

(11) Número de Publicação: **PT 1511768 E**

(51) Classificação Internacional:

C07K 14/315 (2006.01) **A61K 39/385** (2006.01)
C12N 5/10 (2006.01) **C12N 15/62** (2006.01)
C12N 15/63 (2006.01) **A61K 38/16** (2006.01)

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

(22) Data de pedido: **2003.06.06**

(30) Prioridade(s): **2002.06.11 GB 0213365**
2003.01.15 GB 0300914

(43) Data de publicação do pedido: **2005.03.09**

(45) Data e BPI da concessão: **2007.04.11**
037/2007

(73) Titular(es):

GLAXO GROUP LIMITED
GLAXO WELLCOME HOUSE BERKELEY
AVENUE GREENFORD MIDDLESEX UB6 ONN
GB
GLAXOSMITHKLINE BIOLOGICALS S.A. **BE**

(72) Inventor(es):

CABEZON SILVA **BE**
JONATHAN H. ELLIS **GB**
C.M.G. GERARD **BE**
PAUL A. HAMBLIN **GB**
R.M. PALMANTIER **BE**

(74) Mandatário:

PEDRO DA SILVA ALVES MOREIRA
RUA DO PATROCÍNIO, N.º 94 1399-019 LISBOA **PT**

(54) Epígrafe: **COMPOSIÇÕES IMUNOGÉNICAS**

(57) Resumo:

RESUMO

"COMPOSIÇÕES IMUNOGÉNICAS"

A presente invenção refere-se a parceiros de fusão que actuam como parceiros de fusão imunológicos, como intensificadores de expressão, e de um modo preferido, a parceiros de fusão que têm ambas as funções. Em particular, os parceiros de fusão contêm um denominado domínio de ligação à colina, por exemplo, fusões compreendendo LytA de *Streptococcus pneumoniae* ou a lisozima CP1 de fago pneumocócico (CPL1), em que o domínio de ligação à colina é modificado para incluir um epitopo heterólogo de auxiliar T, e estão fundidos a antígenos, particularmente antígenos pouco imunogénicos, tais como auto-antígenos, e. g., antígenos específicos de tumor ou específicos de tecido. A invenção também se refere a proteínas de fusão contendo as mesmas, à sua produção, à sua utilização em composições imunogénicas e de vacina e à sua utilização em medicamentos.

DESCRIÇÃO

"COMPOSIÇÕES IMUNOGÉNICAS"

A presente invenção refere-se a parceiros de fusão que actuam como parceiros de fusão imunológicos, como intensificadores de expressão e, de um modo preferido, a parceiros de fusão que têm ambas as funções. A invenção também se refere a proteínas de fusão que os contêm, à sua produção, à sua utilização em vacinas e à sua utilização em medicamentos. Em particular, são proporcionados parceiros de fusão que contêm um denominado domínio de ligação à colina, por exemplo fusões compreendendo LytA de *Streptococcus pneumoniae* ou a lisozima CP1 de fago pneumocócico (CPL1), em que o domínio de ligação à colina é modificado para incluir um epitopo heterólogo de auxiliar T. É mostrado que esses parceiros de fusão melhoram o nível de expressão da proteína heteróloga aí ligada e também têm utilidade particular quando fundidos a proteínas ou péptidos pouco imunogénicos que são, de outro modo, úteis como antigénios de vacina. Mais particularmente, esses parceiros de fusão são úteis em construções compreendendo auto-antigénios, e. g., antigénios específicos de tumor ou específicos de tecido.

Antecedentes da invenção

O *streptococcus pneumoniae* sintetiza uma N-acetil-L-alanina amidase, LytA, uma autolisina que degrada especificamente a estrutura básica de peptidoglicano da parede celular conduzindo, eventualmente, à lise celular. A sua cadeia polipeptídica tem dois domínios. O domínio do N-terminal é responsável pela

actividade catalítica, enquanto que o domínio do C-terminal da LytA é responsável pela afinidade à colina e ancoragem à parede celular. Este domínio do C-terminal é conhecido por se ligar a colina e a análogos de colina e também se irá ligar a amins terciárias, tais como DEAE (dietilaminoetilo), habitualmente utilizadas em cromatografia.

A LytA é uma proteína de 318 aminoácidos e a parte do C-terminal compreende um tandem de seis repetições imperfeitas de 20 ou 21 aminoácidos e uma extremidade do COOH-terminal curta. As repetições estão localizadas nas seguintes posições:

R1: 177-191

R2: 192-212

R3: 213-234

R4: 235-254

R5: 255-275

R6: 276-298

Estas repetições são previstas serem numa conformação de beta-volta. O C-terminal é responsável pela ligação à colina. Do mesmo modo, o C-terminal de CPL1 é responsável pela afinidade de ligação e os resíduos aromáticos na repetição contribuem para essa ligação. Estas proteínas foram utilizadas como marcadores de afinidade para permitir purificação rápida (Sanchez Puelles, Eur J Biochem. 1992, 203, 153-9).

Foram também estudadas outras proteínas com um domínio de ligação à colina em *Streptococcus pneumoniae*.

Uma delas, PspA (ou Proteína A de Superfície Pneumocócica), é um factor de virulência (Yother J e Briles (1992) J Bacteriol 174(2) p 601). Esta proteína é antigénica e imunogénica. Tem um

domínio do C-terminal consistindo de 10 repetições de 20 aminoácidos, homólogas às repetições da LytA.

A CbpA (ou Proteína A de Ligação à Colina) está envolvida na aderência dos pneumococos às células humanas (Rosenow *et al.* (1997) *Mol Microbiol* 25 (5) p 819). Apresenta 10 repetições de 20 aminoácidos no domínio do C-terminal que são quase idênticas àsquelas da PspA.

A LytB e a LytC têm uma organização modular diferente das proteínas acima mencionadas uma vez que os seus domínios de ligação à colina, compostos de 15 repetições e 11 repetições respectivamente, estão situados na extremidade do N-terminal, não na extremidade do C-terminal (Garcia P *Mol Microbiol* (1999) 31 (4) p1275 e Garcia P *et al.* (1999) *Mol Microbiol* do 33(1) p128). A comparação de sequência mostra que LytB tem actividade de glucosamidase. A LytC mostra *in vitro* uma actividade de tipo lisozima. Adicionalmente, em 1995 foram clonados três genes denominados PepA, PepB e PepC. Embora a sua função seja desconhecida, estes genes também têm um número variável de repetições homólogas àsquelas da LytA.

No seu ciclo de infecção, os fagos sintetizam mureína hidrolases que facilitam a sua passagem para a bactéria. Estas hidrolases têm um domínio de ligação à colina.

A muramidase CPL1 do fago Cp-1 tem sido bem estudada. Esta apresenta 6 repetições de 20 aminoácidos no C-terminal envolvidas no reconhecimento específico da colina (Garica J. L. *J. Virol* 61 (8) p2573-80; (1987) e Garcia E *Proc Natl Acad Sci* (1988) p914). Uma comparação das repetições de LytA e CPL1 permite que seja feito um consenso inicial daquelas repetições.

As mureína hidrolases de fagos Dp-1 (Garcia P *et al.* (1983) J Gen Microbiol 129 (2) p489, Cpl-9 (Garcia P *et al.* (1989) Biochem Biophys Res Commun 158(1) p 251, Hb-3 Romero *et al.* 1990 J Bacteriol 172 (9) p 5064-5070) e EJ-1 Diaz (1992) J Bacteriol 174 (17) p 5516), também apresentam as características de domínios de ligação à colina.

Esta propriedade é também partilhada pela lisozima codificada pelo fago pneumocócico Cp-1. O documento WO 99/10375 descreve *inter alia*, as proteínas do vírus do papiloma humano E6 ou E7 ligadas a um marcador His e à porção do C-terminal da LytA (aqui C-LytA) e a purificação das proteínas através de cromatografia de afinidade diferencial. O documento WO 99/40188 descreve *inter alia*, proteínas de fusão compreendendo antigénios MAGE com extremidades His e uma porção de C-LytA no N-terminal da molécula.

Foi agora surpreendentemente verificado que parceiros de fusão de acordo com a presente invenção, quando fundidos a uma proteína heteróloga, foram capazes de intensificar a imunogenicidade das proteínas heterólogas aí ligadas. Foi também verificado que pode ser intensificado o nível de expressão das proteínas heterólogas aí ligadas. De acordo com o exposto, a presente invenção proporciona, numa forma de realização preferida, um parceiro de fusão imunológico melhorado que também pode actuar como um intensificador de expressão.

Sumário da invenção

Consequentemente, a presente invenção compreende uma molécula de parceiro de fusão compreendendo um domínio de ligação à colina ou um seu fragmento ou um seu análogo, e um

epitopo de auxiliar T heterólogo promíscuo, de um modo preferido, um epitopo de T de MHC Classe II promíscuo. O referido parceiro de fusão mostra uma capacidade de actuação como um parceiro de fusão imunológico ou como um intensificador de expressão e, de um modo preferido, como um parceiro imunológico e como intensificador de expressão. Um epitopo de auxiliar T promíscuo é um epitopo que se liga a mais de um alelo de MHC Classe II, de um modo preferido, a mais de 3 alelos de MHC Classe II. Em particular, esses epitopos são capazes de induzir uma resposta por célula auxiliar T num grande número de indivíduos que expressam diversos haplotipos de MHC. Opcionalmente, a proteína de fusão pode reter a sua capacidade de ligação à colina.

Numa forma de realização preferida, a porção de ligação à colina é derivada do C-terminal da LytA. De um modo preferido, a C-LytA ou derivados compreendem, pelo menos, quatro repetições de qualquer uma das repetições R1 a R6 apresentadas na figura 1 (SEQ ID N°:1 a 6). Numa forma de realização muito preferida, a C-LytA estende-se dos aminoácidos 177-298 que contêm uma porção da primeira repetição e os cinco outros completos.

Num aspecto adicional da invenção é proporcionado um parceiro de fusão como aqui definido, compreendendo ainda uma proteína heteróloga. A proteína heteróloga pode ser quimicamente conjugada ou fundida ao parceiro de fusão. De um modo preferido, a proteína heteróloga é um antigénio associado a tumor ou um seu fragmento imunogénico.

Num aspecto adicional da invenção é proporcionada uma sequência de ácidos nucleicos codificando as proteínas como aqui definidas. É também proporcionado um vector de expressão compreendendo o referido ácido nucleico e um hospedeiro transformado com o referido ácido nucleico ou vector.

Num aspecto adicional da invenção é proporcionada uma composição imunogénica compreendendo uma proteína ou uma sequência de ácidos nucleicos, como aqui descrita, e um excipiente, diluente ou veículo farmacêuticamente aceitável. De um modo preferido, a composição imunogénica compreende ainda um adjuvante indutor de Th-1.

Ainda noutra forma de realização, a invenção proporciona a composição imunogénica ou proteína e ácidos nucleicos para utilização em medicina. Em particular, é proporcionada uma proteína ou um ácido nucleico da invenção, na produção de um medicamento para induzir uma resposta imunológica num doente ou para utilização no tratamento ou profilaxia de doenças infecciosas ou doenças cancerígenas.

A invenção proporciona, ainda, métodos de tratamento de um doente que sofre de uma doença infecciosa ou uma doença cancerígena, particularmente carcinoma da mama, pulmonar (particularmente, carcinoma pulmonar das células não pequenas), colorrectal, ovárico, da próstata, gástrico e outros GI (gastrointestinais) através da administração de uma quantidade segura e eficaz de uma composição ou ácido nucleico como aqui descrito.

Ainda noutra forma de realização, a invenção proporciona um método de produção de uma composição imunogénica como aqui descrita, por mistura de um ácido nucleico ou proteína da invenção com um excipiente, diluente ou veículo farmacêuticamente aceitável.

Descrição detalhada da invenção

Como aí descrito, numa forma de realização da presente invenção o domínio modificado de ligação à colina (parceiro de fusão) tem uma capacidade de actuação como um intensificador de expressão, com a resultante proteína de fusão a ser expressa com um rendimento mais elevado numa célula hospedeira comparativamente com a proteína não fundida, de um modo preferido, com um rendimento superior em cerca de 100% (2 vezes superior) ou 150% ou mais, como medido através de SDS-PAGE seguida por coloração com azul de Coomassie ou coloração com prata, opcionalmente seguida por digitalização de gel. O domínio modificado de ligação à colina, de acordo com a invenção tem também a capacidade de actuação como um parceiro imunológico com a resultante proteína de fusão com uma proteína heteróloga que irá ser mais imunogénica num hospedeiro comparativamente à proteína heteróloga não fundida.

Noutra forma de realização da presente invenção, o domínio modificado de ligação à colina tem a capacidade de actuar como um parceiro de fusão imunológico, permitindo que seja obtida uma resposta imunológica intensificada com a proteína de fusão comparativamente à proteína heteróloga isolada.

Numa forma de realização preferida, o domínio modificado de ligação à colina tem uma função dupla, tendo a capacidade para actuar como um parceiro de fusão imunológico e como um intensificador de expressão.

