	Leu	Asn	Gly	His	His	Pro	Gly	Asp			290	295	300
Val	Ile	Ser	Val	Thr	Trp	Gln	Thr	Lys	Ser	Gly	Gly	Thr	Arg	Thr	Gly			305	310	315
Asn	Val	Thr	Leu	Ala	Glu	Gly	Pro	Pro	Ala									325	330	

<212> ADN

<213> Ra35FLMutSA

<400> SEQ ID N°: 5

```
catatgcac accatcacca tcacgccccg ccggccttgt cgcaggaccg gttcgccgac 60
ttccccgcgc tgccccctcga ccggtccgcg atgggtcgccc aagtggggcc acagggtggtc 120
aacatcaaca ccaaactggg ctacaacaac gccgtgggcg ccgggaccgg catcgtcac 180
gatcccaacg gtgtcgtgct gaccaacaac cacgtgatcg cgggcgccac cgacatcaat 240
gcgttcagcg tcggtccgg ccaaacctac ggcgtcgatg tggtcgggta tgaccgcacc 300
caggatgtcg cgtgctgca gctgcgcggt gccggtggcc tgccgtcggc ggcgatcgg 360
ggcggcgctc cgttggtga gcccgctcgtc gcgatgggca acagcggtgg gcagggcgga 420
acgccccgtg cgtgacctg caggggtggtc gcgctcggcc aaaccgtgca ggcgtcggat 480
tcgctgaccg gtgccgaaga gacattgaac ggggtgatcc agttcgatgc cgcgatccag 540
cccgggtgat cgggccccgc cgtcgtcaac ggcctaggac aggtggtcgg tatgaacacg 600
gccggttcg ataacttcca gctgtcccag ggtgggcagg gattcgccat tccgatcggg 660
caggcgatgg cgatcgcggg ccagatccga tcgggtgggg ggtcaccac cgttcatatc 720
gggcctaccg ccttcctcgg cttgggtgtt gtcgacaaca acggcaacgg cgcacgagtc 780
caacgcgtgg tcgggagcgc tccggcgga agtctcggca tctccaccgg cgacgtgatc 840
accgcggtcg acggcgctcc gatcaactcg gccaccgca tggcggacgc gcttaacggg 900
catcatcccg gtgacgtcat ctcggtgacc tggcaaacca agtcgggcgg cacgcgtaca 960
gggaacgtga cattggccga gggacccccg gcctgagaat tc 1002
```

<212> PRT

<213> Ra35FLMutSA

<400> SEQ ID N°: 6

```
Met His His His His His His Ala Pro Pro Ala Leu Ser Gln Asp Arg
      5                      10                      15

Phe Ala Asp Phe Pro Ala Leu Pro Leu Asp Pro Ser Ala Met Val Ala
      20                      25                      30

Gln Val Gly Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn
      35                      40                      45

Asn Ala Val Gly Ala Gly Thr Gly Ile Val Ile Asp Pro Asn Gly Val
      50                      55                      60

Val Leu Thr Asn Asn His Val Ile Ala Gly Ala Thr Asp Ile Asn Ala
      65                      70                      75                      80

Phe Ser Val Gly Ser Gly Gln Thr Tyr Gly Val Asp Val Val Gly Tyr
      85                      90                      95

Asp Arg Thr Gln Asp Val Ala Val Leu Gln Leu Arg Gly Ala Gly Gly
      100                      105                      110

Leu Pro Ser Ala Ala Ile Gly Gly Gly Val Ala Val Gly Glu Pro Val
      115                      120                      125

Val Ala Met Gly Asn Ser Gly Gly Gln Gly Gly Thr Pro Arg Ala Val
      130                      135                      140

Pro Gly Arg Val Val Ala Leu Gly Gln Thr Val Gln Ala Ser Asp Ser
      145                      150                      155                      160
```

Leu	Thr	Gly	Ala	Glu	Glu	Thr	Leu	Asn	Gly	Leu	Ile	Gln	Phe	Asp	Ala	165	170	175	
Ala	Ile	Gln	Pro	Gly	Asp	Ala	Gly	Gly	Pro	Val	Val	Asn	Gly	Leu	Gly	180	185	190	
Gln	Val	Val	Gly	Met	Asn	Thr	Ala	Ala	Ser	Asp	Asn	Phe	Gln	Leu	Ser	195	200	205	
Gln	Gly	Gly	Gln	Gly	Phe	Ala	Ile	Pro	Ile	Gly	Gln	Ala	Met	Ala	Ile	210	215	220	
Ala	Gly	Gln	Ile	Arg	Ser	Gly	Gly	Gly	Ser	Pro	Thr	Val	His	Ile	Gly	225	230	235	240
Pro	Thr	Ala	Phe	Leu	Gly	Leu	Gly	Val	Val	Asp	Asn	Asn	Gly	Asn	Gly	245	250	255	
Ala	Arg	Val	Gln	Arg	Val	Val	Gly	Ser	Ala	Pro	Ala	Ala	Ser	Leu	Gly	260	265	270	
Ile	Ser	Thr	Gly	Asp	Val	Ile	Thr	Ala	Val	Asp	Gly	Ala	Pro	Ile	Asn	275	280	285	
Ser	Ala	Thr	Ala	Met	Ala	Asp	Ala	Leu	Asn	Gly	His	His	Pro	Gly	Asp	290	295	300	
Val	Ile	Ser	Val	Thr	Trp	Gln	Thr	Lys	Ser	Gly	Gly	Thr	Arg	Thr	Gly	305	310	315	320
Asn	Val	Thr	Leu	Ala	Glu	Gly	Pro	Pro	Ala							325	330		

(2) INFORMAÇÃO PARA SEQ ID N°: 7: Ra35 (MTB32A N-term)

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 615 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

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

gccccgccccgcttggtcgaggaccgggttcgcccacttccccgcgctgccccctcgaccggtccgcg
atggtcgccccagtgggggccacaggtgggtcaacatcaacaccaaactgggctacaacaacgcccgtg
ggcgccgggaccggcatcgatcgatcccaacgggtgctgctgaccaacaaccacgtgatcgcg
ggcgccaccgacatcaatgcgttcagcgctcggtccggccaaacctacggcgatgtgggtcggg
tatgaccgcacccaggatgtcgcggtgctgcagctgcgcggtgccggtggcctgccgtcggcggcg
atcggtggcggtcgcggttggtgagcccgtcgctcgcgatgggcaacagcggtgggcagggcgga
acgccccgtcggtgcctggcaggggtgggtcgcgctcggccaaacctgcagggcgatcggtg
accggtgccgaagagacattgaacgggtgatccagttcgatgccgcgatccagccgggtgaggcg
ggcggggcccgctcgtcaacggcctaggacaggtgggtcggtatgaacacggccgcgtcc

(2) INFORMAÇÃO PARA SEQ ID N°: 8: Ra35 (MTB32A N-term)

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 205 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

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

Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala Leu
Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Ala Pro Gln Val Val
Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly Thr
Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val
Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln
Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala
Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly
Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser Gly
Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala Leu
Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu Thr
Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser

Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr
Ala Ala Ser

(2) INFORMAÇÃO PARA SEQ ID N°: 9: Ra12

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 447 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

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

CGGTATGAAC	ACGGCCGCGT	CCGATAACTT	CCAGCTGTCC	CAGGGTGGGC	AGGGATTGCG	60
CATTCCGATC	GGGCAGGCGA	TGGCGATCGC	GGGCCAGATC	CGATCGGGTG	GGGGGTCACC	120
CACCGTTCAT	ATCGGGCCTA	CCGCCTTCCT	CGGCTTGGGT	GTTGTCGACA	ACAACGGCAA	180
CGGCGCACGA	GTCCAACGCG	TGGTCGGGAG	CGCTCCGGCG	GCAAGTCTCG	GCATCTCCAC	240
CGGCGACGTG	ATCACCGCGG	TCGACGGGCG	TCCGATCAAC	TCGGCCACCG	CGATGGCGGA	300
CGCGCTTAAC	GGGCATCATC	CCGGTGACGT	CATCTCGGTG	AACTGGCAAA	CCAAGTCGGG	360
CGGCACGCGT	ACAGGGAACG	TGACATTGGC	CGAGGGACCC	CCGGCCTGAT	TTCGTCGYGG	420
ATACCACCCG	CCGGCCGGCC	AATTGGA				447

(2) INFORMAÇÃO PARA SEQ ID N°: 10: Ra12

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 132 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

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

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	Asn	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												

(2) INFORMAÇÃO PARA SEQ ID N°: 11: TbH9

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 851 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

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

CTGCAGGGTG	GCGTGGATGA	GCGTCACCGC	GGGGCAGGCC	GAGCTGACCG	CCGCCCAGGT	60
CCGGGTGCT	GCGGCGGCT	ACGAGACGGC	GTATGGGCTG	ACGGTGCCCC	CGCCGGTGAT	120
CGCCGAGAAC	CGTGCTGAAC	TGATGATTCT	GATAGCGACC	AACCTCTTGG	GGCAAAACAC	180
CCCGGCGATC	GCGGTCAACG	AGGCCGAATA	CGGCGAGATG	TGGGCCCAAG	ACGCCGCCGC	240
GATGTTTGGC	TACGCCGCGG	CGACGGCGAC	GGCGACGGCG	ACGTTGCTGC	CGTTCGAGGA	300
GGCGCCGGAG	ATGACCAGCG	CGGGTGGGCT	CCTCGAGCAG	GCCGCCGCGG	TCGAGGAGGC	360
CTCCGACACC	GCCGCGGCGA	ACCAGTTGAT	GAACAATGTG	CCCCAGGCGC	TGAAACAGTT	420
GGCCCAGCCC	ACGCAGGGCA	CCACGCCTTC	TTCCAAGCTG	GGTGGCCTGT	GGAAGACGGT	480
CTCGCCGCAT	CGGTCGCCGA	TCAGCAACAT	GGTGTGATG	GCCAACAACC	ACATGTCGAT	540
GACCAACTCG	GGTGTGTCGA	TGACCAACAC	CTTGAGCTCG	ATGTTGAAGG	GCTTTGCTCC	600
GGCGGCGGCC	GCCAGGCCG	TGCAAACCGC	GGCGCAAAC	GGGGTCCGGG	CGATGAGCTC	660
GCTGGGCAGC	TCGCTGGGTT	CTTCGGGTCT	GGGCGGTGGG	GTGGCCGCCA	ACTTGGGTCTG	720
GGCGGCCTCG	GTACGGTATG	GTCACCGGGA	TGGCGGAAAA	TATGCANAGT	CTGGTCGGCG	780
GAACGGTGGT	CCGGCGTAAG	GTTTACCCCC	GTTTTCTGGA	TGCGGTGAAC	TTCGTCAACG	840
GAAACAGTTA	C					851

(2) INFORMAÇÃO PARA SEQ ID N°: 12:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 263 aminoácidos

(B) TIPO: aminoácido

(C) TIPO DE CADEIA: simples

(D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 12: TbH9

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

(2) INFORMAÇÃO PARA SEQ ID N°: 13: TBH9FL

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 3058 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADEIA: simples

(D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 13:

GATCGTACCC GTGCGAGTGC TCGGGCCGTT TGAGGATGGA GTGCACGTGT CTTTCGTGAT	60
GGCATAACCA GAGATGTTGG CGGCGGCGGC TGACACCCTG CAGAGCATCG GTGCTACCAC	120
TGTGGCTAGC AATGCCGCTG CGGCGGCCCC GACGACTGGG GTGGTGCCCC CCGCTGCCGA	180
TGAGGTGTCG GCGCTGACTG CGGCGCACTT CGCCGCACAT GCGGCGATGT ATCAGTCCGT	240
GAGCGCTCGG GCTGCTGCGA TTCATGACCA GTTCGTGGCC ACCCTTGCCA GCAGCGCCAG	300
CTCGTATGCG GCCACTGAAG TCGCCAATGC GCGGCGGCCC AGCTAAGCCA GGAACAGTCG	360
GCACGAGAAA CCACGAGAAA TAGGGACACG TAATGGTGGA TTTCGGGGCG TTACCACCGG	420
AGATCAACTC CGCGAGGATG TACGCCGGCC CGGGTTCGGC CTCGCTGGTG GCCGCGGCTC	480
AGATGTGGGA CAGCGTGGCG AGTGACCTGT TTTCGGCCGC GTCGGCGTTT CAGTCGGTGG	540
TCTGGGGTCT GACGGTGGGG TCGTGATAG GTTCGTGGC GGGTCTGATG GTGGCGGCGG	600
CCTCGCCGTA TGTGGCGTGG ATGAGCGTCA CCGCGGGCA GGCCGAGCTG ACCGCCGCC	660
AGGTCGGGT TGCTGCGGCG GCCTACGAGA CGGCGTATGG GCTGACGGTG CCCCCGCCG	720
TGATCGCCGA GAACCGTGCT GAACTGATGA TTCTGATAGC GACCAACCTC TTGGGGCAAA	780
ACACCCGGC GATCGCGGTC AACGAGGCCG AATACGGCGA GATGTGGGCC CAAGACGCCG	840
CCGCGATGTT TGGCTACGCC GCGGCGACGG CGACGGCGAC GCGGACGTTG CTGCCGTTCC	900
AGGAGGCGCC GGAGATGACC AGCGCGGGTG GGCTCCTCGA GCAGGCCGCC GCGGTCGAGG	960
AGGCCTCCGA CACCGCCGCG GCGAACCAGT TGATGAACAA TGTGCCCCAG GCGCTGCAAC	1020
AGCTGGCCCA GCCACGCGAG GGCACCACGC CTCTTCCAA GCTGGGTGGC CTGTGGAAGA	1080
CGGTCTCGCC GCATCGGTCG CCGATCAGCA ACATGGTGTC GATGGCCAAC AACCACATGT	1140
CGATGACCAA CTCGGGTGTG TCGATGACCA ACACCTGAG CTCGATGTTG AAGGGCTTTG	1200
CTCCGGCGGC GGCCGCCAG GCCGTGCAA CCGCGGCGCA AAACGGGGTC CGGGCGATGA	1260
GCTCGCTGGG CAGCTCGCTG GGTCTTTCGG GTCTGGGCGG TGGGGTGGCC GCCAACTTGG	1320
GTCGGGCGGC CTCGGTCGGT TCGTTGTCGG TGCCGCAGGC CTGGGCCGCG GCCAACCAGG	1380
CAGTACCCCC GCGGCGCGCG GCGCTGCCGC TGACCAGCCT GACCAGCGCC GCGGAAAGAG	1440
GGCCCCGGCA GATGCTGGGC GGGCTGCCGG TGGGGCAGAT GGGCGCCAGG GCCGGTGGTG	1500
GGCTCAGTGG TGTGCTGCGT GTTCCGCCGC GACCCTATGT GATGCCGCAT TCTCCGGCGG	1560
CCGGCTAGGA GAGGGGGCGC AGACTGTCGT TATTTGACCA GTGATCGGCG GTCTCGGTGT	1620
TTCCGCGGCC GGCTATGACA ACAGTCAATG TGCATGACAA GTTACAGGTA TTAGGTCCAG	1680
GTTCAACAAG GAGACAGGCA ACATGGCCTC ACGTTTTATG ACGGATCCGC ACGCGATGCG	1740