Numa forma de realização preferida, a unidade de ligação à colina é derivada do C-terminal da LytA. De um modo preferido, a C-LytA ou derivados compreendem, pelo menos, duas repetições, de um modo preferido, pelo menos, quatro repetições. Neste contexto, os derivados de C-LytA referem-se a uma variante de C-LytA de acordo com a presente invenção, isto é, variantes que

retiveram a capacidade de actuação como um parceiro imunológico e como um intensificador de expressão. As variantes preferidas incluem, por exemplo, péptidos compreendendo uma sequência de aminoácidos tendo, pelo menos, 85% de identidade, de um modo preferido, pelo menos, 90% de identidade, de um modo mais preferido, pelo menos, 95% de identidade, de um modo muito preferido, pelo menos, 97-99% de identidade, com quaisquer das repetições R1 a R6 apresentadas na figura 1 (SEQ ID N°:1 a 6) ou um péptido compreendendo uma sequência de aminoácidos tendo, pelo menos, 15, 20, 30, 40, 50 ou 100 aminoácidos contíguos da sequência de aminoácidos apresentada na figura 1 (SEQ ID N°:1 a 8).

De acordo com o exposto, num aspecto da invenção, é proporcionada uma proteína parceira de fusão compreendendo um domínio modificado de ligação à colina e um epitopo heterólogo de auxiliar T promíscuo, em que o domínio de ligação à colina é seleccionado do grupo compreendendo:

a) o domínio do C-terminal da LytA como apresentado em SEQ ID N°:7;

b) a sequência de SEQ ID N°:8;

c) uma sequência peptídica compreendendo uma sequência de aminoácidos tendo, pelo menos, 85% de identidade, de um modo preferido, pelo menos, 90% de identidade, de um modo mais preferido, pelo menos, 95% de identidade, de um modo muito preferido, pelo menos, 97-99% de identidade, com qualquer de SEQ ID N°:1 a 6;

d) uma sequência peptídica compreendendo uma sequência de aminoácidos que tem, pelo menos, 15, 20, 30, 40, 50 ou 100

aminoácidos contíguos da sequência de aminoácidos de SEQ ID a N°:7 ou SEQ ID N°:8.

Numa forma de realização muito preferida, a C-LyTA estende-se dos aminoácidos 177-298 que contém uma porção da primeira repetição e os cinco outros completos, como apresentado na figura 1.

O segundo componente do parceiro de fusão, o epitopo de célula T heterólogo é seleccionado, de um modo preferido, do grupo de epitopos que se irão ligar a um número de indivíduos que expressam mais do que uma molécula MHC II em humanos. Por exemplo, os epitopos que são especificamente contemplados são os epitopos P2 e P30 do toxóide de tétano, Panina - Bordignon Eur. J. Immunol 19 (12), 2237 (1989). Numa forma de realização preferida, o epitopo heterólogo de célula T é P2 ou P30 da toxina do Tétano.

O epitopo P2 tem a sequência QYIKANSKFIGITE e corresponde aos aminoácidos 830-843 da toxina do Tétano. O epitopo P30 (resíduos 947-967 da Toxina do Tétano) tem a sequência FNNFTVSFWLRVPKVSASHLE. A sequência FNNFTV pode ser, opcionalmente, removida. Os outros epitopos de T universais podem ser derivados da proteína circunsporozoite de Plasmodium falciparum - em particular a região 378-398 que tem a sequência DIEKKIAKMEKASSVFNWNS (Alexander J, (1994) Immunity 1 (9), p 751-761). Um outro epitopo é derivado da proteína de fusão do vírus do sarampo nos resíduos 288-302 que têm a sequência LSEIKGVIVHRLEGV (Partidos CD, 1990, J. Gen Virol 71(9) 2099-2105). Ainda outro epitopo é derivado do antigénio de superfície do vírus da hepatite B, em particular os aminoácidos que têm a sequência FLLTRILTIPQSLD. Outro conjunto de epitopos é derivado da toxina da difteria. Quatro destes péptidos (aminoácidos 271-290, 321-340, 331-350, 351-370) estão mapeados

dentro do domínio T do fragmento B da toxina e os restantes 2 estão mapeados no domínio R (411-430, 431-450):

PVFAGANYAAWAVNVAQVI

VHHNTEEIVAQSIALSSLMV

QSIALSSLMVAQAIPLVGEL

VDIGFAAYNFVESII NLFQV

QGESGHDIKITAENTPLPIA

GVLLPTIPGKLDVNVKSKTHI

(Raju R., Navaneetham D., Okita D., Diethelm-Okita B., McCormick D., Conti-Fine B. M. (1995) Eur. J. Immunol. 25: 3207-14.)

O epitopo T heterólogo é, de um modo preferido fundido a C-LytA contendo, pelo menos, 4 repetições, de um modo preferido repetições 2-5 inclusive. Uma ou mais repetições subsequentes podem ser, opcionalmente, fundidas ao C-terminal do epitopo T. Alternativamente, o epitopo T heterólogo é, de um modo preferido, introduzido entre duas repetições consecutivas de C-LytA contendo um total de, pelo menos, 4 repetições ou introduzido numa das repetições de C-LytA contendo um total de, pelo menos, 4 repetições. De um modo mais preferido, a C-LytA contém 6 repetições e o epitopo heterólogo é introduzido dentro e no início da sexta repetição de C-LytA.

A presente invenção proporciona ainda, noutros aspectos, proteínas de fusão que compreendem, pelo menos, um polipéptido como acima descrito, assim como polinucleótidos codificando essas proteínas de fusão, tipicamente na forma de composições farmacêuticas, e. g., composições de vacina, compreendendo um veículo fisiologicamente aceitável e/ou um imunoestimulante. Assim, uma auto-proteína ou outra proteína pouco imunogénica pode ser fundida à extremidade N- ou C-terminal do parceiro de fusão resultante. Alternativamente, a auto-proteína ou a

proteína pouco imunogénica pode ser introduzida no parceiro de fusão. Numa forma de realização opcional, um marcador de histidina ou, pelo menos, quatro, de um modo preferido mais de 6 resíduos de histidina podem ser fundidos à extremidade alternativa da proteína pouco imunogénica. Isto permitiria à proteína ser purificada através de passos de cromatografia de afinidade, como uma extremidade de histidina compreendendo tipicamente, pelo menos, quatro, de um modo preferido seis ou mais resíduos de ligação a iões metálicos e é, deste modo, adequada para cromatografia de afinidade de ião metálico (IMAC) com metal imobilizado.

As construções típicas compreenderiam assim:

- Proteína pouco imunogénica - repetições₁₋₄ de C-LytA - epitopo P₂ (introduzido ou substituindo a repetição₅ de C-LytA) - repetições C-LytA
- Repetições₁₋₄ de C-LytA - epitopo P₂ (introduzido na ou substituindo a repetição₅ de C-LytA) - repetição₆ de C-LytA - Proteína pouco imunogénica
- Proteína pouco imunogénica - repetições₂₋₅ de C-LytA epitopo P₂ (introduzido na repetição₆ de C-LytA)
- C-LytA₂₋₅ - epitopo P₂ (introduzido na repetição₆ de C-LytA) - Proteína pouco imunogénica.
- Proteína pouco imunogénica - repetições₁₋₅ de C-LytA - epitopo P₂ introduzido na repetição₆ de C-LytA
- Repetições₁₋₅ de C-LytA - epitopo P₂ introduzido na repetição₆ de C-LytA - Proteína pouco imunogénica
- Proteína pouco imunogénica - epitopo P₂ introduzido na repetição₁ de C-LytA - repetições₂₋₅ de C-LytA
- Epitopo P₂ introduzido na repetição₁ de C-LytA - repetições₂₋₅ de C-LytA - Proteína pouco imunogénica
- Proteína pouco imunogénica - epitopo P₂ introduzido na repetição₁ de C-LytA - repetições₂₋₆ de C-LytA

- Epitopo P₂ introduzido na repetição₁ de C-LytA - repetições₂₋₆ de C-LytA - Proteína pouco imunogénica
- Proteína pouco imunogénica - repetição₁ de C-LytA - epitopo P₂ introduzido na repetição₂ de C-LytA - repetições₃₋₆ de C-LytA
- Repetição₁ de C-LytA - epitopo P₂ introduzido na repetição₂ de C-LytA - repetições₃₋₆ de C-LytA - Proteína pouco imunogénica;

em que "introduzido" significa em qualquer local na repetição, por exemplo, entre o resíduo 1 e 2 ou entre 2 e 3, etc.

O epitopo de auxiliar T promíscuo pode ser introduzido dentro de uma região de repetição, por exemplo, repetições₂₋₅ de C-LytA - repetição 6a de C-LytA - epitopo P₂ - repetição 6b de C-LytA, em que o epitopo P₂ é introduzido dentro da sexta repetição (ver figura 2).

Noutras formas de realização preferidas, a extremidade C-terminal de CPL1 (C-CPL1) pode ser utilizada como uma alternativa a C-LytA.

Alternativamente, o epitopo P₂ nas construções acima pode ser substituído por outros epitopos T promíscuos, por exemplo, P₃₀. Numa forma de realização da invenção, dois ou mais epitopos promíscuos são parte da construção de fusão. É contudo preferido manter o parceiro de fusão tão pequeno quanto possível, limitando assim o número de epitopos CD8+ e B, potencialmente interferentes. Assim, o parceiro de fusão é, de um modo preferido, não maior que 100-140 aminoácidos, de um modo preferido, não maior que 120 aminoácidos, tipicamente cerca de 100 aminoácidos.

O antigénio ao qual o parceiro de fusão é fundido pode ser de fontes bacterianas, virais, de protozoários, fúngicas ou de mamífero, incluindo humano.

O parceiro de fusão da presente invenção é fundido, de um modo preferido, a um auto-antigénio, tal como antigénios específicos de tecido ou associados a tumor, tais como aqueles para cancros da próstata, mama, colorrectal, pulmonar, pancreático, ovárico, renal ou melanoma. Os fragmentos dos referidos antigénios auto ou tumorais são expressamente contemplados para serem fundidos ao parceiro de fusão da invenção. Tipicamente, o fragmento conterá, pelo menos, 20, de um modo preferido 50, de um modo mais preferido 100 aminoácidos contíguos da sequência de tamanho total. Tipicamente, esses fragmentos serão desprovidos de um ou mais domínios transmembranares ou podem ter deleções no N-terminal ou C-terminal de cerca de 3, 5, 8, 10, 15, 20, 28, 33, 50, 54 aminoácidos. Esses fragmentos serão capazes, quando apropriadamente apresentados, de gerar respostas imunológicas que reconhecem a proteína de tamanho total. Os polipéptidos particularmente ilustrativos da presente invenção compreendem uma sequência de, pelo menos, 10 aminoácidos contíguos, de um modo preferido 20, de um modo mais preferido 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180 aminoácidos de uma proteína específica de tecido ou associada a tumor fundida ao parceiro de fusão.

Os polipéptidos da invenção são imunogénicos, *i. e.*, reagem detectavelmente num ensaio imunológico (tal como um ensaio de estimulação de células T ou ELISA) com anti-soros e/ou células T de um doente com cancro expressando cripto. O rastreio para a actividade imunogénica pode ser realizado utilizando técnicas bem conhecidas do especialista. Por exemplo, esses rastreios podem ser realizados utilizando métodos tais como aqueles

descritos em Harlow e Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. Num exemplo ilustrativo, um polipéptido pode ser imobilizado num suporte sólido e colocado em contacto com soros de doentes para permitir a ligação de anticorpos dos soros ao polipéptido imobilizado. Os soros não ligados podem ser depois removidos e os anticorpos ligados detectados utilizando, por exemplo, Proteína A marcada com ¹²⁵I. Como seria reconhecido por um especialista, as porções imunogénicas de antigénio associado a tumor ou específico de tumor estão também abrangidas pela presente invenção. Uma "porção imunogénica" como aqui utilizado, é um fragmento que ele próprio é imunologicamente reactivo (*i. e.*, liga-se especificamente) com os receptores antigénicos de superfície de células B e/ou células T que reconhecem o polipéptido. As porções imunogénicas podem ser geralmente identificadas utilizando técnicas bem conhecidas, tais como aquelas resumidas em Paul, *Fundamental Immunology*, 3ª ed., 243-247 (Raven Press, 1993) e referências aí citadas. Essas técnicas incluem rastreio de polipéptidos para a capacidade de reagirem com anticorpos específicos de antigénio, anti-soros e/ou linhas de células T ou clones. Como aqui utilizado, os anti-soros e anticorpos são "específicos de antigénio" se se ligarem especificamente a um antigénio (*i. e.*, reagem com a proteína num ELISA ou outro ensaio imunológico, e não reagem detectavelmente com proteínas não relacionadas). Esses anti-soros e anticorpos podem ser preparados como aqui descrito, e utilizando técnicas bem conhecidas. Numa forma de realização preferida, uma porção imunogénica de um polipéptido é uma porção que reage com anti-soros e/ou células T a um nível que não é substancialmente menor do que a reactividade do polipéptido de tamanho total (*e. g.*, num ensaio de reactividade de célula T e/ou ELISA). De um modo preferido, o nível de actividade imunogénica da porção imunogénica é, pelo menos, cerca de 50%, de um modo preferido, pelo menos, cerca de 70% e de um modo muito preferido maior do

que cerca de 90% da imunogenicidade para o polipéptido de tamanho total. Nalguns casos, serão identificadas as porções imunogénicas preferidas que têm um nível de actividade imunogénica maior do que aquela do polipéptido de tamanho total correspondente, e. g., tendo mais do que cerca de 100% ou 150% ou mais de actividade imunogénica.