GGACATGGCG	GGCCGTTTTG	AGGTGCACGC	CCAGACGGTG	GAGGACGAGG	CTCGCCGGAT	1800
GTGGGCGTCC	GCGCAAAACA	TTTCCGGTGC	GGGCTGGAGT	GGCATGGCCG	AGGCGACCTC	1860
GCTAGACACC	ATGGCCCAGA	TGAATCAGGC	GTTTCGCAAC	ATCGTGAACA	TGCTGCACGG	1920
GGTGCGTGAC	GGGCTGGTTC	GCGACGCCAA	CAACTACGAG	CAGCAAGAGC	AGGCCTCCCA	1980
GCAGATCCTC	AGCAGCTAAC	GTCAGCCGCT	GCAGCACAAT	ACTTTTACAA	GCGAAGGAGA	2040
ACAGGTTCGA	TGACCATCAA	CTATCAATTC	GGGGATGTCT	ACGCTCACGG	CGCCATGATC	2100
CGCGCTCAGG	CCGGGTTGCT	GGAGGCCGAG	CATCAGGCCA	TCATTCTGTA	TGTGTTGACC	2160
GCGAGTGAAT	TTTGGGGCGG	CGCCGGTTCG	GCGGCCTGCC	AGGGGTTTCAT	TACCCAGTTG	2220
GGCCGTAACT	TCCAGGTGAT	CTACGAGCAG	GCCAACGCCC	ACGGGCAGAA	GGTGCAGGCT	2280
GCCGGCAACA	ACATGGCGCA	AACCGACAGC	GCCGTCGGCT	CCAGCTGGGC	CTGACACCAG	2340
GCCAAGGCCA	GGGACGTGGT	GTACGAGTGA	AGTTCCTCGC	GTGATCCTTC	GGGTGGCAGT	2400
CTAAGTGGTC	AGTGCTGGGG	TGTTGGTGGT	TTGCTGCTTG	GCGGGTTCTT	CGGTGCTGGT	2460
CAGTGCTGCT	CGGGCTCGGG	TGAGGACCTC	GAGGCCCAGG	TAGCGCCGTC	CTTCGATCCA	2520
TTCGTCGTGT	TGTTCCGGCA	GGACGGCTCC	GACGAGGCGG	ATGATCGAGG	CGCGGTCGGG	2580
GAAGATGCCC	ACGACGTCGG	TTCGGCGTCG	TACCTCTCGG	TTGAGGCGTT	CCTGGGGGTT	2640
GTGAGACCAG	ATTTGGCGCC	AGATCTGCTT	GGGGAAGGCG	GTGAACGCCA	GCAGGTCGGT	2700
GCGGGCGGTG	TCGAGGTGCT	CGGCCACCGC	GGGGAGTTTG	TCGGTCAGAG	CGTCGAGTAC	2760
CCGATCATAT	TGGGCAACAA	CTGATTCGGC	GTCGGGCTGG	TCGTAGATGG	AGTGCAGCAG	2820
GGTGCGCACC	CACGGCCAGG	AGGGCTTCGG	GGTGGCTGCC	ATCAGATTGG	CTGCGTAGTG	2880
GGTTCTGCAG	CGCTGCCAGG	CCGCTGCGGG	CAGGGTGGCG	CCGATCGCGG	CCACCAGGCC	2940
GGCGTGGGCG	TCGCTGGTGA	CCAGCGCGAC	CCCGGACAGG	CCGCGGGCGA	CCAGGTCGCG	3000
GAAGAACGCC	AGCCAGCCGG	CCCCGTCTCT	GGCGGAGGTG	ACCTGGATGC	CCAGGATC	3058

(2) INFORMAÇÃO PARA SEQ ID N°: 14: TbH9FL

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 391 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 14:

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

Met	Ser	Ser	Leu	Gly	Ser	Ser	Leu	Gly	Ser	Ser	Gly	Leu	Gly	Gly	Gly	
290						295					300					
Val	Ala	Ala	Asn	Leu	Gly	Arg	Ala	Ala	Ser	Val	Gly	Ser	Leu	Ser	Val	
305					310					315					320	
Pro	Gln	Ala	Trp	Ala	Ala	Ala	Asn	Gln	Ala	Val	Thr	Pro	Ala	Ala	Arg	
				325					330					335		
Ala	Leu	Pro	Leu	Thr	Ser	Leu	Thr	Ser	Ala	Ala	Glu	Arg	Gly	Pro	Gly	
			340					345					350			
Gln	Met	Leu	Gly	Gly	Leu	Pro	Val	Gly	Gln	Met	Gly	Ala	Arg	Ala	Gly	
	355						360					365				
Gly	Gly	Leu	Ser	Gly	Val	Leu	Arg	Val	Pro	Pro	Arg	Pro	Tyr	Val	Met	
370						375					380					
Pro	His	Ser	Pro	Ala	Ala	Gly										
385					390											

<210> SEQ ID N°: 15

<211> 2287

<212> ADN

<213> Sequência Artificial

<223> Descrição da Sequência Artificial: proteína de tri-fusão
Mtb72F (fusão Ral2-TbH9-Ra35 ou Mtb32-Mtb39)

tctagaaata attttgttta ctttaagaan ganatataca t atg cat cac cat cac	56
Met His His His His	
1 5	
cat cac acg gcc gcg tcc gat aac ttc cag ctg tcc cag ggt ggg cag	104
His His Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln	
10 15 20	
gga ttc gcc att ccg atc ggg cag gcg atg gcg atc gcg ggc cag atc	152
Gly Phe Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile	
25 30 35	
cga tcg ggt ggg ggg tca ccc acc gtt cat atc ggg cct acc gcc ttc	200
Arg Ser Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe	
40 45 50	
ctc ggc ttg ggt gtt gtc gac aac aac ggc aac ggc gca cga gtc caa	248
Leu Gly Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln	
55 60 65	

cgc	gtg	gtc	ggg	agc	gct	ccg	gcg	gca	agt	ctc	ggc	atc	tcc	acc	ggc	296
Arg	Val	Val	Gly	Ser	Ala	Pro	Ala	Ala	Ser	Leu	Gly	Ile	Ser	Thr	Gly	
70					75					80					85	
gac	gtg	atc	acc	gcg	gtc	gac	ggc	gct	ccg	atc	aac	tgc	gcc	acc	gcg	344
Asp	Val	Ile	Thr	Ala	Val	Asp	Gly	Ala	Pro	Ile	Asn	Ser	Ala	Thr	Ala	
				90					95					100		
atg	gcg	gac	gcg	ctt	aac	ggg	cat	cat	ccc	ggt	gac	gtc	atc	tgc	gtg	392
Met	Ala	Asp	Ala	Leu	Asn	Gly	His	His	Pro	Gly	Asp	Val	Ile	Ser	Val	
			105					110					115			
acc	tgg	caa	acc	aag	tgc	ggc	ggc	acg	cgt	aca	ggg	aac	gtg	aca	ttg	440
Thr	Trp	Gln	Thr	Lys	Ser	Gly	Gly	Thr	Arg	Thr	Gly	Asn	Val	Thr	Leu	
		120					125					130				
gcc	gag	gga	ccc	ccg	gcc	gaa	ttc	atg	gtg	gat	ttc	ggg	gcg	tta	cca	488
Ala	Glu	Gly	Pro	Pro	Ala	Glu	Phe	Met	Val	Asp	Phe	Gly	Ala	Leu	Pro	
	135					140					145					
ccg	gag	atc	aac	tcc	gcg	agg	atg	tac	gcc	ggc	ccg	ggt	tgc	gcc	tgc	536
Pro	Glu	Ile	Asn	Ser	Ala	Arg	Met	Tyr	Ala	Gly	Pro	Gly	Ser	Ala	Ser	
150					155					160					165	
ctg	gtg	gcc	gcg	gct	cag	atg	tgg	gac	agc	gtg	gcg	agt	gac	ctg	ttt	584
Leu	Val	Ala	Ala	Ala	Gln	Met	Trp	Asp	Ser	Val	Ala	Ser	Asp	Leu	Phe	
				170					175					180		
tgc	gcc	gcg	tgc	gcg	ttt	cag	tgc	gtg	gtc	tgg	ggt	ctg	acg	gtg	ggg	632
Ser	Ala	Ala	Ser	Ala	Phe	Gln	Ser	Val	Val	Trp	Gly	Leu	Thr	Val	Gly	
			185					190					195			
tgc	tgg	ata	ggt	tgc	tgc	gcg	ggt	ctg	atg	gtg	gcg	gcg	gcc	tgc	ccg	680
Ser	Trp	Ile	Gly	Ser	Ser	Ala	Gly	Leu	Met	Val	Ala	Ala	Ala	Ser	Pro	
		200					205					210				
tat	gtg	gcg	tgg	atg	agc	gtc	acc	gcg	ggg	cag	gcc	gag	ctg	acc	gcc	728
Tyr	Val	Ala	Trp	Met	Ser	Val	Thr	Ala	Gly	Gln	Ala	Glu	Leu	Thr	Ala	
	215					220					225					
gcc	cag	gtc	cgg	gtt	gct	gcg	gcg	gcc	tac	gag	acg	gcg	tat	ggg	ctg	776
Ala	Gln	Val	Arg	Val	Ala	Ala	Ala	Ala	Tyr	Glu	Thr	Ala	Tyr	Gly	Leu	
230					235					240				245		
acg	gtg	ccc	ccg	ccg	gtg	atc	gcc	gag	aac	cgt	gct	gaa	ctg	atg	att	824
Thr	Val	Pro	Pro	Pro	Val	Ile	Ala	Glu	Asn	Arg	Ala	Glu	Leu	Met	Ile	
				250					255					260		
ctg	ata	gcg	acc	aac	ctc	ttg	ggg	caa	aac	acc	ccg	gcg	atc	gcg	gtc	872
Leu	Ile	Ala	Thr	Asn	Leu	Leu	Gly	Gln	Asn	Thr	Pro	Ala	Ile	Ala	Val	
			265				270						275			

aac gag gcc gaa tac ggc gag atg tgg gcc caa gac gcc gcc gcg atg	920
Asn Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln Asp Ala Ala Ala Met	
280 285 290	
ttt ggc tac gcc gcg gcg acg gcg acg gcg acg gcg acg ttg ctg ccg	968
Phe Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr Leu Leu Pro	
295 300 305	
ttc gag gag gcg ccg gag atg acc agc gcg ggt ggg ctc ctc gag cag	1016
Phe Glu Glu Ala Pro Glu Met Thr Ser Ala Gly Gly Leu Leu Glu Gln	
310 315 320 325	
gcc gcc gcg gtc gag gag gcc tcc gac acc gcc gcg gcg aac cag ttg	1064
Ala Ala Ala Val Glu Glu Ala Ser Asp Thr Ala Ala Ala Asn Gln Leu	
330 335 340	
atg aac aat gtg ccc cag gcg ctg caa cag ctg gcc cag ccc acg cag	1112
Met Asn Asn Val Pro Gln Ala Leu Gln Gln Leu Ala Gln Pro Thr Gln	
345 350 355	
ggc acc acg cct tct tcc aag ctg ggt ggc ctg tgg aag acg gtc tcg	1160
Gly Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser	
360 365 370	
ccg cat cgg tcg ccg atc agc aac atg gtg tcg atg gcc aac aac cac	1208
Pro His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His	
375 380 385	
atg tcg atg acc aac tcg ggt gtg tcg atg acc aac acc ttg agc tcg	1256
Met Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser	
390 395 400 405	
atg ttg aag ggc ttt gct ccg gcg gcg gcc cgc cag gcc gtg caa acc	1304
Met Leu Lys Gly Phe Ala Pro Ala Ala Ala Arg Gln Ala Val Gln Thr	
410 415 420	
gcg gcg caa aac ggg gtc ccg gcg atg agc tcg ctg ggc agc tcg ctg	1352
Ala Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu	
425 430 435	
ggt tct tcg ggt ctg ggc ggt ggg gtg gcc gcc aac ttg ggt cgg gcg	1400
Gly Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala	
440 445 450	
gcc tcg gtc ggt tcg ttg tcg gtg ccg cag gcc tgg gcc gcg gcc aac	1448
Ala Ser Val Gly Ser Leu Ser Val Pro Gln Ala Trp Ala Ala Ala Asn	
455 460 465	
cag gca gtc acc ccg gcg gcg cgg gcg ctg ccg ctg acc agc ctg acc	1496
Gln Ala Val Thr Pro Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr	
470 475 480 485	
agc gcc gcg gaa aga ggg ccc ggg cag atg ctg ggc ggg ctg ccg gtg	1544
Ser Ala Ala Glu Arg Gly Pro Gly Gln Met Leu Gly Gly Leu Pro Val	
490 495 500	
ggg cag atg ggc gcc agg gcc ggt ggt ggg ctc agt ggt gtg ctg cgt	1592
Gly Gln Met Gly Ala Arg Ala Gly Gly Gly Leu Ser Gly Val Leu Arg	
505 510 515	

gtt ccg ccg cga ccc tat gtg atg ccg cat tct ccg gca gcc ggc gat	1640
Val Pro Pro Arg Pro Tyr Val Met Pro His Ser Pro Ala Ala Gly Asp	
520 525 530	
atc gcc ccg ccg gcc ttg tgc cag gac cgg ttc gcc gac ttc ccc gcg	1688
Ile Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala	
535 540 545	
ctg ccc ctc gac ccg tcc gcg atg gtc gcc caa gtg ggg cca cag gtg	1736
Leu Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Gly Pro Gln Val	
550 555 560 565	
gtc aac atc aac acc aaa ctg ggc tac aac aac gcc gtg ggc gcc ggg	1784
Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly	
570 575 580	
acc ggc atc gtc atc gat ccc aac ggt gtc gtg ctg acc aac aac cac	1832
Thr Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His	
585 590 595	
gtg atc gcg ggc gcc acc gac atc aat gcg ttc agc gtc ggc tcc ggc	1880
Val Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly	
600 605 610	
caa acc tac ggc gtc gat gtg gtc ggg tat gac cgc acc cag gat gtc	1928
Gln Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val	
615 620 625	
gcg gtg ctg cag ctg cgc ggt gcc ggt ggc ctg ccg tgc gcg gcg atc	1976
Ala Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile	
630 635 640 645	
ggt ggc ggc gtc gcg gtt ggt gag ccc gtc gtc gcg atg ggc aac agc	2024
Gly Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser	
650 655 660	
ggt ggg cag ggc gga acg ccc cgt gcg gtg cct ggc agg gtg gtc gcg	2072
Gly Gly Gln Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala	
665 670 675	
ctc ggc caa acc gtg cag gcg tgc gat tgc ctg acc ggt gcc gaa gag	2120
Leu Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu	
680 685 690	
aca ttg aac ggg ttg atc cag ttc gat gcc gcg atc cag ccc ggt gat	2168
Thr Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp	
695 700 705	
tgc ggc ggg ccc gtc gtc aac ggc cta gga cag gtg gtc ggt atg aac	2216
Ser Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn	
710 715 720 725	
acg gcc gcg tcc taggatatcc atcacactgg cggccgctcg agcagatccg	2268
Thr Ala Ala Ser	
gntgtaacaa agccccgaaa	2287

<210> SEQ ID N°: 16

<211> 729

<212> PRT

<213> Sequência Artificial

<223> Descrição da Sequência Artificial: proteína de tri-fusão
Mtb72F (fusão Ra12-TbH9-Ra35 ou Mtb32-Mtb39)

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	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
 385 390 395 400
 Asn Thr Leu Ser Ser Met Leu Lys Gly Phe Ala Pro Ala Ala Ala Arg
 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
 465 470 475 480
 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					560
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					640
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					720
Val	Val	Gly	Met	Asn	Thr	Ala	Ala	Ser							
				725											

<210> SEQ ID N°: 17

<211> 2190

<212> ADN

<213> Mtb72FMutSA

```

atgcatcacc atcaccatca cacggccgcg tccgataact tccagctgtc ccaggggtggg 60
cagggattcg ccattccgat cgggcaggcg atggcgatcg cgggccagat ccgatcgggt 120
gggggggtcac ccaccgttca tatcgggcct accgccttcc tcggccttggg tgttgtcgac 180
aacaacggca acggcgcacg agtccaacgc gtggtcggga gcgctccggc ggcaagtctc 240
ggcatctcca cggcgacgt gatcaccgcg gtcgacggcg ctccgatcaa ctcgccacc 300
gcgatggcgg acgcgttaa cgggcatcat cccggtgacg tcattctcgt gacctggcaa 360
accaagtccg gcggcacgcg tacagggaac gtgacattgg ccgagggacc cccggccgaa 420
ttcatggtgg atttcggggc gttaccaccg gagatcaact ccgcgaggat gtacgccggc 480
ccgggttcgg cctcgctggg ggccgcggct cagatgtggg acagcgtggc gagtgcctg 540
ttttcgggcg cgtcggcggt tcagtcgggtg gtctgggggtc tgacgggtggg gtcgtggata 600

```

```

gggttcgctcgg cggtgtctgat ggtggcgggcg gcctcgccgt atgtggcggt gatgagcgtc 660
accgcgggggc agggccgagct gaccgcccgc caggtcggg ttgctgcggc ggcctacgag 720
acggcggtatg ggctgacggt gcccccgccg gtgatcgccg agaaccgtgc tgaactgatg 780
attctgatag cgaccaacct cttggggcaa aacaccccgg cgatcgcggt caacgaggcc 840
gaatacggcg agatgtgggc ccaagacgcc gcccgatgt ttggctacgc cgcggcgacg 900
gcgacggcga cggcgacgtt gctgccgttc gaggaggcgc cggagatgac cagcgcggt 960
gggctcctcg agcaggccgc cgcggtcgag gaggcctccg acaccgcccgc ggcgaaccag 1020
ttgatgaaca atgtgcccga ggcgctgcaa cagctggccc agcccacgca gggcaccacg 1080
ccttcttcca agctgggtgg cctgtggaag acggtctcgc cgcacggtc gccgatcagc 1140
aacatggtgt cgatggccaa caaccacatg tcgatgacca actcgggtgt gtcgatgacc 1200
aacacctga gctcgatgtt gaagggcttt gctccggcgg cggccgcccga ggcggtgcaa 1260
accgcggcgc aaaacgggggt ccgggcgatg agctcgctgg gcagctcgtt gggttcttcg 1320
ggctcgggcg gtgggggtggc cgccaacttg ggtcggggcg cctcgggtcg ttcgttgtcg 1380
gtgcccgcag cctgggcccgc ggccaaccag gcagtcaccc cggcggcgcg ggcgctgccg 1440
ctgaccagcc tgaccagcgc cgcggaaaga gggcccgggc agatgctggg cgggctgccg 1500
gtggggcaga tgggcgccag ggccggtggt gggctcagtg gtgtgctgcg tgttccgcg 1560
cgacctatg tgatgccgca ttctccggca gccggcgata tcgcccgcgc ggccttgtcg 1620
caggaccggt tcgccgactt ccccgcgctg cccctcgacc cgtccgcgat ggtcgcccaa 1680
gtggggccac aggtggtcaa catcaacacc aaactgggct acaacaacgc cgtgggcgcc 1740
gggaccggca tcgtcatcga tcccaacggt gtcgtgctga ccaacaacca cgtgatcgcg 1800
ggcgccaccg acatcaatgc gttcagcgtc ggctccggcc aaacctacgg cgtcgatgtg 1860
gtcgggtatg accgcaccca ggatgtcgcg gtgctgcagc tgcgcggtgc cgggtggcctg 1920
ccgtcggcgg cgatcggtgg cggcgctcgc gttggtgagc ccgtcgtcgc gatgggcaac 1980
agcgggtgggc agggcggaac gcccgtgcg gtgcctggca ggggtggtcg gctcggccaa 2040
accgtgcagg cgtcggatc gctgaccggt gccgaagaga cattgaacgg gttgatccag 2100
ttcgatgccg cgatccagcc cggtgatgcg ggcggggccc tcgtcaacgg cctaggacag 2160
gtggtcggta tgaacacggc cgcgtcctag . 2190

```

<210> SEQ ID N°: 18

<211> 729

<212> PRT

<213> Mtb72FMutSA

```

Met His His His His His His Thr Ala Ala Ser Asp Asn Phe Gln Leu
      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
 385 390 395 400
 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
 465 470 475 480
 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
 545 550 555 560
 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
 625 630 635 640
 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 Ala Gly Gly Pro Val Val Asn Gly Leu Gly Gln
 705 710 715 720

Val Val Gly Met Asn Thr Ala Ala Ser
725

<210> SEQ ID N°: 19

<211> 1797

<212> ADN

<213> Sequência Artificial

<223> Descrição da Sequência Artificial: proteína de bi-fusão
TbH9-Ra35 (designada Mtb59f)

<222> (1)..(1791)

cat atg cat cac cat cac cat cac atg gtg gat ttc ggg gcg tta cca	48
His Met His His His His His His Met Val Asp Phe Gly Ala Leu Pro	
1 5 10 15	
ccg gag atc aac tcc gcg agg atg tac gcc ggc ccg ggt tcg gcc tcg	96
Pro Glu Ile Asn Ser Ala Arg Met Tyr Ala Gly Pro Gly Ser Ala Ser	
20 25 30	
ctg gtg gcc gcg gct cag atg tgg gac agc gtg gcg agt gac ctg ttt	144
Leu Val Ala Ala Ala Gln Met Trp Asp Ser Val Ala Ser Asp Leu Phe	
35 40 45	
tcg gcc gcg tcg gcg ttt cag tcg gtg gtc tgg ggt ctg acg gtg ggg	192
Ser Ala Ala Ser Ala Phe Gln Ser Val Val Trp Gly Leu Thr Val Gly	
50 55 60	
tcg tgg ata ggt tcg tcg gcg ggt ctg atg gtg gcg gcg gcc tcg ccg	240
Ser Trp Ile Gly Ser Ser Ala Gly Leu Met Val Ala Ala Ala Ser Pro	
65 70 75 80	
tat gtg gcg tgg atg agc gtc acc gcg ggg cag gcc gag ctg acc gcc	288
Tyr Val Ala Trp Met Ser Val Thr Ala Gly Gln Ala Glu Leu Thr Ala	
85 90 95	
gcc cag gtc cgg gtt gct gcg gcg gcc tac gag acg gcg tat ggg ctg	336
Ala Gln Val Arg Val Ala Ala Ala Tyr Glu Thr Ala Tyr Gly Leu	
100 105 110	
acg gtg ccc ccg ccg gtg atc gcc gag aac cgt gct gaa ctg atg att	384
Thr Val Pro Pro Pro Val Ile Ala Glu Asn Arg Ala Glu Leu Met Ile	
115 120 125	
ctg ata gcg acc aac ctc ttg ggg caa aac acc ccg gcg atc gcg gtc	432
Leu Ile Ala Thr Asn Leu Leu Gly Gln Asn Thr Pro Ala Ile Ala Val	
130 135 140	
aac gag gcc gaa tac ggc gag atg tgg gcc caa gac gcc gcc gcg atg	480
Asn Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln Asp Ala Ala Ala Met	
145 150 155 160	

ttt ggc tac gcc gcg gcg acg gcg acg gcg acg gcg acg ttg ctg ccg	528
Phe Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr Ala Thr Leu Leu Pro	
165 170 175	
ttc gag gag gcg ccg gag atg acc agc gcg ggt ggg ctc ctc gag cag	576
Phe Glu Glu Ala Pro Glu Met Thr Ser Ala Gly Gly Leu Leu Glu Gln	
180 185 190	
gcc gcc gcg gtc gag gag gcc tcc gac acc gcc gcg gcg aac cag ttg	624
Ala Ala Ala Val Glu Glu Ala Ser Asp Thr Ala Ala Ala Asn Gln Leu	
195 200 205	
atg aac aat gtg ccc cag gcg ctg caa cag ctg gcc cag ccc acg cag	672
Met Asn Asn Val Pro Gln Ala Leu Gln Gln Leu Ala Gln Pro Thr Gln	
210 215 220	
ggc acc acg cct tct tcc aag ctg ggt ggc ctg tgg aag acg gtc tcg	720
Gly Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser	
225 230 235 240	
ccg cat cgg tcg ccg atc agc aac atg gtg tcg atg gcc aac aac cac	768
Pro His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His	
245 250 255	
atg tcg atg acc aac tcg ggt gtg tcg atg acc aac acc ttg agc tcg	816
Met Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser	
260 265 270	
atg ttg aag ggc ttt gct ccg gcg gcg gcc gcc cag gcc gtg caa acc	864
Met Leu Lys Gly Phe Ala Pro Ala Ala Ala Ala Gln Ala Val Gln Thr	
275 280 285	
gcg gcg caa aac ggg gtc cgg gcg atg agc tcg ctg ggc agc tcg ctg	912
Ala Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu	
290 295 300	
ggt tct tcg ggt ctg ggc ggt ggg gtg gcc gcc aac ttg ggt cgg gcg	960
Gly Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala	
305 310 315 320	
gcc tcg gtc ggt tcg ttg tcg gtg ccg cag gcc tgg gcc gcg gcc aac	1008
Ala Ser Val Gly Ser Leu Ser Val Pro Gln Ala Trp Ala Ala Ala Asn	
325 330 335	
cag gca gtc acc ccg gcg gcg cgg gcg ctg ccg ctg acc agc ctg acc	1056
Gln Ala Val Thr Pro Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr	
340 345 350	
agc gcc gcg gaa aga ggg ccc ggg cag atg ctg ggc ggg ctg ccg gtg	1104
Ser Ala Ala Glu Arg Gly Pro Gly Gln Met Leu Gly Gly Leu Pro Val	
355 360 365	
ggg cag atg ggc gcc agg gcc ggt ggt ggg ctc agt ggt gtg ctg cgt	1152
Gly Gln Met Gly Ala Arg Ala Gly Gly Gly Leu Ser Gly Val Leu Arg	
370 375 380	

gtt ccg ccg cga ccc tat gtg atg ccg cat tct ccg gca gcc ggc gat	1200
Val Pro Pro Arg Pro Tyr Val Met Pro His Ser Pro Ala Ala Gly Asp	
385 390 395 400	
atc gcc ccg ccg gcc ttg tcg cag gac cgg ttc gcc gac ttc ccc gcg	1248
Ile Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala	
405 410 415	
ctg ccc ctc gac ccg tcc gcg atg gtc gcc caa gtg ggg cca cag gtg	1296
Leu Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Gly Pro Gln Val	
420 425 430	
gtc aac atc aac acc aaa ctg ggc tac aac aac gcc gtg ggc gcc ggg	1344
Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly	
435 440 445	
acc ggc atc gtc atc gat ccc aac ggt gtc gtg ctg acc aac aac cac	1392
Thr Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His	
450 455 460	
gtg atc gcg ggc gcc acc gac atc aat gcg ttc agc gtc ggc tcc ggc	1440
Val Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly	
465 470 475 480	
caa acc tac ggc gtc gat gtg gtc ggg tat gac cgc acc cag gat gtc	1488
Gln Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val	
485 490 495	
gcg gtg ctg cag ctg cgc ggt gcc ggt ggc ctg ccg tcg gcg gcg atc	1536
Ala Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile	
500 505 510	
ggt ggc ggc gtc gcg gtt ggt gag ccc gtc gtc gcg atg ggc aac agc	1584
Gly Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser	
515 520 525	
ggt ggg cag ggc gga acg ccc cgt gcg gtg cct ggc agg gtg gtc gcg	1632
Gly Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala	
530 535 540	
ctc ggc caa acc gtg cag gcg tcg gat tcg ctg acc ggt gcc gaa gag	1680
Leu Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu	
545 550 555 560	
aca ttg aac ggg ttg atc cag ttc gat gcc gcg atc cag ccc ggt gat	1728
Thr Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp	
565 570 575	
tcg ggc ggg ccc gtc gtc aac ggc cta gga cag gtg gtc ggt atg aac	1776
Ser Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn	
580 585 590	
acg gcc gcg tcc taggatatc	1797
Thr Ala Ala Ser	
595	

<210> SEQ ID N°: 20

<211> 596

<212> PRT

<213> Sequência Artificial

<223> Descrição da Sequência Artificial: proteína de bi-fusão
TbH9-Ra35 (designada Mtb59f)

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

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

Gly Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala
 530 535 540
 Leu Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu
 545 550 555 560
 Thr Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp
 565 570 575
 Ser Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn
 580 585 590
 Thr Ala Ala Ser
 595

(2) INFORMAÇÃO PARA SEQ ID N°: 21: DPV (MTB8.4)

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 500 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 21:

CGTGGCAATG	TCGTTGACCG	TCGGGGCCGG	GGTCGCCTCC	GCAGATCCCG	TGGACGCGGT	60
CATTAACACC	ACCTGCAATT	ACGGGCAGGT	AGTAGCTGCG	CTCAACGCGA	CGGATCCGGG	120
GGCTGCCGCA	CAGTTCAACG	CCTCACCAGT	GGCGCAGTCC	TATTTGCGCA	ATTTCTCGC	180
CGCACCGCCA	CCTCAGCGCG	CTGCCATGGC	CGCGCAATTG	CAAGCTGTGC	CGGGGGCGGC	240
ACAGTACATC	GGCCTTGTCG	AGTCGGTTGC	CGGCTCCTGC	AACAACCTAT	AAGCCCATGC	300
GGGCCCCATC	CCGCGACCCG	GCATCGTCGC	CGGGGCTAGG	CCAGATTGCC	CCGCTCCTCA	360
ACGGGCCGCA	TCCCGCGACC	CGGCATCGTC	GCCGGGGCTA	GGCCAGATTG	CCCCGCTCCT	420
CAACGGGCCG	CATCTCGTGC	CGAATTCCTG	CAGCCCGGGG	GATCCACTAG	TTCTAGAGCG	480
GCCGCCACCG	CGGTGGAGCT					500

(2) INFORMAÇÃO PARA SEQ ID N°: 22: DPV (MTB8.4)

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 96 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 22:

Val	Ala	Met	Ser	Leu	Thr	Val	Gly	Ala	Gly	Val	Ala	Ser	Ala	Asp	Pro
1				5					10					15	
Val	Asp	Ala	Val	Ile	Asn	Thr	Thr	Cys	Asn	Tyr	Gly	Gln	Val	Val	Ala
			20					25					30		
Ala	Leu	Asn	Ala	Thr	Asp	Pro	Gly	Ala	Ala	Ala	Gln	Phe	Asn	Ala	Ser
		35					40				45				
Pro	Val	Ala	Gln	Ser	Tyr	Leu	Arg	Asn	Phe	Leu	Ala	Ala	Pro	Pro	Pro
	50					55					60				
Gln	Arg	Ala	Ala	Met	Ala	Ala	Gln	Leu	Gln	Ala	Val	Pro	Gly	Ala	Ala
65				70					75					80	
Gln	Tyr	Ile	Gly	Leu	Val	Glu	Ser	Val	Ala	Gly	Ser	Cys	Asn	Asn	Tyr
			85						90					95	

(2) INFORMAÇÃO PARA SEQ ID N°: 23: MSL (MTB9.8)

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 585 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: dupla
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(vi) FONTE ORIGINAL:

(A) ORGANISMO: *Micobacterium tuberculosis*

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 23:

TGGATTCCGA	TAGCGGTTTC	GGCCCCCTCGA	CGGGCGACCA	CGGCGCGCAG	GCCTCCGAAC	60
GGGGGGCCGG	GACGCTGGGA	TTCGCCGGGA	CCGCAACCAA	AGAACGCCGG	GTCCGGGCGG	120
TCGGGCTGAC	CGCACTGGCC	GGTGATGAGT	TCGGCAACGG	CCCCCGGATG	CCGATGGTGC	180
CGGGGACCTG	GGAGCAGGGC	AGCAACGAGC	CCGAGGCGCC	CGACGGATCG	GGGAGAGGGG	240
GAGGCGACGG	CTTACCGCAC	GACAGCAAGT	AACCGAATTC	CGAATCACGT	GGACCCGTAC	300
GGGTGAAAAG	GAGAGATGTT	ATGAGCCTTT	TGGATGCTCA	TATCCACAG	TTGGTGGCCT	360
CCCAGTCGGC	GTTTGCCGCC	AAGGCGGGGC	TGATGCGGCA	CACGATCGGT	CAGGCCGAGC	420
AGGCGGCGAT	GTCGGCTCAG	GCGTTTCACC	AGGGGGAGTC	GTCGGCGGCG	TTTCAGGCCG	480
CCCATGCCCC	GTTTGTGGCG	GCGGCCGCCA	AAGTCAACAC	CTTGTTGGAT	GTCGCGCAGG	540
CGAATCTGGG	TGAGGCCGCC	GGTACCTATG	TGGCCGCCGA	TGCTG		585

(2) INFORMAÇÃO PARA SEQ ID N°: 24: MSL (MTB9.8)

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 97 aminoácidos

(B) TIPO: aminoácido

(C) TIPO DE CADEIA: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 24:

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

(2) INFORMAÇÃO PARA SEQ ID N°: 25: MTI (MTB9.9A)

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 1742 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADEIA: dupla

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genômico)

(vi) FONTE ORIGINAL:

(A) ORGANISMO: *Micobacterium tuberculosis*

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 25:

CCGCTCTCTT	TCAACGTCAT	AAGTTCGGTG	GGCCAGTCGG	CCGCGCGTGC	ATATGGCACC	60
AATAACGCGT	GTCCCATGGA	TACCCGGACC	GCACGACGGT	AGAGCGGATC	AGCGCAGCCG	120
GTGCCGAACA	CTACCGCGTC	CACGCTCAGC	CCTGCCGCGT	TGCGGAAGAT	CGAGCCACAG	180
TTCTCATGGT	CGTTAACGCC	TTCCAACACT	GCGACGGTGC	GCGCCCCGGC	GACCACCTGA	240
GCAACGCTCG	GCTCCGGCAC	CCGGCGCGCG	GCTGCCAACA	CCCCACGATT	GAGATGGAAG	300
CCGATCACCC	GTGCCATGAC	ATCAGCCGAC	GCTCGATAGT	ACGGCGCGCC	GACACCGGCC	360
AGATCATCCT	TGAGCTCGGC	CAGCCGCGCG	TCGGTGCCGA	ACAGCGCCAG	CGGCGTGAAC	420
CGTGAGGCCA	GCATGCGCTG	CACCACCAGC	ACACCTCTCG	CGATCACCAA	CGCCTTGCCG	480
GTCGGCAGAT	CGGGACNACN	GTGATGCTG	TTGAGTTCAC	GGAAATCGTC	GAGCCGTGGG	540
TCGTCGGGAT	CGCAGACGTC	CTGAACATCG	AGGCCGTCGG	GGTGCTGGGC	ACAACGGCCT	600
TCGGTCACGG	GCTTTCGTCG	ACCAGAGCCA	GCATCAGATC	GGCGGCGCTG	CGCAGGATGT	660
CACGCTCGCT	GCGGTTTCAGC	GTGCGAGGCC	GCTCAGCCAG	CCACTCTTGC	AGAGAGCCGT	720
TGCTGGGATT	AATTGGGAGA	GGAGACAGC	ATGTCGTTTC	TGACCACACA	GCCGGAAGCC	780
CTGGCAGCTG	CGGCGGCGAA	CCTACAGGGT	ATTGGCACGA	CAATGAACGC	CCAGAACCGC	840
GCCGCGGCTG	CTCCAACCAC	CGGAGTAGTG	CCCGCAGCCG	CCGATGAAGT	ATCAGCGCTG	900
ACCGCGGCTC	AGTTTGCTGC	GCACGCGCAG	ATGTACCAAA	CGGTCAGCGC	CCAGGCGCGC	960
GCCATTACAG	AAATGTTTCG	GAACACGCTG	GTGGCCAGTT	CTGGCTCATA	CGCGGCCACC	1020
GAGGCGGCCA	ACGCAGCCGC	TGCCGGCTGA	ACGGGCTCGC	ACGAACCTGC	TGAAGGAGAG	1080
GGGGAACATC	CGGAGTTCTC	GGGTCAGGGG	TTGCGCCAGC	GCCCAGCCGA	TTGAGNTATC	1140
GGCGTCCATA	ACAGCAGACG	ATCTAGGCAT	TCAGTACTAA	GGAGACAGGC	AACATGGCCT	1200
CACGTTTAT	GACGGATCCG	CATGCGATGC	GGGACATGGC	GGGCGGTTT	GAGGTGCACG	1260
CCCAGACGGT	GGAGGACGAG	GCTCGCCGGA	TGTGGGCGTC	CGCGCAAAAC	ATTTCCGGTG	1320
CGGGCTGGAG	TGGCATGGCC	GAGGCGACCT	CGTAGACAC	CATGACCTAG	ATGAATCAGG	1380
CGTTTCGCAA	CATCGTGAAC	ATGCTGCACG	GGGTGCGTGA	CGGGCTGGTT	CGCGACGCCA	1440
ACAANTACGA	ACAGCAAGAG	CAGGCCTCCC	AGCAGATCCT	GAGCAGNTAG	CGCCGAAAGC	1500
CACAGCTGNG	TACGNTTCT	CACATTAGGA	GAACACCAAT	ATGACGATTA	ATTACAGTT	1560
CGGGGACGTC	GACGCTCATG	GCGCCATGAT	CCGCGCTCAG	GCGGCGTCGC	TTGAGGCGGA	1620
GCATCAGGCC	ATCGTTCGTG	ATGTGTTGGC	CGCGGGTGAC	TTTTGGGGCG	GCGCCGGTTC	1680
GGTGGCTTGC	CAGGAGTTCA	TTACCCAGTT	GGGCCGTAAC	TTCCAGGTGA	TCTACGAGCA	1740
GG						1742

(2) INFORMAÇÃO PARA SEQ ID N°: 26. MTI (MTB9.9A)

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 2836 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADEIA: dupla

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN (genômico)

(vi) FONTE ORIGINAL:

(A) ORGANISMO: *Micobacterium tuberculosis*

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 26:

GTTGATTCCG	TTCCGCGGCG	CGCCGAAGAC	CACCAACTCC	GCTGGGGTGG	TCCACAGGC	60
GGTTGCGTCG	GTCAGCTGGC	CGAATCCCAA	TGATTGGTGG	CTCNGTGCGG	TTGCTGGGCT	120
CGATTACCCC	CACGGAAAGG	ACGACGATCG	TTCGTTTGCT	CGGTCAGTCG	TACTTGGCGA	180
CGGGCATGGC	GCGGTTTCTT	ACCTCGATCG	CACAGCAGCT	GACCTTCGGC	CCAGGGGGCA	240
CAACGGCTGG	CTCCGGCGGA	GCCTGGTACC	CAACGCCACA	ATTCCGCCGC	CTGGGTGCAG	300
GCCCCGGCGT	GTCGGCGAGT	TTGGCGCGGG	CGGAGCCGGT	CGGGAGGTTG	TGGGTGCCGC	360
CAAGTTGGGC	CGTCCGCGCT	CCGGCCTTCG	CGGAGAAGCC	TGAGGCGGGC	ACGCCGATGT	420
CGTTCATCGG	CGAAGCGTCC	AGCTGCGGTC	AGGGAGGCCT	GCTTCGAGGC	ATACCGCTGG	480
CGAGAGCGGG	GCCGCGTACA	GGCGCCTTCG	CTCACCGATA	CGGGTTCCGC	CACAGCGTGA	540
TTACCCGGTC	TCCGTCGGCG	GGATAGCTTT	CGATCCGGTC	TGCGCGGCCG	CCGGAATGTC	600
TGCAGATAGC	GATCGACCGC	GCCGCTCGGT	AAACGCCGCA	CACGGCACTA	TCAATGCGCA	660
CGGCGGGCGT	TGATGCCAAA	TTGACCGTCC	CGACGGGGCT	TTATCTGCGG	CAAGATTTCA	720
TCCCCAGCCC	GGTCGGTGGG	CCGATAAATA	CGCTGGTCAG	CGCGACTCTT	CCGGCTGAAT	780
TCGATGCTCT	GGGCGCCCCG	TCGACGCCGA	GTATCTCGAG	TGGGCCGCAA	ACCCGGTCAA	840
ACGCTGTTAC	TGTGGCGTTA	CCACAGGTGA	ATTTCGGGTG	CCAACGGGTG	AACACTTGCG	900
AACGGGTGGC	ATCGAAATCA	ACTTGTTCGG	TTGCAGTGAT	CTACTCTCTT	GCAGAGAGCC	960
GTTCGCTGGG	TTAATTGGGA	GAGGAAGACA	GCATGTCGTT	CGTGACCACA	CAGCCGGAAG	1020
CCCTGSCAGC	TTCGGCGGGC	AACCTACAGG	GTATTGGCAC	GACAATGAAC	GCCCAGAACG	1080
CGGCCGCGGC	TGCTCCAACC	ACCGGAGTAG	TGCCCCGAGC	CGCCGATGAA	GTATCAGCGC	1140
TGACCGCGGC	TCAGTTTGCT	GCGCACGCGC	AGATGTACCA	AACGGTCAGC	GCCCAGGCCG	1200
CGGCCATTCA	CGAAATGTTT	GTGAACACGC	TGGTGGCCAG	TTCTGGCTCA	TACGCGGCCA	1260
CCGAGSGCGC	CAACGCAGCC	GCTGCGGGCT	GAACGGGCTC	GCACGAACCT	GCTGAAGGAG	1320
AGGGGGAACA	TCCGGAGTTC	TGGGTTCAGG	GGTTGCCGCA	GCGCCGAGCC	GATTCAGCTA	1380
TCGGCTTCCA	TAAACAGCAG	CGATCTAGGC	ATTCACTACT	AAGGAGACAG	GCAACATGGC	1440
CTCAGTPTT	ATGACGGATC	CGCATGCGAT	GCGGAGCATG	GCGGGCCGTT	TTGAGGTGCA	1500
CGCCCGAGCG	GTGGAGGACG	AGGCTCGCCG	GATGTGGGCG	TCCGCGCAAA	ACATTTCCGG	1560
TGCGGGCTGG	AGTGGCATGG	CCGAGGCGAC	CTCGCTAGAC	ACCATGACCT	AGATGAATCA	1620
GGCGTTTCGC	AACATCGTGA	ACATGCTGCA	CGGGGTGCGT	GACGGGCTGG	TTGCGGACGC	1680
CAACAACCTAC	GAACAGCAAG	AGCAGGCCTC	CCAGCAGATC	CTGAGCAGCT	AGCGCCGAAA	1740
GCCACAGCTG	CGTACGCTTT	CTCACATTAG	GAGAACACCA	ATATGACGAT	TAATTACCAG	1800
TTCCGGGAGC	TCGACGCTCA	TGGCGCCATG	ATCCGCGCTC	AGGCGGGGTC	GCTTGAGGCG	1860