Nalgumas outras formas de realização, as porções imunogénicas ilustrativas podem incluir péptidos em que foi removida uma sequência líder do N-terminal e/ou domínio transmembranar. Outras porções imunogénicas ilustrativas irão conter uma pequena deleção do N- e/ou C-terminal (e. g., cerca de 1-50 aminoácidos, de um modo preferido cerca de 1-30 aminoácidos, de um modo mais preferido cerca de 5-15 aminoácidos), relativamente à proteína madura.

Os antigénios exemplificativos ou fragmentos daí derivados incluem antigénios MAGE 1, Mage 3 e MAGE 4 ou outros MAGE, tal como divulgados no documento WO 99/40188, PRAME (documento WO 96/10577), BAGE, RAGE, LAGE 1 (documento WO 98/32855), LAGE 2 (também conhecido como NY-ESO-1, documento WO 98/14464), XAGE (Liu et al., Cancer Res, 2000, 60:4752-4755; documento WO 02/18584) SAGE, e HAGE (documento WO 99/53061) ou GAGE (Robbins e Kawakami, 1996, Current Opinions in Immunology 8, pps 628-636; Van den Eynde et al., International Journal of Clinical e Laboratory Research (submetido em 1997); Correale et al. (1997), Journal of the National Cancer Institute 89, p293. De facto estes antigénios são expressos numa vasta gama de tipos de tumor, tais como melanoma, carcinoma pulmonar, sarcoma e carcinoma da bexiga.

Numa forma de realização preferida são utilizados antigénios da próstata, tais como antigénio específico de Próstata (PSA),

PAP, PSCA (PNAS 95(4) 1735-1740 1998), PSMA ou o antigénio conhecido como prostase.

Numa forma de realização particularmente preferida, o antigénio de próstata é P501S ou um seu fragmento. O P501S, também denominado prostateína (Xu et al., Cancer Res. 61, 2001, 1563-1568), é conhecida como SEQ ID N° 113 do documento WO98/37814 e é uma proteína de 553 aminoácidos. Os fragmentos imunogénicos e suas porções compreendendo, pelo menos, 20, de um modo preferido 50, de um modo mais preferido 100 aminoácidos contíguos como divulgados no pedido de patente acima referenciado e são especificamente contemplados através da presente invenção. Os fragmentos preferidos são divulgados no documento WO 98/50567 (antigénio PS108) e como proteína associada ao cancro da próstata (SEQ ID N°: 9 do documento WO 99/67384). Outros fragmentos preferidos são os aminoácidos 51-553, 34-553 ou 55-553 da proteína P501S de tamanho total. Em particular, as construções 1, 2 e 3 (ver figura 2, SEQ ID N° 27-32) são contempladas e podem ser expressas em sistemas de levedura, por exemplo, sequências de ADN codificando esses polipéptidos podem ser expressos em sistema de levedura.

A prostase é uma protease de serina específica da próstata (de tipo tripsina), de 254 aminoácidos de comprimento, com uma tríade H-D-S catalítica conservada de protease de serina e uma sequência pré-propéptido do amino-terminal, indicando uma potencial função excretora (P. Nelson, Lu Gan, C. Ferguson, P. Moss, R. Linas, L. Hood e K. Wand, "Molecular cloning and characterisation of prostase, an androgen-regulated serine protease with prostate restricted expression, em Proc. Natl. Acad. Sci. USA (1999) 96, 3114-3119). Foi descrito um local putativo de glicosilação. A estrutura prevista é muito semelhante a outras proteases de serina conhecidas, mostrando que o polipéptido maduro se enrola num único domínio. A proteína

madura tem 224 aminoácidos, com um epitopo A2 mostrado ser processado naturalmente. A sequência de nucleótidos da prostase e a sequência polipeptídica deduzida e homólogos são divulgados em Ferguson, et al. (Proc. Natl. Acad. Sci. USA 1999, 96, 3114-3119) e nos Pedidos de Patente Internacional N° WO 98/12302 (e também a correspondente patente concedida U.S. 5955306), WO 98/20117 (e também as correspondentes patentes concedidas U.S. 5840871 e U.S. 5786148) (calicreína específica da próstata) e WO 00/04149 (P703P).

Outros antigénios específicos da próstata são conhecidos dos documentos WO98/37418 e WO/004149. Outro é STEAP (PNAS 96 14523 14528 7-12 1999).

Outros antigénios associados a tumor, úteis no contexto da presente invenção incluem: Plu-1 J. Biol. Chem 274 (22) 15633 - 15645, 1999, HASH-1, HASH-2 (Alders, M. et al., Hum. Mol. Genet. 1997, 6, 859-867). Cripto (Salomon et al. Bioessays 199, 21 61 - 70, patente U.S. 5654140), CASB616 (documento WO 00/53216), Criptina (documento U.S. 5981215). Adicionalmente, os antigénios particularmente relevantes para vacinas na terapia do cancro também compreendem tirosinase, telomerase, P53, NY-Br1.1 (documento WO 01/47959) e os seus fragmentos, como divulgado no documento WO 00/43420, B726 (documento WO 00/60076, SEQ ID N° 469 e 463; documento WO 01/79286, SEQ ID N° 474 e 475), P510 (documento WO 01/34802 SEQ ID N° 537 e 538) e survivina.

A presente invenção é também útil em combinação com antigénios do cancro da mama, tais como Her-2/neu, mamaglobina (patente U.S. 5668267), B305D (documento WO 00/61753 SEQ ID N° 299, 304, 305 e 315) ou aqueles divulgados nos documentos WO 00/52165, WO 99/33869, WO 99/19479, WO 98/45328. Os antigénios Her-2/neu são divulgados *inter alia*, na patente U.S. 5801005. De um modo preferido, o Her-2/neu compreende o

domínio extracelular completo (compreendendo, aproximadamente, 1-645 aminoácidos) ou os seus fragmentos e, pelo menos, uma porção imunogénica de ou o domínio intracelular inteiro, aproximadamente, de 580 aminoácidos do C-terminal. Em particular, a porção intracelular deve compreender o domínio de fosforilação ou os seus fragmentos. Essas construções são divulgadas no documento WO 00/44899. Uma construção particularmente preferida é conhecida como ECD-PhD, uma segunda é conhecida como ECD deltaPhD (ver documento WO 00/44899). O Her-2/neu como aqui utilizado, pode ser derivado de rato, murganho ou humano.

Determinados antigénios tumorais são pequenos antigénios peptídicos (*i. e.*, menos do que cerca de 50 aminoácidos). Estes antigénios podem ser quimicamente conjugados com a proteína de ligação à colina modificada da presente invenção.

Os péptidos exemplificativos incluem péptidos derivados de Mucina, tais como MUC-1 (ver, por exemplo, documentos U.S. 5744144; U.S. 5827666; WO 88/05054, U.S. 4963484). São especificamente contemplados péptidos derivados de MUC-1 que compreendem, pelo menos, uma unidade de repetição do péptido MUC-1, de um modo preferido, pelo menos, duas dessas repetições e que é reconhecida pelo anticorpo SM3 (documento U.S. 6054438). Outros péptidos derivados de mucina incluem péptidos de MUC-5.

Alternativamente, o referido antigénio é uma interleucina, tal como IL13 e IL14, as quais são preferidas. Ou o referido antigénio talvez uma hormona auto-peptídica, tal como a hormona libertadora da hormona Gonadotrofina de tamanho total (GnRH, documento WO 95/20600), um péptido curto de 10 aminoácidos, úteis no tratamento de muitos cancros ou em imunocastração.

Outros antigénios específicos de tumor são adequados para serem acoplados com a proteína de ligação à Colina modificada da presente invenção incluem, mas não são restritos a, gangliósidos específicos de tumor, tais como GM2 e GM3.

O acoplamento covalente do péptido à proteína de ligação à colina modificada pode ser realizada de um modo bem conhecido na técnica. Assim, por exemplo, para acoplamento covalente directo é possível utilizar uma carbodiimida, glutaraldeído ou um éster de (N-[γ -maleimidobutirilo]succinimida, utilizando ligantes heterobifuncionais comuns comercialmente disponíveis, tais como CDAP e SPDP (utilizando as instruções dos fabricantes). Após a reacção de acoplamento, o imunogénio pode ser facilmente isolado e purificado através de um método de diálise, um método de filtração em gel, um método de fraccionamento, etc.

O antigénio pode também ser derivado de fontes que são patogénicas aos humanos, tais como o vírus da Imunodeficiência Humana HIV-1 (tal como tat, nef, transcriptase reversa, gag, gp120 e gp160), vírus do herpes simplex humano, tais como gD ou os seus derivados ou proteína Imediatamente Precoce, tal como ICP27 de HSV1 ou HSV2, citomegalovírus ((esp Humano) (tais como gB ou os seus derivados), Rotavírus (incluindo vírus vivos atenuados), vírus de Epstein-Barr (tal como gp350 ou os seus derivados), Vírus Varicella Zoster (tal como gpI, II e IE63) ou de um vírus da hepatite, tal como o vírus da hepatite B (por exemplo, o antigénio de Superfície da Hepatite B ou um seu derivado), vírus da hepatite A, vírus da hepatite C e vírus da hepatite E ou de outros patogéneos virais, tais como paramixovírus: Vírus Sincicial Respiratório (tais como as proteínas F e G ou os seus derivados), vírus da parainfluenza, vírus do sarampo, vírus da parotidite, vírus do papiloma humano (por exemplo, HPV6, 11, 16, 18,.), flavivírus (e. g., Vírus da Febre Amarela, Vírus do Dengue, vírus da encefalite da Febre da

Carraça, Vírus da Encefalite Japonesa) ou vírus Influenza (vírus inteiro vivo ou inativado, vírus influenza dividido, cultivado em ovos ou células MDCK ou virossomas inteiros da gripe (como descrito por R. Gluck, Vaccine, 1992, 10, 915-920) ou as suas proteínas purificadas ou recombinantes, tais como proteínas HA, NP, NA ou M ou as suas combinações) ou derivados de patógenos bacterianos, tais como *Neisseria spp*, incluindo *N. gonorrhea* e *N. meningitidis* (por exemplo, polissacáridos capsulares e os seus conjugados, proteínas de ligação a transferrina, proteínas de ligação a lactoferrina, PIIC, adesinas); *S. Pyogenes* (por exemplo, proteínas M ou os seus fragmentos, protease C5A, ácidos lipoteicóicos), *S. agalactiae*, *S. mutans*; *H. Ducreyi*; *Moraxella spp*, incluindo *M. catarrhalis*, também conhecida como *Branhamella catarrhalis* (por exemplo, adesinas e invasinas de peso molecular elevado e baixo); *Bordetella spp*, incluindo *B. pertussis* (por exemplo, pertactina, toxina pertussis ou os seus derivados, hemaglutinina filamentosa, adenilato ciclase, fimbriae), *B. parapertussis* e *B. bronchiseptica*; *Mycobacterium spp.*, incluindo *M. tuberculosis* (por exemplo, ESAT6, Antígeno 85A, -B ou -C), *M. bovis*, *M. leprae*, *M. avium*, *M. paratuberculosis*, *M. smegmatis*; *Legionella spp*, incluindo *L. pneumophila*; *Escherichia spp*, incluindo *E. coli* enterotóxica (por exemplo, factores de colonização, toxinas ou os seus derivados, toxinas estáveis ao calor ou os seus derivados), *E. coli* entero-hemorrágica, *E. coli* enteropatogénica (por exemplo, toxina do tipo toxina shiga ou os seus derivados); *Vibrio spp*, incluindo *V. cholera* (por exemplo, toxina da cólera ou os seus derivados); *Shigella spp*, incluindo *S. sonnei*, *S. dysenteriae*, *S. flexnerii*; *Yersinia spp*, incluindo *Y. enterocolitica* (por exemplo, uma proteína Yop), *Y. pestis*, *Y. pseudotuberculosis*; *Campylobacter spp*, incluindo *C. jejuni* (por exemplo, toxinas, adesinas e invasinas) e *C. coli*, *Salmonella spp*, incluindo *S. typhi*, *S. paratyphi*, *S. choleraesuis*, *S. enteritidis*; *Listeria spp.*, incluindo

L. monocytogenes; *Helicobacter* spp, incluindo *H. pylori* (por exemplo, urease, catalase, toxina vacuolante); *Pseudomonas* spp, incluindo *P. aeruginosa*; *Staphylococcus* spp., incluindo *S. aureus*, *S. epidermidis*; *Enterococcus* spp., *E. faecalis*, incluindo *E. faecium*; *Clostridium* spp., incluindo *C. tetani* (por exemplo, toxina do tétano e os seus derivados), *C. botulinum* (por exemplo, toxina botulina e os seus derivados), *C. difficile* (por exemplo, toxinas A ou B de clostridium e os seus derivados); *Bacillus* spp., incluindo *B. anthracis* (por exemplo, toxina botulina e os seus derivados); *Corynebacterium* spp., incluindo *C. diphtheriae* (por exemplo, toxina da difteria e os seus derivados); *Borrelia* spp., incluindo *B. burgdorferi* (por exemplo, OspA, OspC, DbpA, DbpB), *B. garinii* (por exemplo, OspA, OspC, DbpA, DbpB), *B. afzelii* (por exemplo, OspA, OspC, DbpA, DbpB), *B. andersonii* (por exemplo, OspA, OspC, -DbpA, DbpB), *B. hermsii*, *Ehrlichia* spp., incluindo *E. equi* e o agente da Erliquiose Granulocítica Humana; *Rickettsia* spp, incluindo *R. rickettsii*; *Chlamydia* spp., incluindo *C. trachomatis* (por exemplo, proteínas MOMP, de ligação a heparina), *C. pneumoniae* (por exemplo, proteínas MOMP, de ligação a heparina), *C. psittaci*; *Leptospira* spp., incluindo *L. interrogans*; *Treponema* spp., incluindo *T. pallidum* (por exemplo, as proteínas raras da membrana externa), *T. denticola*, *T. hyodysenteriae*; ou derivados de parasitas, tais como *Plasmodium* spp., incluindo *P. falciparum*, *Toxoplasma* spp., incluindo *T. gondii* (por exemplo, SAG2, SAG3, Tg34); *Entamoeba* spp., incluindo *E. histolytica*; *Babesia* spp., incluindo *B. microti*; *Trypanosoma* spp, incluindo *T. cruzi*; *Giardia* spp., incluindo *G. lamblia*; *Leshmania* spp., incluindo *L. major*; *Pneumocystis* spp., incluindo *P. carinii*; *Trichomonas* spp., incluindo *T. vaginalis*; *Schistosoma* spp., incluindo *S. mansoni* ou derivados de levedura,

tal como *Candida spp.*, incluindo *C. lbicans*; *Cryptococcus spp.*, incluindo *C. neoformans*.

Outros antígenos específicos preferidos para *M. tuberculosis* são, por exemplo, Tb Ra12, Tb H9, Tb Ra35, Tb38-1, Erd 14, DPV, MTI, MSL, mTTC2 e hTCC1 (documento WO 99/51748). As proteínas para *M. tuberculosis* também incluem proteínas de fusão e suas variantes, em que, pelo menos, dois, de um modo preferido três polipéptidos de *M. tuberculosis* estão fundidos numa proteína maior. As fusões preferidas incluem Ra12-TbH9-Ra35, Erd14-DPV-MTI, DPV-MTI-MSL, Erd14-DPV-MTI-MSL-mTCC2, Erd14-DPV-MTI-MSL, DPV-MTI-MSL-mTCC2, TbH9-DPV-MTI (documento WO 99/51748).

Os antígenos muito preferidos para *Chlamydia* incluem, por exemplo, a Proteína de Elevado Peso Molecular (HWMP) (documento WO 99/17741), ORF3 (documento EP 366412) e proteínas putativas de membrana (Pmps). Outros antígenos de *Chlamydia* da formulação de vacina podem ser seleccionados do grupo descrito no documento WO 99/28475.

Os antígenos bacterianos preferidos são derivados de *Streptococcus spp.*, incluindo *S. pneumoniae* (por exemplo, polissacáridos capsulares e os seus conjugados, PsaA, PspA, estreptolisina, proteínas de ligação à colina) e o antígeno proteico Pneumolisina (Biochem Biophys Acta, 1989, 67, 1007; Rubins *et al.*, Microbial Pathogenesis, 25, 337-342), e os seus derivados mutantes desintoxicados (documentos WO 90/06951; WO 99/03884). Outros antígenos bacterianos preferidos são derivados de *Haemophilus spp.*, incluindo *H. influenzae* tipo B (por exemplo, PRP e os seus conjugados), *H. influenzae* não tipável, por exemplo, OMP26, adesinas de peso molecular elevado, P5, P6, proteína D e lipoproteína D, e fimbrina e péptidos

derivados de fimbrina (documento U.S. 5843464) ou variantes de múltiplas cópias ou suas proteínas de fusão.

Os derivados do antígeno de Superfície da Hepatite B são bem conhecidos na técnica e incluem, *inter alia*, aqueles antígenos S PreS1, PreS2 apresentados, descritos nos pedidos de Patente Europeia Ep-A-414374; EP-A-0304578 e EP198-474. Num preferido, o antígeno HBV é a HBV polimerase (Ji Hoon Jeong *et al*, 1996, BBRC 223, 264-271; Lee H.J. *et al*, Biotechnol. Lett. 15. 821-826). Noutro aspecto preferido, o antígeno dentro da fusão é um antígeno HIV-1, gp120, especialmente quando expresso em células CHO. Numa forma de realização adicional, o antígeno compreende gD2t como aqui definido acima.

Numa forma de realização preferida da presente invenção, as fusões compreendem um antígeno derivado do Vírus do Papiloma Humano (HPV 6a, 6b, 11, 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 e 68), em particular aqueles serotipos de HPV considerados serem responsáveis por verrugas genitais (HPV 6 ou HPV 11 e outros), e os vírus HPV responsáveis pelo cancro cervical (HPV16, HPV18 e outros).

Os antígenos adequados de HPV são E1, E2, E4, E5, E6, E7, L1 e L2. Particularmente formas preferidas de fusões profiláticas ou terapêuticas de verruga genital, compreendem partículas L1 ou capsómeros, e proteínas de fusão compreendendo um ou mais antígenos seleccionados das proteínas E6, E7, L1, e L2 de HPV 6 e HPV 11.

As formas mais preferidas de proteína de fusão são: L2E7 como divulgada no documento WO 96/26277, e proteínaD(1/3)-E7 divulgada no documento GB 9717953.5 (documento PCT/EP98/05285).

Uma composição preferida de vacina terapêutica ou profilática de cancro ou infecção cervical de HPV pode compreender antígenos 16 ou 18 de HPV. Por exemplo, monómeros de antígeno L1 ou L2 ou antígenos L1 ou L2 apresentados conjuntamente como uma partícula de tipo viral (VLP) ou só a proteína L1 apresentada isolada numa VLP ou estrutura de capsómero. Esses antígenos, partículas de tipo viral e capsómero são conhecidos por si. Ver, por exemplo, documentos WO94/00152, WO94/20137, WO94/05792 e WO93/02184.

Podem ser incluídas proteínas precoces adicionais isoladas ou como proteínas de fusão, tais como E7, E2 ou, de um modo preferido, E5, por exemplo; formas de realização particularmente preferidas destas incluem uma VLP compreendendo proteínas de fusão L1 E7 (documento WO 96/11272). Antígenos particularmente preferidos de HPV 16 compreendem as proteínas precoces E6 ou E7 em fusão com uma proteína D de veículo para formar as fusões Proteína D - E6 ou E7 a partir de HPV 16 ou suas combinações; ou combinações de E6 ou E7 com L2 (documento WO 96/26277). Alternativamente, as proteínas precoces E6 e E7 de HPV 16 ou 18, podem ser apresentadas numa única molécula, de um modo preferido, uma fusão Proteína D - E6/E7. Outras fusões contêm opcionalmente cada uma ou ambas as proteínas E6 e E7 de HPV 18, de um modo preferido, na forma de uma proteína de fusão Proteína D - E6 ou Proteína D - E7 ou proteína de fusão Proteína D - E6/E7. As fusões podem compreender antígenos de outras estirpes de HPV, de um modo preferido de estirpes 31 ou 33 de HPV.

As fusões, de acordo com a presente invenção, compreendem antígenos derivados de parasitas que causam a Malária. Por exemplo, os antígenos preferidos de *Plasmodia falciparum* incluem RTS,S e TRAP. A RTS é uma proteína híbrida compreendendo substancialmente toda a porção do C-terminal da proteína

circunsporozoítica (CS) de *P. falciparum* ligada através de quatro aminoácidos da porção preS2 do antígeno de superfície da Hepatite B ao antígeno de superfície (S) do vírus da hepatite B. A sua estrutura completa é divulgada no Pedido de Patente Internacional N° PCT/EP92/02591, publicado com o Número WO 93/10152 reivindicando prioridade do pedido de patente UK N° 9124390.7. Quando expressa em levedura, a RTS é produzida como uma partícula de lipoproteína, e quando é co-expressa com o antígeno S de HBV produz uma partícula mista conhecida como RTS,S. Os antígenos TRAP são descritos no Pedido de Patente Internacional N° PCT/GB89/00895, publicado no documento WO 90/01496. Uma forma de realização preferida da presente invenção é uma fusão em que a preparação antigénica compreende uma combinação de antígenos RTS,S e TRAP. Outros antígenos de plasmodia que são candidatos prováveis a serem componentes da fusão são MSP1, AMA1, MSP3, EBA, GLURP, RAP1, RAP2, Sequestrina, PfEMP1, Pf332, LSA1, LSA3, STARP, SALSA, PfEXP1, Pfs25, Pfs28, PFS27/25, Pfs16, Pfs48/45, Pfs230 de *P. faciparum* e os seus análogos em Plasmodium spp.

A presente invenção também proporciona um polinucleótido codificando o parceiro de fusão de acordo com a presente invenção. A invenção refere-se ainda a um polinucleótido que hibrida com a sequência de polinucleótidos aqui proporcionada na figura 1 (SEQ ID N°:9 a 16). A este respeito, a invenção refere-se especialmente aos polinucleótidos que hibridam sob condições restritivas para o polinucleótido aqui descrito. Como aqui utilizado, os termos "condições restritivas" e "condições de hibridação restritivas" significam ocorrer hibridação apenas se existir identidade de, pelo menos, 95% e de um modo preferido de, pelo menos, 97% entre as sequências. Um exemplo específico de condições de hibridação restritivas é incubação de um dia para o outro a 42 °C numa solução compreendendo: 50% de formamida, 5x SSC (150 mM de NaCl, 15 mM de citrato trissódico),

50 mM de fosfato de sódio (pH 7,6), 5x solução de Denhardt, 10% de sulfato de dextrano e 20 microgramas/mL de ADN de esperma de salmão cortado e desnaturado, seguido por lavagem do suporte de hibridação em 0,1x SSC a cerca de 65 °C. As condições de hibridação e lavagem são bem conhecidas e exemplificadas em Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Segunda Edição, Cold Spring Harbor, N.Y., (1989), particularmente no seu Capítulo 11. A solução de hibridação pode também ser utilizada com as sequências de polinucleótidos proporcionadas através da invenção.

A presente invenção também proporciona um polinucleótido codificando o polipéptido, compreendendo o parceiro de fusão de acordo com a presente invenção fundido a um antigénio associado a tumor ou seu fragmento. Em particular, a presente invenção proporciona sequências de polinucleótidos codificando uma proteína parceira de fusão compreendendo um domínio de ligação à colina e um epitopo T auxiliar promiscuo heterólogo, de um modo preferido em que o domínio de ligação à colina é derivado a partir do C-terminal da LytA. Numa forma de realização mais preferida, a unidade de C-LytA dos polinucleótidos de acordo com a invenção compreende, pelo menos, quatro repetições de qualquer de SEQ ID N° 9-14, de um modo mais preferido compreende a sequência de SEQ ID N° 15, de um modo ainda mais preferido a sequência de SEQ ID N° 16. Noutras formas de realização relacionadas, a presente invenção proporciona variantes de polinucleótidos que têm identidade substancial com as sequências aqui divulgadas em SEQ ID N°:9-16, por exemplo, aquelas compreendendo, pelo menos, 70% de identidade de sequência, de um modo preferido, pelo menos, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% ou 99% ou superior, identidade de sequência comparativamente a uma sequência de polinucleótidos desta invenção utilizando métodos convencionais, e. g., análise BLAST utilizando parâmetros padrão. Ainda noutra forma de realização, o

polinucleótido como reivindicado compreende ainda uma proteína heteróloga.

Essas sequências de polinucleótidos podem ser introduzidas num vector de expressão adequado e expressas numa hospedeiro adequado. Podem ser proporcionados vectores que codificam a proteína modificada de ligação à colina da invenção e que contêm um sítio de restrição adequado no qual pode ser introduzido um ADN codificando uma proteína pouco imunogénica para produzir uma proteína de fusão.

Noutras formas de realização da invenção, as sequências de polinucleótidos ou os seus fragmentos que codificam fusões polipeptídicas da invenção, podem ser utilizadas em moléculas de ADN recombinante para direccionar a expressão de um polipéptido em células hospedeiras apropriadas. 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 compreendido pelos especialistas na técnica, pode ser vantajoso nalguns casos produzir sequências de nucleótidos codificando polipéptidos que possuem codões de ocorrência não natural. O código de ADN tem 4 letras (A, T, C e G) e utiliza-as para soletrar "codões" de três letras que representam os aminoácidos que as proteínas codificam em genes de um organismo. A sequência linear de codões ao longo da molécula de ADN é traduzida na sequência linear de aminoácidos na(s) proteína(s) codificada(s) por aqueles genes. O código é altamente degenerado, com 61 codões codificando para 20 aminoácidos naturais e 3 codões representando sinais de "terminação". Assim, a maioria dos aminoácidos são codificados por mais de um codão -

de facto diversos são codificados por quatro ou mais códons diferentes.