```

GAGCATCAGG CCATCGTTTCG TGATGTGTTG GCCGCGGGTG ACTTTTGGGG CGGCGCCGGT 1920
TCGGTGGCTT GCCAGGAGTT CAFTACCCAG TTGGGCGGTA ACTTCCAGGT GATCTACGAG 1980
CAGGCCAAGC CCCACGGGCA GAAGGTGCAG GCTGCCGGCA ACAACATGGC GCAAACCGAC 2040
AGCGCGGTGC GCTCCAGCTG GGCCTAAAC TGAACCTCAG TCGCGGCAGC ACACCAACCA 2100
GCCGGTGTGC TGCTGTGTCC TGCAGTTAAC TAGCACTCGA CCGCTGAGGT AGCGATGGAT 2160
CAACAGAGTA CCCGCACCGA CATCACCGTC AACGTCGACG GCTTCTGGAT GCTTCAGGCG 2220
CTACTGGATA TCCGCCACGT TCGCCCTGAG TTAGCTTGCC GCGCGTACGT CTCCACCGAT 2280
TCCAATGACT GGCTAAACGA GCACCCGGGG ATGGCGGTCA TGCSCGAGCA GGGCATTGTC 2340
GTCAACGACG CGGTCAACGA ACAGGTGCGT GCCCGGATGA AGGTGCTTGC CGCACCTGAT 2400
CTTGAAGTCG TCGCCCTGCT GTCACGCGGC AAGTTGCTGT ACGGGGTCAT AGACGACGAG 2460
AACCAGCCGC CGGGTTCCGC TGACATCCCT GACAATGAGT TCCGGGTGGT GTTGGCCCGG 2520
CGAGGCCAGC ACTGGGTGTC GGCGGTACGG GTTGGCAATG ACATCACCGT CGATGACGTG 2580
ACGGTCTCGG ATAGCGCCTC GATCGCCGCA CTGGTAATGG ACGGTCTGGA GTCGATTAC 2640
CACGCCGACC CAGCCGCGAT CAACGCGGTC AACGTGCCAA TGGAGGAGAT CTCGTGCCGA 2700
ATTCCGGCAG AGGCACGAGG CGGTGTCCGT GACGACGGGA TCGATCAGCA TCATCGACCG 2760
GCCCGGATCC TTGGCGATCT CGTTGAGCAC GACCCGGGCC CGCGGGAAGC TCTGCGACAT 2820
CCATGGGTTC TTCCCG 2836

```

(2) INFORMAÇÃO PARA SEQ ID N°: 27: MTI (MTB9.9A)

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 94 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: péptido

(vi) FONTE ORIGINAL:

- (A) ORGANISMO: *Micobacterium tuberculosis*

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 27:

```

Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
1           5           10           15
Ile Arg Ala Leu Ala Gly Leu Leu Glu Ala Glu His Gln Ala Ile Ile
20           25           30
Ser Asp Val Leu Thr Ala Ser Asp Phe Trp Gly Gly Ala Gly Ser Ala
35           40           45
Ala Cys Gln Gly Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
50           55           60
Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn
65           70           75           80
Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
85           90

```

(2) INFORMAÇÃO PARA SEQ ID N°: 28: HTCC#1

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1200 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 28:

CAGGCATGAG	CAGAGCGTTC	ATCATCGATC	CAACGATCAG	TGCCATTGAC	GGCTTGACG	60
ACCTTCTGGG	GATTGGAATA	CCCAACCAAG	GGGGTATCCT	TTACTCCTCA	CTAGAGTACT	120
TCGAAAAAGC	CCTGGAGGAG	CTGGCAGCAG	CGTTTCCGGG	TGATGGCTGG	TTAGGTTCCG	180
CCGCGGACAA	ATACGCCGGC	AAAAACCGCA	ACCACGTGAA	TTTTTTCCAG	GAAGTGGCAG	240
ACCTCGATCG	TCAGCTCATC	AGCCTGATCC	ACGACCAGGC	CAACGCGGTC	CAGACGACCC	300
GCGACATCCT	GGAGGGCGCC	AAGAAAGGTC	TCGAGTTCGT	GCGCCCGGTG	GCTGTGGACC	360
TGACCTACAT	CCCGTTCGTC	GGGCACGCC	TATCGGCCGC	CTTCCAGGCG	CCGTTTTGCG	420
CGGGCGCGAT	GGCCGTAGTG	GGCGGCGCGC	TTGCCTACTT	GGTCGTGAAA	ACGCTGATCA	480
ACGCGACTCA	ACTCCTCAAA	TTGCTTGCCA	AATTGGCGGA	GTTGGTCGCG	GCCGCCATTG	540
CGGACATCAT	TTCCGATGTG	GCGGACATCA	TCAAGGGCAC	CCTCGGAGAA	GTGTGGGAGT	600
TCATCACAAA	CGCGTCAAC	GGCCTGAAAG	AGCTTTGGGA	CAAGCTCACG	GGGTGGGTGA	660
CCGGACTGTT	CTCTCGAGGG	TGGTCGAACC	TGGAGTCCTT	CTTTGCGGGC	GTCCCCGGCT	720
TGACCGGCGC	GACCAGCGGC	TTGTCGCAAG	TGACTGGCTT	GTTCGGTGCG	GCCGGTCTGT	780
CCGCATCGTC	GGGCTTGGCT	CACGCGGATA	GCCTGGCGAG	CTCAGCCAGC	TTGCCCGCCC	840
TGGCCGGCAT	TGGGGGCGGG	TCCGGTTTGT	GGGGCTTGCC	GAGCCTGGCT	CAGGTCCATG	900
CCGCCTCAAC	TCGGCAGGCG	CTACGGCCCC	GAGCTGATGG	CCCGGTCGGC	GCCGCTGCCG	960
AGCAGGTCGG	CGGGCAGTCG	CAGCTGGTCT	CCGCGCAGGG	TTCCCAAGGT	ATGGGCGGAC	1020
CCGTAGGCAT	GGGCGGCATG	CACCCCTCTT	CGGGGGCGTC	GAAAGGGACG	ACGACGAAGA	1080
AGTACTCGGA	AGGCGCGGCG	GCGGGCACTG	AAGACGCCGA	GCGCGCGCCA	GTCGAAGCTG	1140
ACGCGGGCGG	TGGGCAAAAG	GTGCTGGTAC	GAAACGTCGT	CTAACCGCAT	GGCGAGCCAA	1200

(2) INFORMAÇÃO PARA SEQ ID N°: 29: HTCC#1

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 392 aminoácidos

(B) TIPO: aminoácido

(C) TIPO DE CADEIA: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 29:

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

Ser	Ala	Ser	Leu	Pro	Ala	Leu	Ala	Gly	Ile	Gly	Gly	Gly	Ser	Gly	Phe
		275					280					285			
Gly	Gly	Leu	Pro	Ser	Leu	Ala	Gln	Val	His	Ala	Ala	Ser	Thr	Arg	Gln
	290					295					300				
Ala	Leu	Arg	Pro	Arg	Ala	Asp	Gly	Pro	Val	Gly	Ala	Ala	Ala	Glu	Gln
305					310					315					320
Val	Gly	Gly	Gln	Ser	Gln	Leu	Val	Ser	Ala	Gln	Gly	Ser	Gln	Gly	Met
			325						330					335	
Gly	Gly	Pro	Val	Gly	Met	Gly	Gly	Met	His	Pro	Ser	Ser	Gly	Ala	Ser
			340					345					350		
Lys	Gly	Thr	Thr	Thr	Lys	Lys	Tyr	Ser	Glu	Gly	Ala	Ala	Ala	Gly	Thr
	355						360					365			
Glu	Asp	Ala	Glu	Arg	Ala	Pro	Val	Glu	Ala	Asp	Ala	Gly	Gly	Gly	Gln
	370					375					380				
Lys	Val	Leu	Val	Arg	Asn	Val	Val								
385					390										

(2) INFORMAÇÃO PARA SEQ ID N°: 30: MTCC#2

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1441 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNC

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 30:

GAGGTTGCTG	GCAATGGATT	TCGGGCTTTT	ACCTCCGGAA	GTGAATTCAA	GCCGAATGTA	60
TTCCGGTCCG	GGGCCGGAGT	CGATGCTAGC	CGCCGCGGCC	GCCTGGGACG	GTGTGGCCGC	120
GGAGTTGACT	TCCGCCGCGG	TCTCGTATGG	ATCGGTGGTG	TCGACGCTGA	TCGTTGAGCC	180
GTGGATGGGG	CCGGCGGGCG	CCGCGATGGC	GGCCGCGGCA	ACGCCGTATG	TGGGGTGGCT	240
GGCATCTGCG	GCCGCGTCCG	CGTTGCCGCC	GTTCACTCCA	CCCGTGCAAG	GCACCGGCCC	540
GGCCGGGCCC	GCGGCCGCG	CCGCGGCGAC	CCAAGCCGCC	GGTGCGGGCG	CCGTTGCGGA	600
TGCACAGGCG	ACACTGGCCC	AGCTGCCCCC	GGGGATCCTG	AGCGACATTC	TGTCCGCATT	660
GGCCGCCAAC	GCTGATCCGC	TGACATCGSG	ACTGTTGGGG	ATCGCGTCGA	CCCTCAACCC	720
GCAAGTCGGA	TCCGCTCAGC	CGATAGTGAT	CCCCACCCCG	ATAGGGGAAT	TGGACGTGAT	780
CGCGCTCTAC	ATTGCATCCA	TCGCGACCGG	CAGCATTGCG	CTCGCGATCA	CGAACACGGC	840
CAGACCCTGG	CACATCGGCC	TATACGGGAA	CGCCGGCGGG	CTGGGACCGA	CGCAGGGCCA	900
TCCACTGAGT	TCGGCGACCG	ACGAGCCGGA	GCCGCACTGG	GGCCCCCTCG	GGGGCGCGGC	960
GCCGGTGTC	GCGGGCGTCG	GCCACGCAGC	ATTAGTCGGA	GCGTTGTCCG	TGCCGCACAG	1020
CTGGACCACG	GCCGCCCCCG	AGATCCAGCT	CGCCGTTTCG	GCAACACCCA	CCTTCAGCTC	1080
CAGCGCCGGC	GCCGACCCGA	CGGCCCTAAA	CGGGATGCCG	GCAGGCCTGC	TCAGCGGGAT	1140
GGCTTTGGCG	AGCCTGGCCG	CACGCGGCAC	GACGGGCGGT	GGCGGCACCC	GTAGCGGCAC	1200
CAGCACTGAC	GGCCAAGAGG	ACGGCCGCAA	ACCCCCGGTA	GTTGTGATTA	GAGAGCAGCC	1260
GCCGCCCGGA	AACCCCCCGC	GGTAAAAGTC	CGGCAACCGT	TCGTCCCGCG	GCGGAAAATG	1320
CCTGGTGAGC	GTGGCTATCC	GACGGGCCGT	TCACACCGCT	TGTAGTAGCG	TACGGCTATG	1380
GACGACGGTG	TCTGGATTCT	CGGCGGCTAT	CAGAGCGATT	TTGCTCGCAA	CCTCAGCAAA	1440
G						1441

(2) INFORMAÇÃO PARA SEQ ID N°: 31: MTCC#2

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 423 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID Nº: 31:

Met	Asp	Phe	Gly	Leu	Leu	Pro	Pro	Glu	Val	Asn	Ser	Ser	Arg	Met	Tyr	1	5	10	15
Ser	Gly	Pro	Gly	Pro	Glu	Ser	Met	Leu	Ala	Ala	Ala	Ala	Ala	Trp	Asp	20	25	30	
Gly	Val	Ala	Ala	Glu	Leu	Thr	Ser	Ala	Ala	Val	Ser	Tyr	Gly	Ser	Val	35	40	45	
Val	Ser	Thr	Leu	Ile	Val	Glu	Pro	Trp	Met	Gly	Pro	Ala	Ala	Ala	Ala	50	55	60	
Met	Ala	Ala	Ala	Ala	Thr	Pro	Tyr	Val	Gly	Trp	Leu	Ala	Ala	Thr	Ala	65	70	75	80
Ala	Leu	Ala	Lys	Glu	Thr	Ala	Thr	Gln	Ala	Arg	Ala	Ala	Ala	Glu	Ala	85	90	95	
Phe	Gly	Thr	Ala	Phe	Ala	Met	Thr	Val	Pro	Pro	Ser	Leu	Val	Ala	Ala	100	105	110	
Asn	Arg	Ser	Arg	Leu	Met	Ser	Leu	Val	Ala	Ala	Asn	Ile	Leu	Gly	Gln	115	120	125	
Asn	Ser	Ala	Ala	Ile	Ala	Ala	Thr	Gln	Ala	Glu	Tyr	Ala	Glu	Met	Trp	130	135	140	
Ala	Gln	Asp	Ala	Ala	Val	Met	Tyr	Ser	Tyr	Glu	Gly	Ala	Ser	Ala	Ala	145	150	155	160
Ala	Ser	Ala	Leu	Pro	Pro	Phe	Thr	Pro	Pro	Val	Gln	Gly	Thr	Gly	Pro	165	170	175	
Ala	Gly	Pro	Ala	Ala	Ala	Ala	Ala	Ala	Thr	Gln	Ala	Ala	Gly	Ala	Gly	180	185	190	
Ala	Val	Ala	Asp	Ala	Gln	Ala	Thr	Leu	Ala	Gln	Leu	Pro	Pro	Gly	Ile	195	200	205	
Leu	Ser	Asp	Ile	Leu	Ser	Ala	Leu	Ala	Ala	Asn	Ala	Asp	Pro	Leu	Thr	210	215	220	
Ser	Gly	Leu	Leu	Gly	Ile	Ala	Ser	Thr	Leu	Asn	Pro	Gln	Val	Gly	Ser	225	230	235	240
Ala	Gln	Pro	Ile	Val	Ile	Pro	Thr	Pro	Ile	Gly	Glu	Leu	Asp	Val	Ile	245	250	255	
Ala	Leu	Tyr	Ile	Ala	Ser	Ile	Ala	Thr	Gly	Ser	Ile	Ala	Leu	Ala	Ile	260	265	270	
Thr	Asn	Thr	Ala	Arg	Pro	Trp	His	Ile	Gly	Leu	Tyr	Gly	Asn	Ala	Gly	275	280	285	
Gly	Leu	Gly	Pro	Thr	Gln	Gly	His	Pro	Leu	Ser	Ser	Ala	Thr	Asp	Glu	290	295	300	
Pro	Glu	Pro	His	Trp	Gly	Pro	Phe	Gly	Gly	Ala	Ala	Pro	Val	Ser	Ala	305	310	315	320
Gly	Val	Gly	His	Ala	Ala	Leu	Val	Gly	Ala	Leu	Ser	Val	Pro	His	Ser	325	330	335	
Trp	Thr	Thr	Ala	Ala	Pro	Glu	Ile	Gln	Leu	Ala	Val	Gln	Ala	Thr	Pro	340	345	350	
Thr	Phe	Ser	Ser	Ser	Ala	Gly	Ala	Asp	Pro	Thr	Ala	Leu	Asn	Gly	Met	355	360	365	
Pro	Ala	Gly	Leu	Leu	Ser	Gly	Met	Ala	Leu	Ala	Ser	Leu	Ala	Ala	Arg	370	375	380	
Gly	Thr	Thr	Gly	Gly	Gly	Gly	Thr	Arg	Ser	Gly	Thr	Ser	Thr	Asp	Gly	385	390	395	400
Gln	Glu	Asp	Gly	Arg	Lys	Pro	Pro	Val	Val	Val	Ile	Arg	Glu	Gln	Pro	405	410	415	
Pro	Pro	Gly	Asn	Pro	Pro	Arg										420			

(2) INFORMAÇÃO PARA SEQ ID N°: 32: ESAT-6

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 154 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 32:

ATGACAGAGC	AGCAGTGGAA	TTTCGCGGGT	ATCGAGGCCG	CGGCAAGCGC	AATCCAGGGA	60
AATGTCACGT	CCATTCATTC	CCTCCTTGAC	GAGGGGAAGC	AGTCCCTGAC	CAAGCTCGCA	120
GCGGCCTGGG	GCGGTAGCGG	TTCGGAAGCG	TACC			154

(2) INFORMAÇÃO PARA SEQ ID N°: 33: ESAT-6

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 51 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 33:

Met	Thr	Glu	Gln	Gln	Trp	Asn	Phe	Ala	Gly	Ile	Glu	Ala	Ala	Ala	Ser	
1				5					10					15		
Ala	Ile	Gln	Gly	Asn	Val	Thr	Ser	Ile	His	Ser	Leu	Leu	Asp	Glu	Gly	
			20					25					30			
Lys	Gln	Ser	Leu	Thr	Lys	Leu	Ala	Ala	Ala	Trp	Gly	Gly	Ser	Gly	Ser	
			35				40					45				
Glu	Ala	Tyr														
			50													

(2) INFORMAÇÃO PARA SEQ ID N°: 34: Tb38-1

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 327 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 34:

CGGCACGAGA	GACCGATGCC	GCTACCCTCG	CGCAGGAGGC	AGGTAATTTC	GAGCGGATCT	60
CCGGCGACCT	GAAAACCCAG	ATCGACCAGG	TGGAGTCGAC	GGCAGGTTTC	TTGCAGGGCC	120
AGTGGCGCGG	CGCGGCGGGG	ACGGCCGCCC	AGGCCGCGGT	GGTGCCTTC	CAAGAAGCAG	180
CCAATAAGCA	GAAGCAGGAA	CTCGACGAGA	TCTCGACGAA	TATTTCGTCAG	GCCGGCGTCC	240
AATACTCGAG	GGCCGACGAG	GAGCAGCAGC	AGGCGCTGTC	CTCGCAAATG	GGCTTCTGAC	300
CCGCTAATAC	GAAAAGAAAC	GGAGCAA				327

(2) INFORMAÇÃO PARA SEQ ID N°: 35: Tb38-1

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 95 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 35:

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

(2) INFORMAÇÃO PARA SEQ ID N°: 36: TbRa3

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 542 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 36:

GAATTCGGCA	CGAGAGGTGA	TCGACATCAT	CGGGACCAGC	CCCACATCCT	GGGAACAGGC	60
GGCGGCGGAG	GCGGTCCAGC	GGGCGCGGGA	TAGCGTCGAT	GACATCCGCG	TCGCTCGGGT	120
CATTGAGCAG	GACATGGCCG	TGGACAGCGC	CGGCAAGATC	ACCTACCGCA	TCAAGCTCGA	180
AGTGTCGTTT	AAGATGAGGC	CGGCGCAACC	GCGCTAGCAC	GGGCCGGCGA	GCAAGACGCA	240
AAATCGCACG	GTTTGCGGTT	GATTCGTGCG	ATTTTGTGTC	TGCTCGCCGA	GGCCTACCAG	300
GCGCGGCCCA	GGTCCGCGTG	CTGCCGTATC	CAGGCGTGCA	TCGCGATTCC	GGCGGCCACG	360
CCGGAGTTAA	TGCTTCGCGT	CGACCCGAAC	TGGGCGATCC	GCCGGNGAGC	TGATCGATGA	420
CCGTGGCCAG	CCCGTCGATG	CCCGAGTTGC	CCGAGGAAAC	GTGCTGCCAG	GCCGGTAGGA	480
AGCGTCCGTA	GGCGGCGGTG	CTGACCGGCT	CTGCCTGCGC	CCTCAGTGCG	GCCAGCGAGC	540
GG						542

(2) INFORMAÇÃO PARA SEQ ID N°: 37: TbRa3

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 66 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 37:

```
val ile asp ile ile gly thr ser pro thr ser trp glu gln ala ala
1          5          10          15
ala glu ala val gln arg ala arg asp ser val asp asp ile arg val
20          25          30
ala arg val ile glu gln asp met ala val asp ser ala gly lys ile
35          40          45
thr tyr arg ile lys leu glu val ser phe lys met arg pro ala gln
50          55          60
pro arg
65
```

(2) INFORMAÇÃO PARA SEQ ID N°: 38: 38 kD

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1993 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID Nº: 38:

TGTTCTTCGA	CGGCAGGCTG	GTGGAGGAAG	GGCCCACCGA	ACAGCTGTTT	TCCTCGCCGA	60
AGCATGCGGA	AACCGCCCCG	TACGTGCGCG	GACTGTGCGG	GGACGTCAAG	GACGCCAAGC	120
GCGGAAATTG	AAGAGCACAG	AAAGGTATGG	CGTGAAAATT	CGTTTGCATA	CGCTGTTGGC	180
CGTGTGACC	GCTGCGCCGC	TGCTGCTAGC	AGCGGCGGGC	TGTGGCTCGA	AACCACCGAG	240
CGGTCGCTT	GAAACGGGCG	CCGGCGCCCG	TACTGTGCGG	ACTACCCCCG	CGTCGTGCGC	300
GGTGACGTTG	GCGGAGACCG	GTAGCACGCT	GCTCTACCCG	CTGTTCAACC	TGTGGGGTCC	360
GGCCTTTTAC	GAGAGGTATC	CGAACGTCAC	GATCACCGCT	CAGGGCACCG	GTTCTGGTGC	420
CGGGATCGCG	CAGGCCGCCG	CCGGGACGGT	CAACATTGGG	GCCTCCGACG	CCTATCTGTC	480
GGAAGGTGAT	ATGGCCGCGC	ACAAGGGGGT	GATGAACATC	GCGCTAGCCA	TCTCCGCTCA	540
GCAGGTCAAC	TACAACCTGC	CCGGAGTGAG	CGAGCACCTC	AAGCTGAACG	GAAAAGTCCT	600
GGCGGCCATG	TACCAGGGCA	CCATCAAAAC	CTGGGACGAC	CCGCAGATCG	CTGCGCTCAA	660
CCCCGGCGTG	AACCTGCCCC	GCACCGCGGT	AGTTCCGCTG	CACCGCTCCG	ACGGGTCCGG	720
TGACACCTTC	TTGTTCACCC	AGTACCTGTC	CAAGCAAGAT	CCCGAGGGCT	GGGGCAAGTC	780
GCCCCGGCTT	GGCACCACCG	TCGACTTCCC	GGCGGTGCCG	GGTGCCTGGG	GTGAGAACGG	840
CAACGGCGGC	ATGGTGACCG	GTTGCGCCGA	GACACCGGGC	TGCGTGGCCT	ATATCGGCAT	900
CAGCTTCCTC	GACCAGGCCA	GTCAACGGGG	ACTCGGCGAG	GCCCAACTAG	GCAATAGCTC	960
TGGCAATTTT	TTGTTGCCCC	ACGCGCAAAG	CATTGAGGCC	GCGGCGGCTG	GCTTCGCATC	1020
GAAAACCCCG	GCGAACCAGG	CGATTTCGAT	GATCGACGGG	CCCGCCCCCG	ACGGGTACCC	1080
GATCATCAAC	TACGAGTACG	CCATCGTCAA	CAACCGGCAA	AAGGACGCCG	CCACCGCGCA	1140
GACCTTGCGG	GCATTTCTGC	ACTGGGCGAT	CACCGACGGC	AACAAGGCCT	CGTTCCTCGA	1200
CCAGGTTTAT	TTCCAGCCCG	TGCCGCCCGC	GGTGGTGAAG	TTGTCTGACG	CGTTGATCGC	1260
GACGATTTTC	AGCTAGCCTC	GTTGACCACC	ACGCGACAGC	AACCTCCGTC	GGGCCATCGG	1320
GCTGCTTTGC	GGAGCATGCT	GGCCCGTGCC	GGTGAAGTCG	GCCGCGCTGG	CCCGGCCATC	1380
CGGTGGTTGG	GTGGGATAGG	TGCGGTGATC	CCGCTGCTTG	CGCTGGTCTT	GGTGCTGGTG	1440
GTGCTGGTCA	TCGAGGCGAT	GGGTGCGATC	AGGCTCAACG	GGTTGCATTT	CTTCACCGCC	1500

ACCGAATGGA ATCCAGGCAA CACCTACGGC GAAACCGTTG TCACCGACGC GTCGCCCATC	1560
CGGTCGGCGC CTA CTACTACGGG GCGTTGCCGC TGATCGTCCG GACGCTGGCG ACCTCGGCAA	1620
TGCCCCTGAT CATCGCGGTG CCGGTCTCTG TAGGAGCGGC GCTGGTGATC GTGGAACGGC	1680
TGCCGAAACG GTTGGCCGAG GCTGTGGGAA TAGTCCTGGA ATTGCTCGCC GGAATCCCCA	1740
GCGTGGTTCGT CGGTTTGTGG GGGGCAATGA CGTTCGGGCC GTTCATCGCT CATCACATCG	1800
CTCCGGTGAT CGCTCACAAC GCTCCCGATG TGCCGGTGCT GAACTACTTG CGCGGCGACC	1860
CGGGCAACGG GGAGGGCATG TTGGTGTCOG GTCTGGTGTT GCGGGTGATG GTCGTTCCCA	1920
TTATCGCCAC CACCACTCAT GACCTGTTCC GGCAGGTGCC GGTGTTGCCC CGGGAGGGCG	1980
CGATCGGGAA TTC	1993

(2) INFORMAÇÃO PARA SEQ ID N°: 39: 38 kD

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 374 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA:
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 39:

Met	Lys	Ile	Arg	Leu	His	Thr	Leu	Leu	Ala	Val	Leu	Thr	Ala	Ala	Pro
1				5					10					15	
Leu	Leu	Leu	Ala	Ala	Ala	Gly	Cys	Gly	Ser	Lys	Pro	Pro	Ser	Gly	Ser
			20					25					30		
Pro	Glu	Thr	Gly	Ala	Gly	Ala	Gly	Thr	Val	Ala	Thr	Thr	Pro	Ala	Ser
		35					40					45			
Ser	Pro	Val	Thr	Leu	Ala	Glu	Thr	Gly	Ser	Thr	Leu	Leu	Tyr	Pro	Leu
	50					55					60				

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

Ile Ala Thr Ile Ser Ser
370

(2) INFORMAÇÃO PARA SEQ ID N°: 40: DPEP

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 999 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 40:

ATGCATCACC	ATCACCATCA	CATGCATCAG	GTGGACCCCA	ACTTGACACG	TCGCAAGGGA	60
CGATTGGCGG	CACTGGCTAT	CGCGGCGATG	GCCAGCGCCA	GCCTGGTGAC	CGTTGCGGTG	120
CCCGCGAÇCG	CCAACGCCGA	TCCGGAGCCA	GCGCCCCCGG	TACCCACAAC	GGCCGCCTCG	180
CCGCGGTCTGA	CCGCTGCAGC	GCCACCCGCA	CCGGCGACAC	CTGTTGCCCC	CCCACCACCG	240
GCCGCCGCCA	ACACGCCGAA	TGCCCAGCCG	GGCGATCCCA	ACGCAGCACC	TCCGCCGGCC	300
GACCCGAACG	CACCGCCGCC	ACCTGTCATT	GCCCCAAACG	CACCCCAACC	TGTCCGGATC	360
GACAACCCGG	TTGGAGGATT	CAGCTTCGCG	CTGCCTGCTG	GCTGGGTGGA	GTCTGACGCC	420
GCCCACTTCG	ACTACGGTTC	AGCACTCCTC	AGCAAAACCA	CCGGGGACCC	GCCATTTCCC	480
GGACAGCCGC	CGCCGGTGGC	CAATGACACC	CGTATCGTGC	TCGGCCGGCT	AGACCAAAAG	540
CTTTACGCCA	GCGCCGAAGC	CACCGACTCC	AAGGCCGCGG	CCCGGTTGGG	CTCGGACATG	600
GGTGAGTTCT	ATATGCCCTA	CCCGGGCACC	CGGATCAACC	AGGAAACCGT	CTCGCTCGAC	660
GCCAACGGGG	TGTCTGGAAG	CGCGTCGTAT	TACGAAGTCA	AGTTCAGCGA	TCCGAGTAAG	720
CCGAACGGCC	AGATCTGGAC	GGGCGTAATC	GGCTCGCCCG	CGGCGAACGC	ACCGGACGCC	780
GGGCCCCCTC	AGCGCTGGTT	TGTGGTATGG	CTCGGGACCG	CCAACAACCC	GGTGGACAAG	840
GGCGCGGCCA	AGGCGCTGGC	CGAATCGATC	CGGCCTTTGG	TCGCCCCGCC	GCCGGCGCCG	900
GCACCGGCTC	CTGCAGAGCC	CGCTCCGGCG	CCGGCGCCCG	CCGGGGAAGT	CGCTCCTACC	960
CCGACGACAC	CGACACCGCA	GCGGACCTTA	CCGGCCTGA			999

(2) INFORMAÇÃO PARA SEQ ID N°: 41: DPEP

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 332 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 41:

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

(2) INFORMAÇÃO PARA SEQ ID N°: 42: TbH4

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 702 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 42:

CGGCACGAGG	ATCGGTACCC	CGCGGCATCG	GCAGCTGCCG	ATTCGCCGGG	TTTCCCCACC	60
CGAGGAAAGC	CGCTACCAGA	TGGCGCTGCC	GAAGTAGGGC	GATCCGTTCC	CGATGCCGGC	120
ATGAACGGGC	GGCATCAAAT	TAGTGCAGGA	ACCTTTCAGT	TTAGCGACGA	TAATGGCTAT	180
AGCACTAAGG	AGGATGATCC	GATATGACGC	AGTCGCAGAC	CGTGACGGTG	GATCAGCAAG	240
AGATTTTGAA	CAGGGCCAAC	GAGGTGGAGG	CCCCGATGGC	GGACCCACCG	ACTGATGTCC	300
CCATCACACC	GTGCGAACTC	ACGGNGGNTA	AAAACGCCGC	CCAACAGNTG	GTNTTGTCCG	360
CCGACAACAT	GCGGGAATAC	CTGGCGGCCG	GTGCCAAAGA	GCGGCAGCGT	CTGGCGACCT	420
CGCTGCGCAA	CGCGGCCAAG	GNGTATGGCG	AGGTTGATGA	GGAGGCTGCG	ACCGCGCTGG	480
ACAACGACGG	CGAAGGAAC	GTGCAGGCAG	AATCGGCCGG	GGCCGTCGGA	GGGGACAGTT	540
CGGCCGAAC	AACCGATACG	CCGAGGGTGG	CCACGGCCGG	TGAACCCAAC	TTTATGGATC	600
TCAAAGAAGC	GGCAAGGAAG	CTCGAAACGG	GCGACCAAGG	CGCATCGCTC	GCGCACTGNG	660
GGGATGGGTG	GAACACTTNC	ACCCTGACGC	TGCAAGGCGA	CG		702

(2) INFORMAÇÃO PARA SEQ ID N°: 43: TbH4

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 286 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 43:

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

(2) INFORMAÇÃO PARA SEQ ID N°: 44: DPPD

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 339 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADN Genómico

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 44:

```
ATGAAGTTGA AGTTTGCTCG CCTGAGTACT GCGATACTGG GTTGTGCAGC GGCGCTTGTG      60
TTTCCTGCCT CGGTTGCCAG CGCAGATCCA CCTGACCCGC ATCAGCCGGA CATGACGAAA      120
GGCTATTGCC CGGGTGGCCG ATGGGGTTTT GGCGACTTGG CCGTGTGCGA CGGCGAGAAG      180
TACCCCGACG GCTCGTTTTG GCACCACTGG ATGCAAACGT GGTTTACCGG CCCACAGTTT      240
TACTTCGATT GTGTCAGCGG CGGTGAGCCC CTCCCCGGCC CGCCGCCACC GGGTGGTTGC      300
GGTGGGGCAA TTCCGTCCGA GCAGCCCAAC GCTCCCTGA      339
```

(2) INFORMAÇÃO PARA SEQ ID N°: 45: DPPD

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 112 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 45:

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

<210> SEQ ID N°: 46

<211> 921

<212> ADN

<213> Sequência Artificial

<223> Descrição da Sequência Artificial: proteína de tri-fusão
DPV-MTI-MSL (designada Mtb31f)

<222> (1)..(900)

cat	atg	cat	cac	cat	cac	cat	cac	gat	ccc	gtg	gac	gcg	gtc	att	aac	48
His	Met	His	His	His	His	His	His	Asp	Pro	Val	Asp	Ala	Val	Ile	Asn	
1				5				10					15			
acc	acc	tgc	aat	tac	ggg	cag	gta	gta	gct	gcg	ctc	aac	gcg	acg	gat	96
Thr	Thr	Cys	Asn	Tyr	Gly	Gln	Val	Val	Ala	Ala	Leu	Asn	Ala	Thr	Asp	
			20					25					30			
ccg	ggg	gct	gcc	gca	cag	ttc	aac	gcc	tca	ccg	gtg	gcg	cag	tcc	tat	144
Pro	Gly	Ala	Ala	Ala	Gln	Phe	Asn	Ala	Ser	Pro	Val	Ala	Gln	Ser	Tyr	
		35					40					45				
ttg	cgc	aat	ttc	ctc	gcc	gca	ccg	cca	cct	cag	cgc	gct	gcc	atg	gcc	192
Leu	Arg	Asn	Phe	Leu	Ala	Ala	Pro	Pro	Pro	Gln	Arg	Ala	Ala	Met	Ala	
	50					55					60					
gcg	caa	ttg	caa	gct	gtg	ccg	ggg	gcg	gca	cag	tac	atc	ggc	ctt	gtc	240
Ala	Gln	Leu	Gln	Ala	Val	Pro	Gly	Ala	Ala	Gln	Tyr	Ile	Gly	Leu	Val	
	65				70				75						80	
gag	tcg	gtt	gcc	ggc	tcc	tgc	aac	aac	tat	gag	ctc	atg	acg	att	aat	288
Glu	Ser	Val	Ala	Gly	Ser	Cys	Asn	Asn	Tyr	Glu	Leu	Met	Thr	Ile	Asn	
				85				90						95		
tac	cag	ttc	ggg	gac	gtc	gac	gct	cat	ggc	gcc	atg	atc	cgc	gct	cag	336
Tyr	Gln	Phe	Gly	Asp	Val	Asp	Ala	His	Gly	Ala	Met	Ile	Arg	Ala	Gln	
			100					105					110			

gcg gcg tcg ctt gag gcg gag cat cag gcc atc gtt cgt gat gtg ttg	384
Ala Ala Ser Leu Glu Ala Glu His Gln Ala Ile Val Arg Asp Val Leu	
115 120 125	
gcc gcg ggt gac ttt tgg ggc ggc gcc ggt tcg gtg gct tgc cag gag	432
Ala Ala Gly Asp Phe Trp Gly Gly Ala Gly Ser Val Ala Cys Gln Glu	
130 135 140	
ttc att acc cag ttg ggc cgt aac ttc cag gtg atc tac gag cag gcc	480
Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile Tyr Glu Gln Ala	
145 150 155 160	
aac gcc cac ggg cag aag gtg cag gct gcc ggc aac aac atg gcg caa	528
Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn Asn Met Ala Gln	
165 170 175	
acc gac agc gcc gtc ggc tcc agc tgg gcc act agt atg agc ctt ttg	576
Thr Asp Ser Ala Val Gly Ser Ser Trp Ala Thr Ser Met Ser Leu Leu	
180 185 190	
gat gct cat atc cca cag ttg gtg gcc tcc cag tcg gcg ttt gcc gcc	624
Asp Ala His Ile Pro Gln Leu Val Ala Ser Gln Ser Ala Phe Ala Ala	
195 200 205	
aag gcg ggg ctg atg cgg cac acg atc ggt cag gcc gag cag gcg gcg	672
Lys Ala Gly Leu Met Arg His Thr Ile Gly Gln Ala Glu Gln Ala Ala	
210 215 220	
atg tcg gct cag gcg ttt cac cag ggg gag tcg tcg gcg gcg ttt cag	720
Met Ser Ala Gln Ala Phe His Gln Gly Glu Ser Ser Ala Ala Phe Gln	
225 230 235 240	
gcc gcc cat gcc cgg ttt gtg gcg gcg gcc gcc aaa gtc aac acc ttg	768
Ala Ala His Ala Arg Phe Val Ala Ala Ala Ala Lys Val Asn Thr Leu	
245 250 255	
ttg gat gtc gcg cag gcg aat ctg ggt gag gcc gcc ggt acc tat gtg	816
Leu Asp Val Ala Gln Ala Asn Leu Gly Glu Ala Ala Gly Thr Tyr Val	
260 265 270	
gcc gcc gat gct gcg gcc gcg tcg acc tat acc ggg ttc gat atc cat	864
Ala Ala Asp Ala Ala Ala Ala Ser Thr Tyr Thr Gly Phe Asp Ile His	
275 280 285	
cac act ggc ggc cgc tcg agc aga tcc ggc tgc taacaaagcc cgaaaggaag	917
His Thr Gly Gly Arg Ser Ser Arg Ser Gly Cys	
290 295	
ctga	921

<210> SEQ ID N°: 47

<211> 299

<212> PRT

<213> Sequência Artificial

<223> Descrição da Sequência Artificial: proteína de tri-fusão
DPV-MTI-MSL (designada Mtb31f)

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

<222> (1) .. (2133)

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

atg	tcg	gct	cag	gcg	ttt	cac	cag	ggg	gag	tcg	tcg	gcg	gcg	ttt	cag	720
Met	Ser	Ala	Gln	Ala	Phe	His	Gln	Gly	Glu	Ser	Ser	Ala	Ala	Phe	Gln	
225					230					235					240	
gcc	gcc	cat	gcc	cgg	ttt	gtg	gcg	gcg	gcc	gcc	aaa	gtc	aac	acc	ttg	768
Ala	Ala	His	Ala	Arg	Phe	Val	Ala	Ala	Ala	Ala	Lys	Val	Asn	Thr	Leu	
				245					250					255		
ttg	gat	gtc	gcg	cag	gcg	aat	ctg	ggg	gag	gcc	gcc	ggg	acc	tat	gtg	816
Leu	Asp	Val	Ala	Gln	Ala	Asn	Leu	Gly	Glu	Ala	Ala	Gly	Thr	Tyr	Val	
			260					265					270			
gcc	gcc	gat	gct	gcg	gcc	gcg	tcg	acc	tat	acc	ggg	ttc	gat	atc	atg	864
Ala	Ala	Asp	Ala	Ala	Ala	Ala	Ser	Thr	Tyr	Thr	Gly	Phe	Asp	Ile	Met	
		275					280					285				
gat	ttc	ggg	ctt	tta	cct	ccg	gaa	gtg	aat	tca	agc	cga	atg	tat	tcc	912
Asp	Phe	Gly	Leu	Leu	Pro	Pro	Glu	Val	Asn	Ser	Ser	Arg	Met	Tyr	Ser	
	290					295					300					
ggg	ccg	ggg	ccg	gag	tcg	atg	cta	gcc	gcc	gcg	gcc	gcc	tgg	gac	ggg	960
Gly	Pro	Gly	Pro	Glu	Ser	Met	Leu	Ala	Ala	Ala	Ala	Ala	Trp	Asp	Gly	
305					310					315					320	
gtg	gcc	gcg	gag	ttg	act	tcc	gcc	gcg	gtc	tcg	tat	gga	tcg	gtg	gtg	1008
Val	Ala	Ala	Glu	Leu	Thr	Ser	Ala	Ala	Val	Ser	Tyr	Gly	Ser	Val	Val	
				325					330					335		
tcg	acg	ctg	atc	gtt	gag	ccg	tgg	atg	ggg	ccg	gcg	gcg	gcc	gcg	atg	1056
Ser	Thr	Leu	Ile	Val	Glu	Pro	Trp	Met	Gly	Pro	Ala	Ala	Ala	Ala	Met	
			340					345					350			
gcg	gcc	gcg	gca	acg	ccg	tat	gtg	ggg	tgg	ctg	gcc	gcc	acg	gcg	gcg	1104
Ala	Ala	Ala	Ala	Thr	Pro	Tyr	Val	Gly	Trp	Leu	Ala	Ala	Thr	Ala	Ala	
			355				360						365			
ctg	gcg	aag	gag	acg	gcc	aca	cag	gcg	agg	gca	gcg	gcg	gaa	gcg	ttt	1152
Leu	Ala	Lys	Glu	Thr	Ala	Thr	Gln	Ala	Arg	Ala	Ala	Ala	Glu	Ala	Phe	
			370				375					380				
ggg	acg	gcg	ttc	gcg	atg	acg	gtg	cca	cca	tcc	ctc	gtc	gcg	gcc	aac	1200
Gly	Thr	Ala	Phe	Ala	Met	Thr	Val	Pro	Pro	Ser	Leu	Val	Ala	Ala	Asn	
385					390					395					400	
cgc	agc	cgg	ttg	atg	tcg	ctg	gtc	gcg	gcg	aac	att	ctg	ggg	caa	aac	1248
Arg	Ser	Arg	Leu	Met	Ser	Leu	Val	Ala	Ala	Asn	Ile	Leu	Gly	Gln	Asn	
				405					410					415		
agt	gcg	gcg	atc	gcg	gct	acc	cag	gcc	gag	tat	gcc	gaa	atg	tgg	gcc	1296
Ser	Ala	Ala	Ile	Ala	Ala	Thr	Gln	Ala	Glu	Tyr	Ala	Glu	Met	Trp	Ala	
			420					425					430			
caa	gac	gct	gcc	gtg	atg	tac	agc	tat	gag	ggg	gca	tct	gcg	gcc	gcg	1344
Gln	Asp	Ala	Ala	Val	Met	Tyr	Ser	Tyr	Glu	Gly	Ala	Ser	Ala	Ala	Ala	
		435					440					445				

tog gcg ttg ccg ccg ttc act cca ccc gtg caa ggc acc-ggc ccg gcc	1392
Ser Ala Leu Pro Pro Phe Thr Pro Pro Val Gln Gly Thr Gly Pro Ala	
450 455 460	
ggg ccc gcg gcc gca gcc gcg gcg acc caa gcc gcc ggt gcg ggc gcc	1440
Gly Pro Ala Ala Ala Ala Ala Ala Thr Gln Ala Ala Gly Ala Gly Ala	
465 470 475 480	
gtt gcg gat gca cag gcg aca ctg gcc cag ctg ccc ccg ggg atc ctg	1488
Val Ala Asp Ala Gln Ala Thr Leu Ala Gln Leu Pro Pro Gly Ile Leu	
485 490 495	
agc gac att ctg tcc gca ttg gcc gcc aac gct gat ccg ctg aca tcc	1536
Ser Asp Ile Leu Ser Ala Leu Ala Ala Asn Ala Asp Pro Leu Thr Ser	
500 505 510	
gga ctg ttg ggg atc gcg tcc acc ctc aac ccg caa gtc gga tcc gct	1584
Gly Leu Leu Gly Ile Ala Ser Thr Leu Asn Pro Gln Val Gly Ser Ala	
515 520 525	
cag ccg ata gtg atc ccc acc ccg ata ggg gaa ttg gac gtg atc gcg	1632
Gln Pro Ile Val Ile Pro Thr Pro Ile Gly Glu Leu Asp Val Ile Ala	
530 535 540	
ctc tac att gca tcc atc gcg acc ggc agc att gcg ctc gcg atc acg	1680
Leu Tyr Ile Ala Ser Ile Ala Thr Gly Ser Ile Ala Leu Ala Ile Thr	
545 550 555 560	
aac acg gcc aga ccc tgg cac atc ggc cta tac ggg aac gcc ggc ggg	1728
Asn Thr Ala Arg Pro Trp His Ile Gly Leu Tyr Gly Asn Ala Gly Gly	
565 570 575	
ctg gga ccg acg cag ggc cat cca ctg agt tcc gcg acc gac gag ccg	1776
Leu Gly Pro Thr Gln Gly His Pro Leu Ser Ser Ala Thr Asp Glu Pro	
580 585 590	
gag ccg cac tgg ggc ccc ttc ggg ggc gcg gcg ccg gtg tcc gcg ggc	1824
Glu Pro His Trp Gly Pro Phe Gly Gly Ala Ala Pro Val Ser Ala Gly	
595 600 605	
gtc ggc cac gca gca tta gtc gga gcg ttg tcc gtg ccg cac agc tgg	1872
Val Gly His Ala Ala Leu Val Gly Ala Leu Ser Val Pro His Ser Trp	
610 615 620	
acc acg gcc gcc ccg gag atc cag ctc gcc gtt cag gca aca ccc acc	1920
Thr Thr Ala Ala Pro Glu Ile Gln Leu Ala Val Gln Ala Thr Pro Thr	
625 630 635 640	
ttc agc tcc agc gcc ggc gcc gac ccg acg gcc cta aac ggg atg ccg	1968
Phe Ser Ser Ser Ala Gly Ala Asp Pro Thr Ala Leu Asn Gly Met Pro	
645 650 655	
gca ggc ctg ctc agc ggg atg gct ttg gcg agc ctg gcc gca cgc ggc	2016
Ala Gly Leu Leu Ser Gly Met Ala Leu Ala Ser Leu Ala Ala Arg Gly	
660 665 670	
acg acg ggc ggt ggc ggc acc cgt agc ggc acc agc act gac ggc caa	2064
Thr Thr Gly Gly Gly Gly Thr Arg Ser Gly Thr Ser Thr Asp Gly Gln	
675 680 685	

gag gac ggc cgc aaa ccc ccg gta gtt gtg att aga gag cag ccg ccg 2112
 Glu Asp Gly Arg Lys Pro Pro Val Val Val Ile Arg Glu Gln Pro Pro
 690 695 700

ccc gga aac ccc ccg cgg taagatttct aaatccatca cactggcggc cgctcgag 2168
 Pro Gly Asn Pro Pro Arg
 705 710

<210> SEQ ID N°: 49

<211> 710

<212> PRT

<213> Sequência Artificial

<223> Descrição da Sequência Artificial: proteína de tetra-fusão
 DPV-MTI-MSL-MTCC2 (designada Mtb71f)

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

Thr Asp Ser Ala Val Gly Ser Ser Trp Ala Thr Ser Met Ser Leu Leu
 180 185 190
 Asp Ala His Ile Pro Gln Leu Val Ala Ser Gln Ser Ala Phe Ala Ala
 195 200 205
 Lys Ala Gly Leu Met Arg His Thr Ile Gly Gln Ala Glu Gln Ala Ala
 210 215 220
 Met Ser Ala Gln Ala Phe His Gln Gly Glu Ser Ser Ala Ala Phe Gln
 225 230 235 240
 Ala Ala His Ala Arg Phe Val Ala Ala Ala Lys Val Asn Thr Leu
 245 250 255
 Leu Asp Val Ala Gln Ala Asn Leu Gly Glu Ala Ala Gly Thr Tyr Val
 260 265 270
 Ala Ala Asp Ala Ala Ala Ala Ser Thr Tyr Thr Gly Phe Asp Ile Met
 275 280 285
 Asp Phe Gly Leu Leu Pro Pro Glu Val Asn Ser Ser Arg Met Tyr Ser
 290 295 300
 Gly Pro Gly Pro Glu Ser Met Leu Ala Ala Ala Ala Ala Trp Asp Gly
 305 310 315 320
 Val Ala Ala Glu Leu Thr Ser Ala Ala Val Ser Tyr Gly Ser Val Val
 325 330 335
 Ser Thr Leu Ile Val Glu Pro Trp Met Gly Pro Ala Ala Ala Ala Met
 340 345 350
 Ala Ala Ala Ala Thr Pro Tyr Val Gly Trp Leu Ala Ala Thr Ala Ala
 355 360 365
 Leu Ala Lys Glu Thr Ala Thr Gln Ala Arg Ala Ala Ala Glu Ala Phe
 370 375 380
 Gly Thr Ala Phe Ala Met Thr Val Pro Pro Ser Leu Val Ala Ala Asn
 385 390 395 400
 Arg Ser Arg Leu Met Ser Leu Val Ala Ala Asn Ile Leu Gly Gln Asn
 405 410 415
 Ser Ala Ala Ile Ala Ala Thr Gln Ala Glu Tyr Ala Glu Met Trp Ala
 420 425 430
 Gln Asp Ala Ala Val Met Tyr Ser Tyr Glu Gly Ala Ser Ala Ala Ala
 435 440 445
 Ser Ala Leu Pro Pro Phe Thr Pro Pro Val Gln Gly Thr Gly Pro Ala
 450 455 460

Gly Pro Ala Ala Ala Ala Ala Thr Gln Ala Ala Gly Ala Gly Ala
 465 470 475 480
 Val Ala Asp Ala Gln Ala Thr Leu Ala Gln Leu Pro Pro Gly Ile Leu
 485 490 495
 Ser Asp Ile Leu Ser Ala Leu Ala Ala Asn Ala Asp Pro Leu Thr Ser
 500 505 510
 Gly Leu Leu Gly Ile Ala Ser Thr Leu Asn Pro Gln Val Gly Ser Ala
 515 520 525
 Gln Pro Ile Val Ile Pro Thr Pro Ile Gly Glu Leu Asp Val Ile Ala
 530 535 540
 Leu Tyr Ile Ala Ser Ile Ala Thr Gly Ser Ile Ala Leu Ala Ile Thr
 545 550 555 560
 Asn Thr Ala Arg Pro Trp His Ile Gly Leu Tyr Gly Asn Ala Gly Gly
 565 570 575
 Leu Gly Pro Thr Gln Gly His Pro Leu Ser Ser Ala Thr Asp Glu Pro
 580 585 590
 Glu Pro His Trp Gly Pro Phe Gly Gly Ala Ala Pro Val Ser Ala Gly
 595 600 605
 Val Gly His Ala Ala Leu Val Gly Ala Leu Ser Val Pro His Ser Trp
 610 615 620
 Thr Thr Ala Ala Pro Glu Ile Gln Leu Ala Val Gln Ala Thr Pro Thr
 625 630 635 640
 Phe Ser Ser Ser Ala Gly Ala Asp Pro Thr Ala Leu Asn Gly Met Pro
 645 650 655
 Ala Gly Leu Leu Ser Gly Met Ala Leu Ala Ser Leu Ala Ala Arg Gly
 660 665 670
 Thr Thr Gly Gly Gly Gly Thr Arg Ser Gly Thr Ser Thr Asp Gly Gln
 675 680 685
 Glu Asp Gly Arg Lys Pro Pro Val Val Val Ile Arg Glu Gln Pro Pro
 690 695 700
 Pro Gly Asn Pro Pro Arg
 705 710

Lisboa, 30 de Outubro de 2009

REIVINDICAÇÕES

1. Polipéptido de fusão compreendendo (i) um antígeno MTB39 (SEQ ID N° 12 ou 14), uma variante do referido antígeno MTB39 tendo, pelo menos, 90% de identidade com este ou um fragmento imunogénico do referido antígeno MTB39 e (ii) um antígeno MTB32A (SEQ ID N° 2 ou 4), uma variante do referido antígeno MTB32A tendo, pelo menos, 90% de identidade com este ou um fragmento imunogénico do referido antígeno MTB32A, caracterizado por o referido antígeno MTB32A, variante ou fragmento terem uma substituição de aminoácidos correspondente à posição 183 da SEQ ID N° 4 ou à posição de aminoácido 208 da SEQ ID N° 2, em que o resíduo de serina normalmente encontrado nesta posição foi substituído por outro aminoácido.
2. Polipéptido de fusão da reivindicação 1, compreendendo, pelo menos, 195 aminoácidos do terminal N de um antígeno MTB32A (SEQ ID N° 2 ou 4) ou uma variante do referido antígeno MTB32A tendo, pelo menos, 90% de identidade com este.
3. Polipéptido de fusão de acordo com as reivindicações 1 ou 2, em que o resíduo de serina correspondente à posição de aminoácido 183 da SEQ ID N° 4 ou à posição de aminoácido 208 da SEQ ID N° 2 foi substituído por alanina.
4. Polipéptido de fusão de qualquer uma das reivindicações 1 a 3, compreendendo adicionalmente um polipéptido compreendendo, pelo menos, cerca de 132 aminoácidos do terminal C do antígeno MTB32A (SEQ ID N° 2 ou 4) ou uma

variante do referido antigénio MTB32A tendo, pelo menos, 90% de identidade com este.

5. Polipéptido de fusão de qualquer uma das reivindicações 1 a 4, em que os antígenos estão covalentemente ligados através de um elemento de ligação químico.
6. Polipéptido de fusão da reivindicação 5, em que o elemento de ligação químico é um elemento de ligação aminoacídico.
7. Polipéptido de fusão da reivindicação 6, em que o polipéptido de fusão compreende a sequência de aminoácidos de MTB72FMutSA (SEQ ID N° 18).
8. Polipéptido de fusão da reivindicação 7, compreendendo adicionalmente, pelo menos, um polipéptido heterólogo adicional de *Mycobacterium* sp.
9. Polipéptido de fusão da reivindicação 8, em que o antígeno adicional é seleccionado do grupo consistindo em antígeno MTB8.4 (SEQ ID N° 22), antígeno MTB9.8 (SEQ ID N° 24), antígeno MTB9.9 (SEQ ID N°27), antígeno MTB40 (SEQ ID N°29), antígeno MTB41 (SEQ ID N° 31), 38-1 (SEQ ID N° 35), TbRa3 (SEQ ID N°37), 38 kD (SEQ ID N°39), DPEP (SEQ ID N° 41), TbH4 (SEQ ID N° 43), DPPD (SEQ ID N°45), Erd14, antígeno ESAT-6 (SEQ ID N°33), antígeno do complexo MTB85 ou antígeno α -cristalino ou um seu fragmento imunogénico.
10. Polipéptido de fusão da reivindicação 7, em que o polipéptido de fusão tem a sequência de aminoácidos de MTB72FMutSA (SEQ ID N° 18).

11. Composição de vacina compreendendo um polipéptido de fusão de acordo com qualquer uma das reivindicações 1 a 10, compreendendo adicionalmente um adjuvante.
12. Composição de vacina da reivindicação 11, em que o adjuvante compreende QS21 e MPL.
13. Composição de vacina da reivindicação 11, em que o adjuvante é seleccionado do grupo consistindo em AS2, ENHANZYN, MPL, 3D-MPL, IFA, QS21, CWS, TDM, AGP, CPG, Leif, saponina e miméticos da saponina.
14. Polinucleótido codificando um polipéptido de fusão de acordo com qualquer uma das reivindicações 1 a 10.
15. Polinucleótido de acordo com a reivindicação 14, o qual codifica MTB72FMutSA (SEQ ID N° 18).
16. Polinucleótido de acordo com a reivindicação 15, tendo a sequência apresentada na SEQ ID N° 17.
17. Composição de vacina compreendendo um polinucleótido de acordo com qualquer uma das reivindicações 14 a 16, compreendendo adicionalmente um adjuvante.
18. Polipéptido de fusão de acordo com qualquer uma das reivindicações 1 a 10, para utilização como um medicamento.
19. Polinucleótido de acordo com qualquer uma das reivindicações 14 a 16, para utilização como um medicamento.

20. Utilização de um polipéptido de fusão de acordo com qualquer uma das reivindicações 1 a 10 na preparação de um medicamento para promover uma resposta imunitária num mamífero.
21. Utilização de um polinucleótido de acordo com qualquer uma das reivindicações 14 a 16 na preparação de um medicamento para promover uma resposta imunitária num mamífero.
22. Método para a produção de uma proteína de fusão de acordo com qualquer uma das reivindicações 1 a 10, em que o método compreende a utilização de uma sequência polinucleotídica que codifica a referida proteína de fusão para controlar a expressão numa célula hospedeira.

Lisboa, 30 de Outubro de 2009

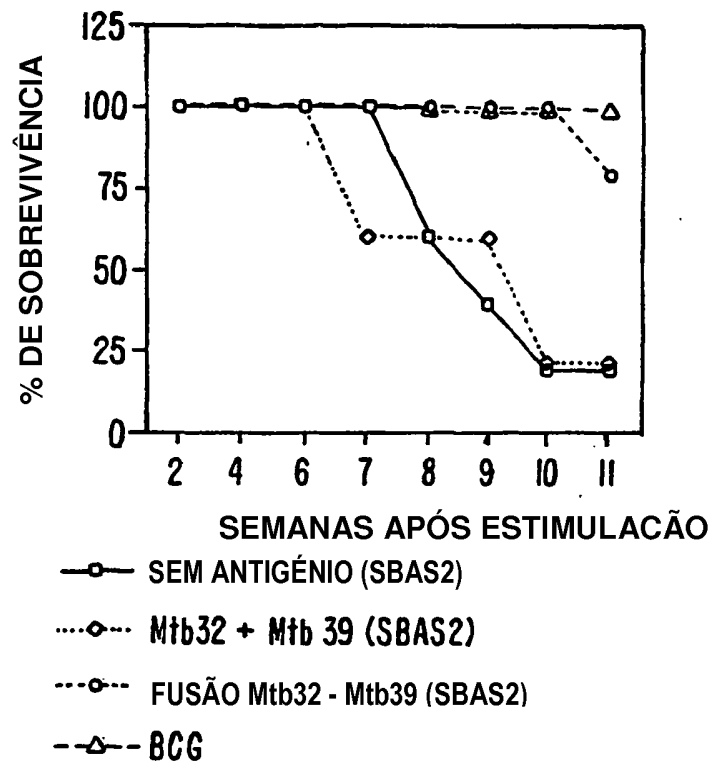
**FIG. 1.**

FIG. 2A.

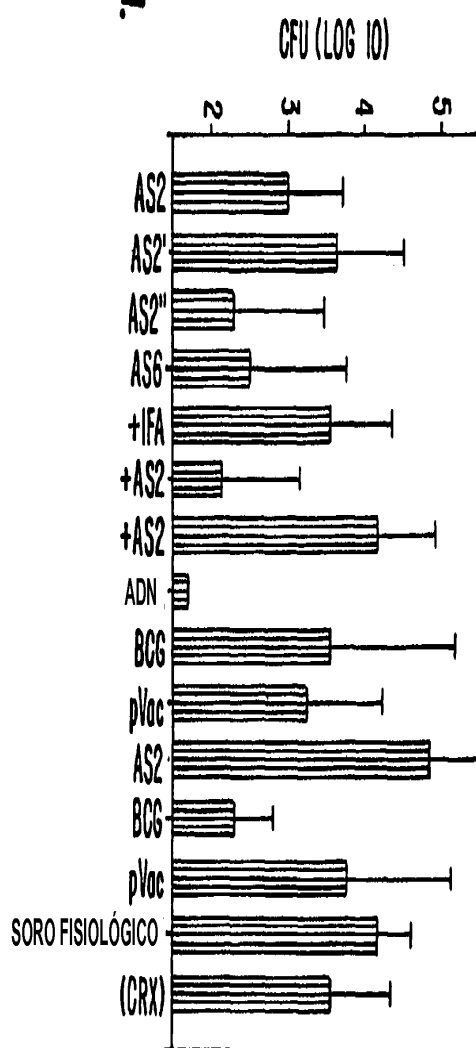
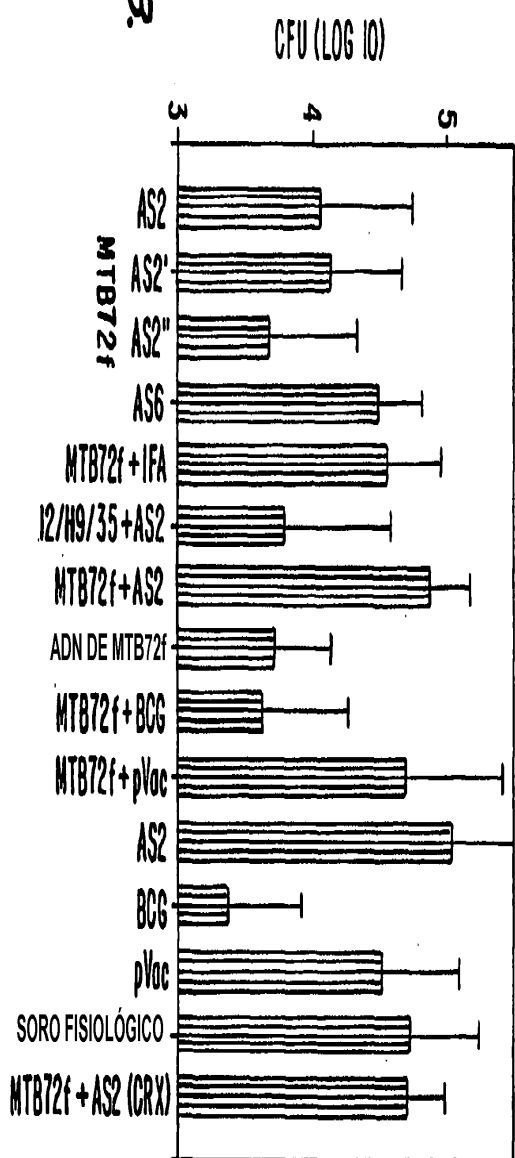
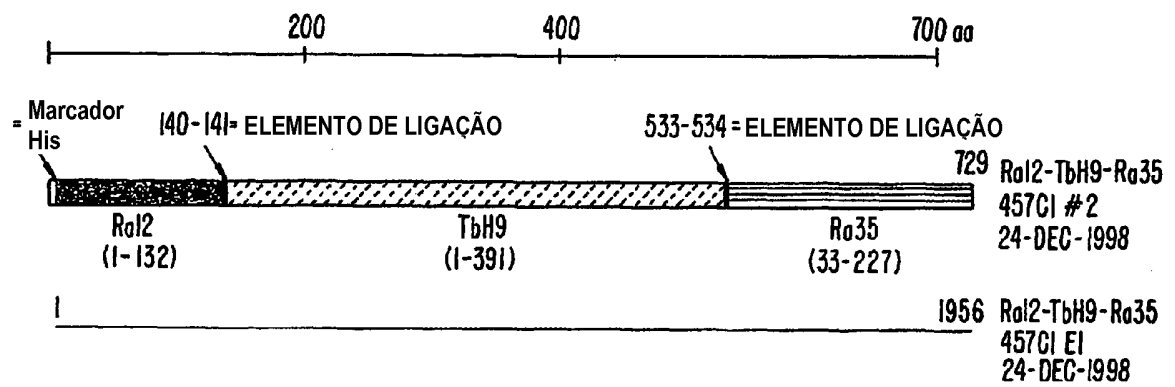


FIG. 2B.



**FIG. 3.**

ADN do terminal N de Ra35