Nos casos em que está disponível mais de um códon para codificar um determinado aminoácido, foi observado que os perfis de utilização do códon de organismos são altamente não aleatórios. Espécies diferentes apresentam uma propensão diferente na sua selecção de códon e, além disso, a utilização de códons pode ser acentuadamente diferente numa única espécie entre genes que são expressos em níveis elevados e baixos. Esta propensão é diferente em vírus, plantas, bactérias e células de mamíferos, e algumas espécies apresentam uma propensão mais forte do que outras de uma selecção aleatória de códon. Por exemplo, os humanos e outros mamíferos são menos fortemente propensos do que determinadas bactérias ou vírus. Por estas razões, existe uma probabilidade significativa de que um gene de mamífero expresso em *E. coli* ou um gene viral expresso em células de mamífero tenham uma distribuição inadequada de códons para expressão eficiente. Acredita-se que a presença numa sequência de ADN heterólogo de aglomerados de códons que são raramente observados no hospedeiro em que a expressão é para ocorrer, é predicativo de baixos níveis de expressão heteróloga nesse hospedeiro.

Em consequência, os códons preferidos por um hospedeiro procariótico particular (por exemplo, *E. coli* ou levedura) ou eucariótico podem ser otimizados, isto é, seleccionados para aumentar a taxa de expressão de proteína, para produzir um transcripto de ARN recombinante possuindo propriedades desejáveis, tal como, por exemplo, uma semi-vida que é mais longa do que a de um transcripto gerado a partir de uma sequência de ocorrência natural ou para otimizar a resposta imunológica em humanos. O processo de optimização de códon pode incluir qualquer sequência, gerada manualmente ou através de

software de computador, em que são modificados alguns ou todos os codões da sequência nativa. Foram publicados diversos métodos (Nakamura *et al.*, Nucleic Acids Research 1996, 24:214-215; documento W098/34640). Um método preferido de acordo com esta invenção é o método Syngene, uma modificação do método Calcgene (R. S. Hale e G Thompson (Protein Expression and Purification Vol. 12 pp.185-188 (1998))).

Deste modo, numa forma de realização preferida a sequência de ADN da proteína tem um RSCU (Utilização relativa de Codão Sinónimo (também conhecido como Índice de Codão CI)) de, pelo menos, 0,65 e tem menos de 85% de identidade com a região de tipo selvagem correspondente.

Este processo de optimização de codão e as construções resultantes são vantajosos uma vez que pode ter alguns ou todos os seguintes benefícios: 1) melhorar a expressão do produto génico por substituição de codões raros ou infrequentemente utilizados com codões mais frequentemente utilizados, 2) remover ou incluir sítios de enzimas de restrição para facilitar a clonagem a jusante e 3) reduzir o potencial para a recombinação homóloga entre a sequência da inserção no vector de ADN e sequências genómicas e 4) melhorar a resposta imunológica em humanos através do aumento de uma resposta celular e/ou de anticorpo (de um modo preferido, ambas as respostas) contra o antígeno alvo. As sequências da presente invenção reduziram vantajosamente o potencial de recombinação, mas expressam, pelo menos, ao mesmo nível que as sequências de tipo selvagem. Devido à natureza dos algoritmos utilizados através do programa SynGene para gerar uma sequência de codão optimizado, é possível gerar um número extremamente grande de sequências optimizadas de codão diferentes que irão realizar uma função semelhante. Resumidamente, os codões são atribuídos utilizando um método estatístico para produzir um gene sintético possuindo uma

frequência de codão mais próxima daquela encontrada naturalmente em genes de *E. coli* e humanos de elevada expressão. Resumidamente, os codões são atribuídos utilizando um método estatístico para produzir um gene sintético possuindo uma frequência de codão mais próxima daquela encontrada naturalmente em genes humanos altamente expressos, tal como β -Actina. Os exemplos ilustrativos, não limitativos, de sequências de codão otimizado adequadas são apresentados em SEQ ID N°:19-22 e SEQ ID N°:24-26.

Nos polinucleótidos da presente invenção, o perfil de utilização de codão é alterado do típico do antígeno alvo para representar, de forma mais próxima, a propensão de codão de um gene altamente expresso num organismo alvo, por exemplo, a β -actina humana. O "coeficiente de utilização de codão" é uma medida de quão próximo o perfil de codão de uma determinada sequência de polinucleótidos se assemelha àquela de uma espécie alvo. As frequências de codão podem ser derivadas de fontes de literatura para os genes altamente expressos de muitas espécies (ver, e. g., Nakamura et al. Nucleic Acids Research 1996, 24:214-215). As frequências de codão para cada um dos 61 codões (expressos como o número de ocorrências que ocorrem por 1000 codões da classe seleccionada de genes) são normalizadas para cada um dos vinte aminoácidos naturais, de modo que o valor para o codão mais frequentemente utilizado para cada aminoácido seja definido como 1 e as frequências para os codões menos comuns são escalonadas de forma a situarem-se entre zero e 1. Assim, a cada um dos 61 codões é atribuído um valor de 1 ou inferior para os genes altamente expressos da espécie alvo. De modo a calcular um coeficiente de utilização de codão para um polinucleótido específico, relativamente a genes altamente expressos daquela espécie, o valor escalonado para cada codão do polinucleótido específico é assinalado e é calculada a média geométrica de todos estes valores (por divisão da soma dos

logaritmos naturais destes valores pelo número total de codões e calculo do antilogaritmo). O coeficiente terá um valor entre zero e 1 e quanto mais alto for o coeficiente, mais codões no polinucleótido são codões frequentemente utilizados. Se uma sequência de polinucleótidos tiver um coeficiente de utilização de codão de 1, todos os codões são os codões "mais" frequentes para genes altamente expressos da espécie alvo.

De acordo com a presente invenção, o perfil de utilização de codão do polinucleótido irá excluir, de um modo preferido, os codões que representam < 10% dos codões utilizados para um aminoácido particular. Um valor de utilização relativa de codão sinónimo (RSCU) é o número observado de codões dividido pelo número esperado, se todos os codões para aquele aminoácido forem utilizados igualmente de um modo frequente. Um polinucleótido da presente invenção irá excluir, de um modo preferido, codões com um valor de RSCU menor que 0,2 em genes altamente expressos do organismo alvo. Um polinucleótido da presente invenção terá geralmente um coeficiente de utilização de codão para genes humanos altamente expressos superior a 0,6, de um modo preferido superior a 0,65, de um modo muito preferido superior a 0,7. As tabelas de utilização de codão para humanos também podem ser encontradas no Genbank.

Em comparação, um gene da beta actina altamente expresso tem um RSCU de 0,747.

A tabela de utilização de codão (Tabela 1) para um *homo sapiens* é apresentada abaixo:

Tabela 1. Utilização de codão para genes humanos (altamente expressos) 1/24/91 (human_high.cod)

Aminoácido	Codão	Número	/1000	Fracção
Gly	GGG	905,00	18,76	0,24
Gly	GGA	525,00	10,88	0,14
Gly	GGT	441,00	9,14	0,12
Gly	GGC	1867,00	38,70	0,50
Glu	GAG	2420,00	50,16	0,75
Glu	GAA	792,00	16,42	0,25
Asp	GAT	592,00	12,27	0,25
Asp	GAC	1821,00	37,75	0,75
Val	GTG	1866,00	38,68	0,64
Val	Gta	134,00	2,78	0,05
Val	GTT	198,00	4,10	0,07
Val	GTC	728,00	15,09	0,25
Ala	GCG	652,00	13,51	0,17
Ala	GCA	488,00	10,12	0,13
Ala	GCT	654,00	13,56	0,17
Ala	GCC	2057,00	42,64	0,53
Arg	AGG	512,00	10,61	0,18
Arg	AGA	298,00	6,18	0,10
Ser	AGT	354,00	7,34	0,10
Ser	AGC	1171,00	24,27	0,34
Lys	AAG	2117,00	43,88	0,82
Lys	Aaa	471,00	9,76	0,18
Asn	AAT	314,00	6,51	0,22
Asn	AAC	1120,00	23,22	0,78

(continuação)

Aminoácido	Codão	Número	/1000	Fracção
Met	ATG	1077,00	22,32	1,00
Ile	ATA	88,00	1,82	0,05
Ile	ATT	315,00	6,53	0,18
Ile	ATC	1369,00	28,38	0,77
Thr	ACG	405,00	8,40	0,15
Thr	ACA	373,00	7,73	0,14
Thr	ACT	358,00	7,42	0,14
Thr	ACC	1502,00	31,13	0,57
Trp	TGG	652,00	13,51	1,00
End	TGA	109,00	2,26	0,55
Cys	TGT	325,00	6,74	0,32
Cys	TGC	706,00	14,63	0,68
End	TAG	42,00	0,87	0,21
End	TAA	46,00	0,95	0,23
Tyr	Tat	360,00	7,46	0,26
Tyr	Tac	1042,00	21,60	0,74
Leu	TTG	313,00	6,49	0,06
Leu	TTA	76,00	1,58	0,02
Phe	TTT	336,00	6,96	0,20
Phe	TTC	1377,00	28,54	0,80
Ser	TCG	325,00	6,74	0,09
Ser	TCA	165,00	3,42	0,05
Ser	TCT	450,00	9,33	0,13
Ser	TCC	958,00	19,86	0,28

(continuação)

Aminoácido	Codão	Número	/1000	Fracção
Arg	CGG	611,00	12,67	0,21
Arg	CGA	183,00	3,79	0,06
Arg	CGT	210,00	4,35	0,07
Arg	CGC	1086,00	22,51	0,37
Gln	CAG	2020,00	41,87	0,88
Gln	CAA	283,00	5,87	0,12
His	Cat	234,00	4,85	0,21
His	CAC	870,00	18,03	0,79
Leu	CTG	2884,00	59,78	0,58
Leu	CTA	166,00	3,44	0,03
Leu	CTT	238,00	4,93	0,05
Leu	Ctc	1276,00	26,45	0,26
Pro	CCG	482,00	9,99	0,17
Pro	CCA	456,00	9,45	0,16
Pro	Cct	568,00	11,77	0,19
Pro	Ccc	1410,00	29,23	0,48

Uma sequência de ADN codificando as proteínas de fusão ou a proteína de ligação à colina modificada da presente invenção pode ser sintetizada utilizando técnicas convencionais de síntese de ADN, tal como por ligação enzimática como descrito por D.M. Roberts *et al.* em *Biochemistry* 1985, 24, 5090-5098, por síntese química, por polimerização enzimática *in vitro* ou por tecnologia de PCR utilizando, por exemplo, uma polimerase termoestável ou por uma combinação destas técnicas.

A polimerização enzimática de ADN pode ser realizada *in vitro* utilizando uma ADN polimerase, tal como a ADN

polimerase I (fragmento *Klenow*) ou Taq polimerase num tampão apropriado contendo os trifosfatos de nucleósido dATP, dCTP, dGTP e dTTP como requerido a uma temperatura de 10-37 °C, geralmente num volume de 50 µL ou menos. A ligação enzimática de fragmentos de ADN pode ser realizada utilizando uma ADN ligase, tal como a T4 ADN ligase num tampão apropriado, tal como 0,05 M de Tris (pH 7,4), 0,01 M de MgCl₂, 0,01 M de ditionitroreitol, 1 mM de espermidina, 1 mM de ATP e 0,1 m/mL de albumina de soro bovino, a uma temperatura de 4 °C até à ambiente, geralmente num volume de 50 µL ou menos. A síntese química do polímero ou fragmento de ADN pode ser realizada por química convencional de fosfotriéster, fosfato ou fosforamidite, utilizando técnicas de fase sólida, tais como aquelas descritas em "Chemical and Enzymatic Synthesis of Gene Fragments - A Laboratory Manual" (ed. H.G. Gassen e A. Lang), Verlag Chemie, Weinheim (1982) ou outras publicações científicas, por exemplo, M.J. Gait, H.W.D. Matthes, M. Singh, B.S. Sproat, e R.C. Titmas, *Nucleic Acids Research*, 1982, 10, 6243; B.S. Sproat, e W. Bannwarth, *Tetrahedron Letters*, 1983, 24, 5771; M.D. Matteucci e M.H. Caruthers, *Tetrahedron Letters*, 1980, 21, 719; M.D. Matteucci e M.H. Caruthers, *Journal of the American Chemical Society*, 1981, 103, 3185; S.P. Adams *et al.*, *Journal of the American Chemical Society*, 1983, 105, 661; N.D. Sinha, J. Biernat, J. McMannus, e H. Koester, *Nucleic Acids Research*, 1984, 12, 4539; e H.W.D. Matthes *et al.*, *EMBO Journal*, 1984, 3, 801.

O processo da invenção pode ser realizado através de técnicas recombinantes convencionais, tais como descritas em Maniatis *et al.*, *Molecular Cloning - A Laboratory Manual*; Cold Spring Harbor, 1982-1989.

Em particular, o processo pode compreender os passos de:

i) preparação de um vector de expressão replicável ou integrante capaz de expressar, numa célula hospedeira, um polímero de ADN compreendendo uma sequência de nucleótidos que codifique a proteína ou um seu derivado imunogénico

ii) transformação de uma célula hospedeira com o referido vector

iii) cultivar a referida célula hospedeira transformada sob condições que permitem a expressão do referido polímero de ADN para produzir a referida proteína; e

iv) recuperação da referida proteína

O termo "transformação" é aqui utilizado para demonstrar a introdução de ADN estranho numa célula hospedeira. Isto pode ser realizado, por exemplo, por transformação, transfecção ou infecção com um vector plasmídico ou viral apropriado utilizando, e. g., técnicas convencionais como descritas em Genetic Engineering; Eds. S.M. Kingsman e A.J. Kingsman; Blackwell Scientific Publications; Oxford, Inglaterra, 1988. O termo "transformado" ou "transformante" aplicar-se-á daqui por diante à célula hospedeira resultante contendo e expressando o gene estranho de interesse.

Os vectores de expressão são novos e também fazem parte da invenção.

Os vectores de expressão replicáveis podem ser preparados de acordo com a invenção, por clivagem de um vector compatível com a célula hospedeira para proporcionar um segmento de ADN linear que tem um replicação intacto, e combinação do referido segmento linear com uma ou mais moléculas de ADN que, conjuntamente com o referido segmento linear codificam o produto desejado, tal como

o polímero de ADN que codifica a proteína da invenção ou os seus derivados, sob condições de ligação.

Assim, o polímero de ADN pode ser realizado ou formado durante a construção do vector, como desejado.

A escolha do vector será determinada em parte pela célula hospedeira, que pode ser procariótica ou eucariótica mas é, de um modo preferido, *E. coli*, levedura ou células CHO. Os vectores adequados incluem plasmídeos, bacteriófagos, cosmídeos e vírus recombinantes. Os vectores de expressão e clonagem contêm, de um modo preferido, um marcador de selecção de modo que apenas irão sobreviver as células hospedeiras que expressam o marcador sob as condições selectivas. Os genes de selecção incluem, mas não estão limitados, àqueles que codificam a proteína que confere uma resistência à ampicilina, tetraciclina ou canamicina. Os vectores de expressão também contêm sequências de controlo que são compatíveis com o hospedeiro designado. Por exemplo, as sequências de controlo de expressão para *E. coli*, e de um modo mais geral para procariotas, incluem promotores e sítios de ligação ao ribossoma. As sequências promotoras podem ser naturais, tais como a β -lactamase (penicilinase) (Weissman 1981, em *Interferon 3* (ed. L. Gresser), lactose (lac) (Chang et al. Nature, 1977, 198: 1056) e triptofano (trp) (Goeddel et al. Nucl. Acids Res. 1980, 8, 4057) e sistema promotor P_L derivado de lambda. Além disso, os promotores sintéticos que não ocorrem na natureza funcionam também como promotores bacterianos. Este é o caso, por exemplo, do promotor híbrido sintético tac que é derivado de sequências dos promotores trp e lac (De Boer et al., Proc. Acad Natl Sci. USA 1983, 80, 21-26). Estes sistemas são particularmente adequados com *E. coli*.

Os vectores compatíveis com levedura também contêm marcadores que permitem a selecção de transformantes bem

sucedidos, por proporcionarem prototrofia a mutantes auxotróficos ou resistência a metais pesados em estirpes de tipo selvagem. As sequências de controlo de expressão para vectores de levedura incluem promotores para enzimas glicolíticas (Hess *et al.*, J. Adv. Enzyme Reg. 1968, 7, 149), gene PH05 codificando a fosfatase ácida, gene CUP1, gene ARG3, promotores de genes GAL e sequências promotoras sintéticas. Outros elementos de controlo úteis na expressão de levedura são terminadores e sequências líder de ARNm. A sequência codificadora 5' é particularmente útil uma vez que codifica tipicamente um péptido sinal compreendido de aminoácidos hidrofóbicos que direccionam a secreção da proteína da célula. As sequências sinal adequadas podem ser codificadas por genes para proteínas de levedura excretadas, tal como o gene da invertase de levedura e o gene do factor a, fosfatase ácida, toxina assassina, o gene do factor alfa-acasalamento e recentemente a sequência sinal da inulinase heteróloga derivada do INUL1A de *Kluyveromyces marxianus*. Foram desenvolvidos vectores adequados para expressão em *Pichia pastoris* e *Saccharomyces cerevisiae*.

Estão disponíveis vários vectores de expressão de *P. pastoris* baseados em vários promotores indutíveis ou constitutivos (Cereghino e Cregg, FEMS Microbiol. Rev. 2000, 24:45-66). Para a produção de proteínas citosólicas e excretadas, os vectores de *P. pastoris* mais habitualmente utilizados contêm o promotor muito forte e firmemente regulado da álcool oxidase (AOX1). Os vectores também contêm o gene da histidinol desidrogenase de *P. pastoris* (HIS4) para a selecção em hospedeiros his4. A secreção da proteína estranha requer a presença de uma sequência sinal e a sequência sinal do factor alfa-acasalamento prepro de *S. cerevisiae* tem sido amplamente utilizado com sucesso no sistema de expressão de *Pichia*. Os vectores de expressão são integrados no genoma de *P. pastoris*

para maximizar a estabilidade de estirpes de expressão. Como em *S. cerevisiae*, a clivagem de um vector de expressão de *P. pastoris* dentro de uma sequência partilhada pelo genoma hospedeiro (AOX1 ou HIS4) estimula os eventos de recombinação homóloga que direccionam eficientemente a integração do vector nesse locus genómico. Em geral, uma estirpe recombinante que contém múltiplas cópias integradas de uma cassette de expressão pode produzir mais proteína heteróloga do que uma estirpe de cópia única. O modo mais eficaz para obter transformantes de elevado número de cópias requer a transformação de uma estirpe receptora de *Pichia* através da técnica do esferoplasto (Cregg *et al.* 1985, Mol. Cell. Biol. 5: 3376-3385).

A preparação do vector de expressão replicável pode ser realizada convencionalmente com enzimas apropriadas para restrição, polimerização e ligação de ADN, através de processos descritos em, por exemplo, Maniatis *et al.* acima referido.

A célula hospedeira recombinante é preparada, de acordo com a invenção, por transformação de uma célula hospedeira com um vector de expressão replicável da invenção sob condições de transformação. As condições de transformação adequadas são convencionais e estão descritas em, por exemplo, Maniatis *et al.* acima referido ou "DNA Cloning" Vol. II, D.M. Glover ed., IRL Press Ltd, 1985.

A escolha das condições de transformação depende da escolha da célula hospedeira a ser transformada. Por exemplo, a transformação *in vivo* utilizando um vector viral vivo como o agente de transformação para os polinucleótidos da invenção é descrita acima. A transformação bacteriana de um hospedeiro, tal como *E. coli*, pode ser realizada através de incorporação directa dos polinucleótidos (que podem ser vectores de expressão

contendo a sequência desejada) depois do hospedeiro ser tratado com uma solução de CaCl_2 (Cohen *et al.*, Proc. Nat. Acad. Sci., 1973, 69, 2110) ou com uma solução compreendendo uma mistura de cloreto de rubídio (RbCl), MnCl_2 , acetato de potássio e glicerol, e depois com ácido 3-[N-morfolino]-propano-sulfónico, RbCl e glicerol ou através de electroporação. A transformação de organismos eucarióticos inferiores, tal como células de levedura em cultura através de incorporação directa pode ser realizada, por exemplo, através da utilização do método Hinnen *et al.* (Proc. Natl. Acad. Sci. 1978, 75: 1929-1933). As células de mamíferos em cultura podem ser transformadas utilizando a coprecipitação com fosfato de cálcio do ADN do vector sobre as células (Graham e Van der Eb, Virology 1978, 52, 546). Outros métodos para introdução de polinucleótidos em células de mamífero incluem transfecção mediada por dextrano, transfecção mediada por polibreno, fusão de protoplastos, electroporação, encapsulação do(s) polinucleótido(s) em lipossomas e microinjecção directa de polinucleótidos em núcleos.

A invenção também se estende a uma célula hospedeira transformada com um ácido nucleico codificando a proteína da invenção ou um vector de expressão replicável da invenção.

O cultivo da célula hospedeira transformada sob condições que permitem a expressão do polímero de ADN é realizada convencionalmente, como descrito em, por exemplo, Maniatis *et al.* e "DNA Cloning", referidos acima. Assim, de um modo preferido a célula é fornecida com nutriente e cultivada a uma temperatura abaixo de 50 °C, de um modo preferido entre 25 °C e 42 °C, de um modo mais preferido entre 25 °C e 35 °C, de um modo muito preferido a 30 °C. O tempo de incubação pode variar desde alguns minutos até algumas horas, de acordo com a proporção do polipéptido na célula bacteriana, como avaliado por SDS-PAGE ou transferência de Western.

O produto pode ser recuperado através de métodos convencionais, de acordo com a célula hospedeira e de acordo com a localização do produto de expressão (intracelular ou excretado no meio de cultura ou no periplasma da célula). Assim, no caso em que a célula hospedeira é bacteriana, tal como *E. coli*, pode, por exemplo, ser lisada fisicamente, quimicamente ou enzimaticamente e o produto proteico isolado do lisado resultante. No caso em que a célula hospedeira é de mamífero, o produto pode ser geralmente isolado do meio nutriente ou de extractos sem células. No caso em que a célula hospedeira é uma levedura, tal como *Saccharomyces cerevisiae* ou *Pichia pastoris*, o produto pode ser geralmente isolado a partir de células lisadas ou do meio de cultura, e depois ainda purificado utilizando técnicas convencionais. A especificidade do sistema de expressão pode ser avaliada através de transferência de Western ou por ELISA utilizando um anticorpo dirigido contra o polipéptido de interesse.

As técnicas convencionais de isolamento de proteína incluem precipitação selectiva, cromatografia de adsorção, e cromatografia de afinidade, incluindo uma coluna de afinidade de anticorpo monoclonal. Quando as proteínas da presente invenção são expressas com uma extremidade de histidina (marcador His), estas podem ser facilmente purificadas através de cromatografia de afinidade utilizando uma coluna de cromatografia de afinidade de ião metálico (IMAC). O ião metálico pode ser qualquer ião adequado, por exemplo, zinco, níquel, ferro, magnésio ou cobre, mas, de um modo preferido, é zinco ou níquel. De um modo preferido, o tampão de IMAC contém detergente, de um modo preferido um detergente aniónico, tal como SDS, de um modo mais preferido um detergente não iónico, tal como Tween 80 ou um detergente anfotérico, tal como Empigen BB, uma vez que isto

pode resultar em níveis inferiores de endotoxina no produto final.

Os passos cromatográficos adicionais incluem, por exemplo, um passo de Q-Sepharose que pode ser realizado antes ou depois da coluna de IMAC. De um modo preferido, o pH está na escala de 7,5 até 10, de um modo mais preferido desde 7,5 até 9,5, optimamente entre 8 e 9.

As proteínas da invenção podem ser assim purificadas de acordo com o seguinte protocolo. Após ruptura das células, os extractos celulares contendo a proteína podem ser solubilizados num tampão Tris a pH 8,5 contendo ureia (por exemplo, 8,0 M) e SDS (por exemplo, desde 0,5% a 1%). Após centrifugação, o sobrenadante resultante pode ser depois carregado sobre uma coluna Sepharose FF (Níquel) de IMAC, equilibrada com um tampão Tris a pH 8,5. A coluna pode ser depois lavada com um tampão contendo sal elevado (e. g., 0,75 - 1,5 M de NaCl, 15 mM de tampão Tris a pH 8,5). A coluna pode ser depois opcionalmente lavada, novamente com tampão de fosfato sem sal. As proteínas da invenção podem ser eluídas da coluna com uma solução tamponada contendo imidazole. As proteínas podem ser depois submetidas a um passo cromatográfico adicional, tal como uma cromatografia de permuta aniónica (por exemplo, Q Sepharose).

As proteínas da presente invenção são proporcionadas solúveis numa forma líquida ou numa forma liofilizada, a qual é a forma preferida. É geralmente esperado que cada dose humana irá compreender 1 a 1000 µg de proteína, e de um modo preferido 30-300 µg. O processo de purificação pode também incluir um passo de carboxiamidação em que a proteína é primeiro reduzida na presença de Glutathione e depois carboximetilada na presença de iodoacetamida. Este passo oferece a vantagem de controlo da agregação oxidativa da molécula com ela própria ou com

contaminantes proteicos da célula hospedeira através de ligação covalente com ligações dissulfureto.