```

gccccgccgg ccttgctgca ggaccgggtc gccgacttcc ccgcgctgcc cctcgaccgc tccgcgatgg 70
tcgccaagt ggggccacag gtggtcaaca tcaacaccaa actgggctac aacaacgcc tgggcgccgg 140
gaccggcatc gtcctgatc ccaacgggtg cgtgctgacc aacaaccacg tgatcgccgg cgccaccgac 210
atcaatgcgt tcagcgtcgg ctccggccaa acctacggcg tcgatgtggt cgggtatgac cgcaccagg 280
atgtcgccgt gctgcagctg ccggtgccc gtggcctacc atcggcgccg atcgggtggc gcgtcgccgt 350
tggtgagccc gtcgtcgca tgggcaacag cgggtggcag ggcggaacgc cccgtgccgt gcctggcagg 420
gtggtcgccg tcggccaaac cgtgcaggcg tcggattcgc tgaccggtgc cgaagagaca ttgaacgggt 490
tgatccagtt cgatgccgag atccagcccg gtgattcggg cgggcccgtc gtcaacggcc taggacaggt 560
ggtcggtatg aacacggccg cgtcctag 588

```

Sequência de aminoácidos do terminal N de Ra35

```

Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala Leu Pro Leu Asp Pro Ser Ala
      5              10              15              20

Met Val Ala Gln Val Gly Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val
      25              30              35              40

Gly Ala Gly Thr Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val Ile Ala
      45              50              55              60              65

Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln Thr Tyr Gly Val Asp Val Val Gly
      70              75              80              85

Tyr Asp Arg Thr Gln Asp Val Ala Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala
      90              95              100              105              110

```

FIG. 4.

Ile	Gly	Gly	Gly	Val	Ala	Val	Gly	Glu	Pro	Val	Val	Ala	Met	Gly	Asn	Ser	Gly	Gly	Gln	Gly	Gly
				115					120					125					130		
Thr	Pro	Arg	Ala	Val	Pro	Gly	Arg	Val	Val	Ala	Leu	Gly	Gln	Thr	Val	Gln	Ala	Ser	Asp	Ser	Leu
		135					140					145					150				
Thr	Gly	Ala	Glu	Glu	Thr	Leu	Asn	Gly	Leu	Ile	Gln	Phe	Asp	Ala	Ala	Ile	Gln	Pro	Gly	Asp	Ser
155					160					165						170				175	
Gly	Gly	Pro	Val	Val	Asn	Gly	Leu	Gly	Gln	Val	Val	Gly	Met	Asn	Thr	Ala	Ala	Ser			
			180					185					190					195			

FIG. 4. (CONTINUAÇÃO)

Ra12

```

1  MHHHHHHHTAASDNFQLSQGGQGFAIPIGQAMAIAGQIRSGGGSPTVHIGPTAFLG Mtb72f
1  MHHHHHHHTAASDNFQLSQGGQGFAIPIGQAMAIAGQIRSGGGSPTVHIGPTAFLG Mtb72f-mutSA

56 LGVVDNNGNGARVQRVVGSAPAASLGISTGDVITAVDGAPINSATAMADALNGHH Mtb72f
56 LGVVDNNGNGARVQRVVGSAPAASLGISTGDVITAVDGAPINSATAMADALNGHH Mtb72f-mutSA

                                TbH9FL
111 PGDVISVTWQTKSFFTRTFNVTLAEGPPAEFFMVDGALPPEINSARMYAGPGSAS Mtb72f
111 PGDVISVTWQTKSFFTRTFNVTLAEGPPAEFFMVDGALPPEINSARMYAGPGSAS Mtb72f-mutSA

166 LVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSSAGLMVAAASPYVAWMSV Mtb72f
166 LVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSSAGLMVAAASPYVAWMSV Mtb72f-mutSA

221 TAGQAELTAAQVRVAAAAYETAYGLTVPPPVIENRAELMILIATNLLGQNTPAI Mtb72f
221 TAGQAELTAAQVRVAAAAYETAYGLTVPPPVIENRAELMILIATNLLGQNTPAI Mtb72f-mutSA

276 AVNEAEYGEWMAQDAAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVE Mtb72f
276 AVNEAEYGEWMAQDAAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVE Mtb72f-mutSA

331 EASDTAAANQLMNNVPQALQQLAQPTQGTTTPSSKLGGWLKTVSPHRSPISNMVSM Mtb72f
331 EASDTAAANQLMNNVPQALQQLAQPTQGTTTPSSKLGGWLKTVSPHRSPISNMVSM Mtb72f-mutSA

386 ANNHMSMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSLGSS Mtb72f
386 ANNHMSMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSLGSS Mtb72f-mutSA

441 GLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAERGPQM Mtb72f
441 GLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAERGPQM Mtb72f-mutSA

```

FIG. 5.

496 LGGLPVGMGARAGGGLSGVLRVPPRPYVMPHSPAAGDIAPPALSQDRFADFPAL Mtb72f
 496 LGGLPVGMGARAGGGLSGVLRVPPRPYVMPHSPAAGDIAPPALSQDRFADFPAL Mtb72f-mutSA
 551 PLDPSAMVAQVGPQVVNINTKLGYNNAVAGAGTGIVDPNGVVLTNNNVIAGATDI Mtb72f
 551 PLDPSAMVAQVGPQVVNINTKLGYNNAVAGAGTGIVDPNGVVLTNNHVIAGATDI Mtb72f-mutSA
 606 NAFSVGSGQTYGVDVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMGN Mtb72f
 606 NAFSVGSGQTYGVDVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMGN Mtb72f-mutSA
 661 SGGQGGTPRAVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVV Mtb72f
 661 SGGQGGTPRAVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVV Mtb72f-mutSA
 716 NGLGQVVMNTAAS Mtb72f
 716 NGLGQVVMNTAAS Mtb72f-mutSA

FIG. 5. (CONTINUAÇÃO)

```

Ra35 N-term
1  MHHHHHH[APPALSQDRFADFPALPLDPSAMVAQVGPOVVNINTKLGYNNA TbRa35_mat
1  MHHHHHH[APPALSQDRFADFPALPLDPSAMVAQVGPOVVNINTKLGYNNA TbRa35 mutSA

51  VGAGTGIVIDPNGVVLTTNNHVIAGATDINAFSVGSGQTYGVDVVGYDRTQ TbRa35_mat
51  VGAGTGIVIDPNGVVLTTNNHVIAGATDINAFSVGSGQTYGVDVVGYDRTQ TbRa35 mutSA

101 DVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMGNSGGQGGTPRAVPGRVVA TbRa35_mat
101 DVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMGNSGGQGGTPRAVPGRVVA TbRa35 mutSA

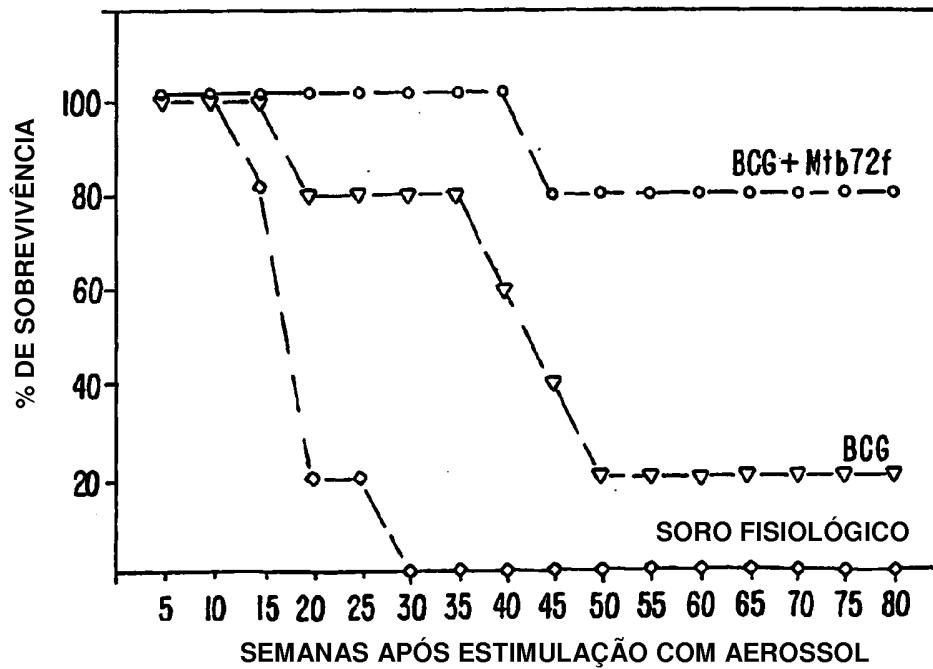
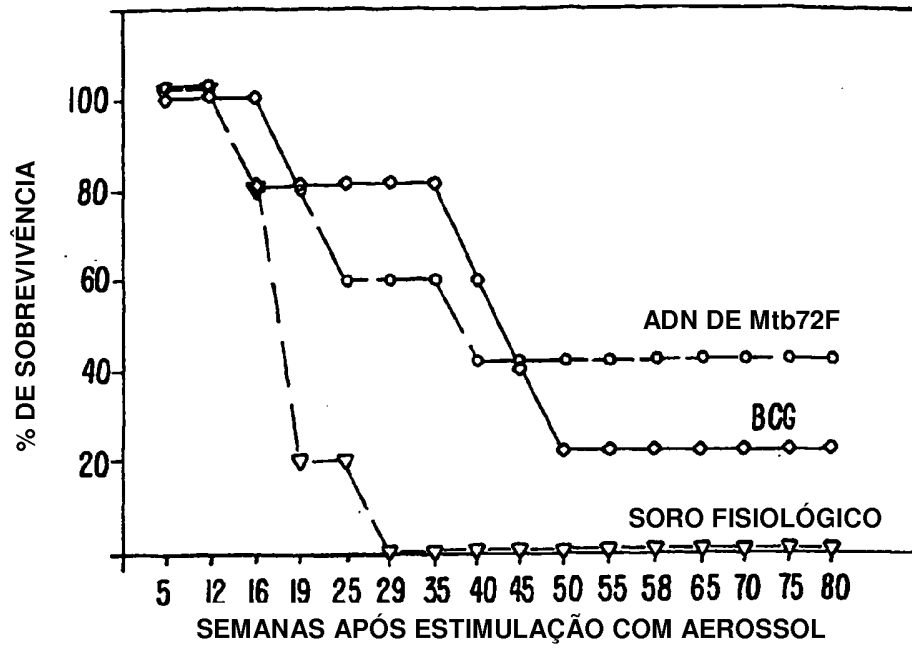
151 LGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNLGQVVGMM[TA TbRa35_mat
151 LGQTVQASDSLTAETLNGLIQFDAAIQPGD[AGGPVVNLGQVVGMM[TA TbRa35 mutSA
end Ra35 Nterm
201 ASDNFQLSQGGQGFAIPIGQAMAIAGQIRSGGGSPTVHIGPTAFLGLGVV TbRa35_mat
201 ASDNFQLSQGGQGFAIPIGQAMAIAGQIRSGGGSPTVHIGPTAFLGLGVV TbRa35 mutSA

251 DNNGNGARVQRVVGSAPAASLGISTGDVITAVDGAPINSATAMADALNGH TbRa35_mat
251 DNNGNGARVQRVVGSAPAASLGISTGDVITAVDGAPINSATAMADALNGH TbRa35 mutSA

301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA] end TbRa35_mat
301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA] Ra12 TbRa35 mutSA

```

FIG. 6.

**FIG. 7.**

RESUMO

"PROTEÍNAS DE FUSÃO DE *MYCOBACTERIUM TUBERCULOSIS*"

A presente invenção refere-se a composições e proteínas de fusão contendo, pelo menos, dois antigénios de *Mycobacterium* sp. e ácidos nucleicos codificando essas composições e proteínas de fusão. As composições da invenção aumentam a sensibilidade serológica de soros de indivíduos infectados com tuberculose e métodos para a sua utilização em diagnóstico, tratamento e prevenção da infecção com tuberculose.