A presente invenção também proporciona composições farmacêuticas e imunogénicas compreendendo uma proteína da presente invenção num excipiente farmaceuticamente aceitável. Uma composição de vacina preferida compreende, pelo menos, uma proteína de acordo com a invenção. A referida proteína tem, de um modo preferido, grupos tiol bloqueados e está altamente purificada, e. g., tem menos do 5% de contaminação da célula hospedeira. Essa vacina pode conter opcionalmente um ou mais outros antígenos associados a tumores e derivados. Por exemplo outro antígeno associado adequado inclui a prostase, PAP-1, PSA (antígeno específico da próstata), PSMA (antígeno de membrana específico da próstata), PSCA (Antígeno da Célula Estaminal da Próstata), STEAP.

Noutra forma de realização, as composições imunogénicas ilustrativas, tais como, por exemplo, composições de vacina, da presente invenção compreendem ADN codificando um ou mais dos polipéptidos de fusão como acima descritos, de modo que o polipéptido de fusão seja gerado *in situ*. Como assinalado acima, o polinucleótido pode ser administrado dentro de qualquer um de uma variedade de sistemas de distribuição conhecidos dos especialistas na técnica. De facto, são bem conhecidas na técnica numerosas técnicas de transferência génica, tal como aquelas descritas por Rolland, *Crit. Rev. Therap. Drug carrier Systems* 15:143-198, 1998, e referências aí citadas. Os sistemas de expressão de polinucleótido apropriados irão, naturalmente, conter as sequências reguladoras de ADN reguladoras necessárias para expressão num doente (tal como um promotor e sinal de terminação adequado). Alternativamente, os sistemas de distribuição bacterianos podem envolver a administração de uma bactéria (tal como o *Bacillus-Calmette-Guerrin*) que expressa uma

porção imunogénica do polipéptido na sua superfície celular ou segrega esse epitopo.

Deste modo, em determinadas formas de realização, os polinucleótidos codificando polipéptidos imunogénicos aqui descritos são introduzidos em células hospedeiras adequadas de mamífero para expressão utilizando qualquer de um número de sistemas baseados em vírus conhecidos. Numa forma de realização ilustrativa, o retrovírus proporciona uma plataforma conveniente e eficaz para sistemas de transferência génica. Uma sequência de nucleótidos seleccionada codificando um polipéptido da presente invenção pode ser introduzida num vector e empacotada em partículas retrovirais utilizando as técnicas conhecidas na matéria. O vírus recombinante pode ser depois isolado e transferido a um individuo. Têm sido descritos uma variedade de sistemas retrovirais ilustrativos (e. g., Pat. U.S. N° 5219740; Miller e Rosman (1989) *BioTechniques* 7:980-990; Miller, A. D. (1990) *Human Gene Therapy* 1:5-14; Scarpa *et al.* (1991) *Virology* 180:849-852; Bums *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:8033-8037; e Boris-Lawrie e Temin (1993) *Cur. Opin. Genet. Develop.* 3:102-109.

Além disso, foi também descrita uma variedade de sistemas baseados em adenovírus ilustrativos. Ao contrário dos retrovírus que se integram no genoma do hospedeiro, os adenovírus persistem extracromossomicamente, minimizando assim os riscos associados com a mutagénese por inserção (Haj-Ahmad e Graham (1986) *J. Virol.* 57:267-274; Bett *et al.* (1993) *J. Virol.* 67:5911-5921; Mittereder *et al.* (1994) *Human Gene Therapy* 5:717-729; Seth *et al.* (1994) *J. Virol.* 68:933-940; Barr *et al.* (1994) *Gene Therapy* 1:51-58; Berkner, K. L. (1988) *BioTechniques* 6:616-629; e Rich *et al.* (1993) *Human Gene Therapy* 4:461-476). Uma vez que os humanos são por vezes infectados com serotipos de adenovírus

humanos comuns, tal como AdHu5, uma proporção significativa da população tem uma resposta de anticorpo neutralizante do adenovírus, que é provável que afecte a resposta imunológica a um antigénio heterólogo num sistema baseado em vacina recombinante. Os vectores adenovirais de primatas não humanos, tal como o adenovírus 68 de chimpanzé (AdC68, Fitzgerald *et al.* (2003) *J. Immunol* 170(3):1416-22)) podem oferecer um sistema adenoviral alternativo sem a desvantagem de uma resposta de anticorpo neutralizante pré-existente.

Foram também desenvolvidos vários sistemas de vector de vírus adeno-associados (AAV) para a distribuição de polinucleótido. Os vectores de AAV podem ser prontamente construídos utilizando técnicas bem conhecidas na matéria. Ver, e. g., Pat. U.S. N° 5173414 e 5139941; Publicação Internacional N° WO 92/01070 e WO 93/03769; Lebkowski *et al.* (1988) *Molec. Cell. Biol.* 8:3988-3996; Vincent *et al.* (1990) *Vaccine* 90 (Cold Spring Harbor Laboratory Press); Carter, B. J. (1992) *Current Opinion in Biotechnology* 3:533-539; Muzyczka, N. (1992) *Current Topics in Microbiol. and Immunol.* 158:97-129; Kotin, R. M. (1994) *Human Gene Therapy* 5:793-801; Shelling e Smith (1994) *Gene Therapy* 1:165-169; e Zhou *et al.* (1994) *J. Exp. Med.* 179:1867-1875.

Os vectores virais adicionais úteis para a transferência de moléculas de ácidos nucleicos codificando polipéptidos da presente invenção, através de transferência génica, incluem aqueles derivados da família pox de vírus, tais como vírus vaccinia e poxvírus de aves. A título de exemplo, podem ser construídos vírus vaccinia recombinantes que expressam as novas moléculas como se segue. O ADN codificando um polipéptido é primeiro introduzido num vector apropriado de modo que esteja adjacente a um promotor de vaccinia e a sequências de ADN de vaccinia flanqueadoras, tal como a sequência codificando a

timidina cinase (TK). Este vector é depois utilizado para transfectar células que são simultaneamente infectadas com vaccinia. A recombinação homóloga serve para introduzir o promotor de vaccinia mais o gene codificando o polipéptido de interesse no genoma viral. O TK.sup.(-) recombinante resultante pode ser seleccionado através de cultivo das células na presença de 5-bromodesoxiuridina e escolhendo as placas virais aí resistentes.

Um sistema de infecção/transfecção baseado em vaccinia pode ser utilizado convenientemente para proporcionar expressão ou coexpressão transitória, indutível, de um ou mais polipéptidos aqui descritos em células hospedeiras de um organismo. Neste sistema particular, as células são primeiro infectadas *in vitro* com um vírus vaccinia recombinante que codifica a T7 ARN polimerase de bacteriófago. Esta polimerase apresenta especificidade subtil uma vez que apenas transcreve moldes contendo promotores T7. Depois da infecção, as células são transfectadas com o polinucleótido ou polinucleótidos de interesse, movidas por um promotor T7. A polimerase expressa no citoplasma a partir do vírus vaccinia recombinante transcreve o ADN transfectado em ARN que é depois traduzido no polipéptido pela maquinaria de tradução do hospedeiro. O método proporciona uma produção citoplasmática transitória de nível elevado, de quantidades grandes de ARN e os seus produtos de tradução. Ver, e. g., Elroy-Stein e Moss, Proc. Natl. Acad. Sci. USA (1990) 87:6743-6747; Fuerst et al. Proc. Natl. Acad. Sci. USA (1986) 83:8122-8126.

Alternativamente, os avipoxvírus, tais como os vírus de *fowlpox* e *canarypox*, podem também ser utilizados para transferir as sequências codificantes de interesse. Os vírus avipox recombinantes, expressando imunogénios de mamíferos patogêneos, são conhecidos por conferirem imunidade protectora quando

administrados à espécies não aviárias. A utilização de um vector de avipox é particularmente desejável em humanos e outras espécies de mamíferos uma vez que os membros do género Avipox podem apenas replicar produtivamente em espécies aviárias susceptíveis e, deste modo, não infectam células de mamíferos. Os métodos para produzir vírus Avipox recombinantes são conhecidos na técnica e utilizam recombinação genética, como acima descrito relativamente à produção de vírus vaccinia. Ver, e. g., documentos WO 91/12882; WO 89/03429; e WO 92/03545.

Qualquer um de uma variedade de vectores alfavírus também pode ser utilizado para a distribuição de composições de polinucleótido da presente invenção, como aqueles vectores descritos nas Patentes U.S. Nº 5843723; 6015686; 6008035 e 6015694. São também utilizados determinados vectores baseados na Encefalite Equina Venezuelana (VEE), cujos exemplos ilustrativos podem ser encontrados nas Patentes U.S. Nº 5505947 e 5643576.

As composições da presente invenção também podem ser distribuídas através de uma variedade de vias, tais como intramuscularmente, subcutaneamente, intraperitonealmente ou intravenosamente.

Noutra forma de realização da invenção, um polinucleótido é administrado/transferido como ADN "nú" como, por exemplo, descrito em Ulmer *et al.*, *Science* 259:1745-1749, 1993 e revisto por Cohen, *Science* 259:1691-1692, 1993. A incorporação de ADN nú pode ser aumentada através do revestimento de esferas biodegradáveis com o ADN, que são eficientemente transportadas para as células. Numa forma de realização preferida, a composição é distribuída intradermicamente. Em particular, a composição é distribuída através de técnicas de administração com pistola de genes (particularmente bombardeamento de partículas) que envolve o revestimento do vector sobre uma

esfera (e. g. ouro) que são depois administradas sob alta pressão na epiderme; tal como, por exemplo, descrito em Haynes *et al.*, J Biotechnology 44: 37-42 (1996).

Num exemplo ilustrativo, a aceleração das partículas impelidas por gás é realizada com dispositivos, tais como aqueles produzidos pela Powderject Pharmaceuticals PLC (Oxford, Reino Unido) e Powderject Vaccines Inc. (Madison, WI), alguns exemplos dos quais são descritos nas Patentes U.S. N° 5846796; 6010478; 5865796; 5584807; e Patente EP N° 0500799. Esta abordagem oferece uma abordagem de distribuição sem agulha em que uma formulação de pó seco de partículas microscópicas, tal como o polinucleótido, é acelerada a alta velocidade dentro de um jacto de gás hélio produzido por um dispositivo agarrado manualmente, propulsionando as partículas num tecido alvo de interesse, tipicamente a pele. As partículas são, de um modo preferido, esferas de ouro de diâmetro de 0,4 - 4,0 μm , de um modo mais preferido 0,6 - 2,0 μm e o ADN conjugado revestido sobre estas e depois encerradas num cartucho ou cassete para colocação na "pistola de genes".

Numa forma de realização relacionada outros dispositivos e métodos que podem ser úteis para a injeção impelida por gás e sem agulha de composições da presente invenção incluem aqueles proporcionado pela Bioject, Inc. (Portland, OR), alguns exemplos dos quais são descritos nas Patentes U.S. N° 4790824; 5064413; 5312335; 5383851; 5399163; 5520639 e 5993412.

É possível para o componente imunogénico compreendendo a sequência de nucleótidos codificando o péptido antigénico, a ser administrado numa base de uma só vez ou a ser administrado repetidamente, por exemplo, entre 1 e 7 vezes, de um modo preferido entre 1 e 4 vezes, em intervalos entre cerca de 1 dia e cerca de 18 meses. Contudo, este regime do tratamento irá

varias significativamente dependendo do tamanho do doente, da doença que está a ser tratada/protegida contra, da quantidade de sequência de nucleótidos administrada, da via de administração, e de outros factores que seriam evidentes a um médico assistente especialista.

É deste modo outro aspecto da presente invenção proporcionar a utilização de uma proteína ou um ADN codificando a referida proteína, como aqui descrita, na produção de uma composição imunogénica para induzir uma resposta imunológica num doente. De um modo preferido, a resposta imunológica é para ser induzida através de administração sequencial i) da referida proteína seguida pela referida sequência de ADN; ou ii) a referida sequência de ADN seguida pela referida proteína. De um modo mais preferido, a sequência de ADN é revestida em esferas biodegradáveis ou distribuída através de uma abordagem de bombardeamento de partículas. De um modo ainda mais preferido, a proteína é adjuvantada, de um modo preferido com um adjuvante de indução de TH-1, de um modo preferido com uma formulação adjuvante baseada em CpG/QS21.

Os vectores que compreendem as sequências de nucleótidos codificando péptidos antigénicos são administrados numa quantidade que será profilaticamente ou terapêuticamente eficaz. A quantidade a ser administrada é geralmente na escala de um picograma até 16 miligrama, de um modo preferido 1 picograma até 10 microgramas para distribuição mediada por partícula, e 10 microgramas até 16 miligrama para outras vias de nucleótido por dose. A quantidade exacta pode variar consideravelmente dependendo do peso do doente a ser imunizado e da via de administração.

As técnicas adequadas para introduzir o vector ou polinucleótido num doente também incluem a aplicação tópica

com um veículo apropriado. O ácido nucleico pode ser administrado topicamente à pele ou a superfícies mucósicas, por exemplo, através de administração intranasal, oral, intravaginal ou intra-rectal. O vector ou polinucleótido nu podem estar presentes conjuntamente com um excipiente farmacologicamente aceitável, tal como soro fisiológico tamponado com fosfato (PBS). A incorporação de ADN pode ser ainda facilitada através da utilização de agentes facilitadores, tal como bupivacaína, separadamente ou incluída na formulação de ADN. Outros métodos de administração do ácido nucleico directamente a um receptor incluem ultra-sons, estimulação eléctrica, electroporação e micro-inoculação que é descrita no documento U.S. 5697901.

A incorporação de construções de ácidos nucleicos pode ser intensificada através de diversas técnicas de transfecção conhecidas, por exemplo, aquelas que incluem a utilização de agentes de transfecção. Os exemplos destes agentes incluem agentes catiónicos, por exemplo, fosfato de cálcio e DEAE-Dextrano e lipofectantes, por exemplo, lipofectam e transfectam. A dosagem do ácido nucleico a ser administrado pode ser alterada.

As proteínas de fusão e polipéptidos codificantes de acordo com a invenção podem também ser formulados como uma composição farmacêutica/imunogénica, e. g., como uma vacina. Deste modo, de acordo com o exposto, a presente invenção também proporciona uma composição farmacêutica/imunogénica compreendendo uma proteína de fusão da presente invenção num excipiente farmacologicamente aceitável. De acordo com o exposto, é também proporcionado um processo para a preparação de uma composição imunogénica de acordo com a presente invenção, compreendendo a mistura da proteína de fusão da invenção ou o polinucleótido codificante com um adjuvante, diluente ou outro veículo farmacologicamente aceitável adequado.

As proteínas de fusão da presente invenção são proporcionadas, de um modo preferido, pelo menos, 80% puras, de um modo mais preferido, 90% puras como visualizado através de SDS-PAGE. De um modo preferido, as proteínas aparecem como uma única banda através de SDS-PAGE.

A preparação de vacina é geralmente descrita em Vaccine Design ("The subunit and adjuvant approach" (eds. Powell M.F. e Newman M.J). (1995) Plenum Press Nova Iorque). A encapsulação em lipossomas é descrita por Fullerton, Patente U.S. 4235877.

As proteínas de fusão da presente invenção e os polinucleótidos codificantes são, de um modo preferido, adjuvantados na formulação de vacina da invenção. Determinados adjuvantes estão comercialmente disponíveis como, por exemplo, Adjuvante Incompleto de Freund e Adjuvante Completo (Difco Laboratories, Detroit, MI); Adjuvante 65 Merck (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Filadélfia, PA); sais de alumínio, tais como gel de hidróxido de alumínio (alúmen) ou fosfato de alumínio; sais de cálcio, ferro ou zinco; uma suspensão insolúvel de tirosina acilada; açúcares acilados; polissacáridos derivados cationicamente ou anionicamente; polifosfazenos; microesferas biodegradáveis; monofosforil-lípido A e quil A. Citocinas, tais como GM-CSF, interleucina-2, -7, -12 e outros factores de crescimento semelhantes, também podem ser utilizados como adjuvantes.

Dentro de determinadas formas de realização da invenção, a composição adjuvante é, de um modo preferido, uma que induz uma resposta imunológica predominantemente de tipo Th1. Níveis elevados de citocinas de tipo Th1 (e. g., IFN- γ , TNF α , IL-2 e IL-12) tendem a favorecer a indução de respostas imunológicas mediadas por célula a um antigénio administrado. Em contraste,

níveis elevados de citocinas de tipo Th2 (e. g., IL-4, IL-5, IL-6 e IL-10) tendem a favorecer a indução de respostas imunológicas humorais. Após aplicação de uma vacina como aqui proporcionada, um doente suportará uma resposta imunológica que inclua respostas de tipo Th1 e Th2. Dentro de uma forma de realização preferida, em que uma resposta é predominantemente do tipo Th1, o nível de citocinas de tipo Th1 aumentará em maior grau do que o nível de citocinas de tipo Th2. Os níveis destas citocinas podem ser prontamente avaliados utilizando ensaios convencionais. Para uma revisão das famílias de citocinas, ver Mosmann e Coffman, *Ann. Rev. Immunol.* 7:145-173, 1989.

Os adjuvantes indutores de Th-1 preferidos são seleccionados do grupo de adjuvantes compreendendo: 3D-MPL, QS21, uma mistura de QS21 e colesterol, e um oligonucleótido CpG ou uma mistura de dois ou mais dos referidos adjuvantes. Determinados adjuvantes preferidos para induzir uma resposta predominantemente de tipo Th1 incluem, por exemplo, uma combinação de monofosforil-lípido A, de um modo preferido monofosforil-lípido A 3-de-O-acilado, conjuntamente com um sal de alumínio. Os adjuvantes MPL® estão disponíveis da Corixa Corporation (Seattle, WA; ver, por exemplo, Patentes U.S. N° 4436727; 4877611; 4866034 e 4912094). Os oligonucleótidos contendo CpG (em que o dinucleótido CpG está não metilado) induzem também predominantemente uma resposta Th1. Esses oligonucleótidos são bem conhecidos e são descritos, por exemplo, nos documentos WO 96/02555, WO 99/33488 e Patentes U.S. N° 6008200 e 5856462. São também descritas sequências de ADN imunoestimuladoras, por exemplo, por Sato *et al.*, *Science* 273:352, 1996. Outro adjuvante preferido compreende uma saponina, tal como Quil A ou os seus derivados, incluindo QS21 e QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escina; Digitonina; ou saponinas de *Gypsophila* ou *Chenopodium quinoa*. Outras formulações preferidas incluem mais do que uma saponina nas combinações adjuvantes da presente invenção, por exemplo,

combinações de, pelo menos, dois do seguinte grupo compreendendo QS21, QS7, Quil A, β -escina ou digitonina.

Alternativamente, as formulações de saponina podem ser combinadas com veículos de vacina compostos de quitosano ou outros polímeros policatiônicos, partículas de polilactida e polilactida-co-glicolida, matriz polimérica baseada em poli-N-acetil-glucosamina, partículas compostas de polissacáridos ou polissacáridos quimicamente modificados, lipossomas e partículas baseadas em lípido, partículas compostas de monoésteres de glicerol, etc. As saponinas podem também ser formuladas na presença de colesterol para formar estruturas particuladas, tais como lipossomas ou ISCOM. Além disso, as saponinas podem ser formuladas conjuntamente com um éter ou éster de polioxietileno, numa solução não particulada ou suspensão ou numa estrutura particulada, tal como um lipossoma *paucilamelar* ou ISCOM. As saponinas também podem ser formuladas com excipientes, tal como Carbopol^R para aumentar a viscosidade ou podem ser formuladas numa forma de pó seco com um excipiente em pó, tal como a lactose.

Numa forma de realização preferida, o sistema adjuvante inclui a combinação de um monofosforil-lípido A e 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 QS21 é neutralizada com colesterol, como descrito no documento WO 96/33739. Outras formulações preferidas compreendem uma emulsão óleo-em-água e tocoferol. Uma outra formulação adjuvante particularmente preferida utilizando QS21, adjuvante 3D-MPL[®] e tocoferol numa emulsão óleo-em-água é descrita no documento WO 95/17210.

Outros sistemas adjuvantes intensificados envolvem a combinação de um oligonucleótido contendo CpG e um derivado de

saponina, particularmente a combinação de CpG e QS21 como divulgada nos documentos WO 00/09159 e WO 00/62800. De um modo preferido, a formulação compreende adicionalmente uma emulsão óleo em água e tocoferol.

Ainda noutra forma de realização, a presente invenção proporciona uma composição imunogénica compreendendo uma proteína de fusão de acordo com a invenção, e compreendendo ainda D3-MPL, uma saponina de um modo preferido QS21 e um oligonucleótido CpG, opcionalmente formulado numa emulsão óleo em água.

Os adjuvantes ilustrativos adicionais para utilização nas composições farmacêuticas da invenção incluem Montanide ISA 720 (Seppic, França), SAF (Chiron, Califórnia, Estados Unidos), ISCOMS (CSL), MF-59 (Chiron), a série de adjuvantes SBAS (e. g., SBAS-2 ou SBAS-4, disponíveis da SmithKline Beecham, Rixensart, Bélgica), Detox (Enhanzyn® (Corixa, Hamilton, TA), RC-529 (Corixa, Hamilton, MT) e outros 4-fosfatos de aminoalquilglucosaminida (AGP), como aqueles descritos nos Pedidos pendentes de Patente U.S. N° de Série 08/853826 e 09/074720, cujas divulgações são aqui incorporadas põe referência na sua totalidade, e adjuvantes de éter de polioxietileno, tais como aqueles descritos no documento WO 99/52549A1.

Outros adjuvantes preferidos incluem moléculas adjuvantes 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 de uma formulação de vacina compreendendo um éter de polioxietileno de fórmula geral (I), em que n está 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 C_4-C_{20} e de um modo muito preferido alquilo C_{12} , e A é uma ligação. A concentração dos éteres de polioxietileno devem estar na gama de 0,1-20%, de um modo preferido a partir de 0,1-10%, e de um modo muito preferido na gama de 0,1-1%. Os éteres de polioxietileno preferidos são seleccionados a partir do seguinte grupo: éter de polioxietileno-9-laurilo, éter de polioxietileno-9-estearilo, éter de polioxietileno-8-estearilo, éter de polioxietileno-4-laurilo, éter de polioxietileno-35-laurilo, e éter de polioxietileno-23-laurilo. Os éteres de polioxietileno, tais como éter polioxietileno-laurilo são descritos no índice da Merck (12^a 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 desejado, ser combinado com outro adjuvante. Por exemplo, uma combinação adjuvante preferida é, de um modo preferido, com CpG como descrito no pedido pendente de Patente UK GB 9820956.2.

É uma forma de realização da invenção que os antigénios, incluindo vector de ácido nucleico, da invenção sejam utilizados com agente imunoestimulador. De um modo preferido, o agente imunoestimulador é administrado ao mesmo tempo que os antigénios da invenção e em formas de realização preferidas são formulados conjuntamente. É outra forma de realização da invenção que o antigénio e o agente imunoestimulador (ou vice-versa) sejam administrados sequencialmente aos mesmos sítios ou adjacentes, separados no tempo por períodos entre 0-100 horas. Esses agentes imunoestimuladores incluem, mas não estão limitados a: imidazoquinolinas sintéticas, tais como imiquimod [S-26308, R-837], (Harrison, et al., Vaccine 19: 1820-1826, 2001; e resiquimod [S-28463, R-848] (Vasilakos, et al., Cellular immunology 204: 64-74, 2000; Bases de Schiff de carbonilos e aminas que são constitutivamente expressas na célula apresentadora de antigénio e superfícies de célula T, tais como

tucaresol (Rhodes, J. et al., Nature 377: 71-75, 1995), citocina, quimiocina e moléculas co-estimuladoras como proteína ou péptido, incluindo, por exemplo, citocinas pró-inflamatórias, tais como Interferão, GM-CSF, IL-1 alfa, IL-1 beta, TGF-alfa e TGF-beta, indutores Th1, tais como interferão gama, IL-2, IL-12, IL-15, IL-18 e IL-21, indutores Th2, tais como IL-4, IL-5, IL-6, IL-10 e IL-13 e outros genes de quimiocina e co-estimuladores, tais como MCP-1, MIP-1 alfa, MIP-1 beta, RANTES, TCA-3, CD80, CD86 e CD40L outros ligandos direccionantes imunoestimuladores, tais como CTLA-4 e L-selectina, proteínas e péptidos estimuladores de apoptose, tais como Fas, (49), adjuvantes sintéticos baseados em lípido, tais como vaxfectin, (Reyes et al., Vaccine 19: 3778-3786, 2001) esqualeno, alfa tocoferol, polissorbato 80, DOPC e colesterol, endotoxina, [LPS], (Beutler, B., Current Opinion in Microbiology 3: 23-30, 2000); Oligo- e di-nucleótidos de CpG (Sato, Y. et al., Science 273 (5273): 352-354, 1996; Hemmi, H. et al., Nature 408: 740-745, 2000) e outros ligandos potenciais que activam receptores Toll para produzir citocinas indutoras de Th1, tais como lipoproteínas Micobacterianas sintéticas, proteína Micobacteriana p19, peptidoglicano, ácido teicóico e lípido A.

Outro adjuvante adequado inclui CT (toxina da cólera, subunidades A e B) e LT (enterotoxina lábil ao calor de *E. coli*, subunidades A e B), família das proteínas de choque térmico (HSP), e LLO (listeriolisina O; documento WO 01/72329).

Nos casos em que o agente imunoestimulador é uma proteína, o agente pode ser administrado como uma proteína ou como um polinucleótido codificando a proteína.

Outros sistemas de distribuição adequados incluem microesferas, em que o material antigénico é incorporado ou

conjugado a polímeros/microesferas biodegradáveis de modo que o material antigénico possa ser misturado com um veículo farmacêutico adequado e utilizado como uma vacina. O termo "microesferas" é geralmente utilizado para descrever partículas coloidais que são substancialmente esféricas e têm um diâmetro na gama de 10 nm a 2 mm. As microesferas feitas a partir de uma vastíssima gama de polímeros naturais e sintéticos encontraram utilização numa variedade de aplicações biomédicas. Este sistema de distribuição é especialmente vantajoso para proteínas que têm semi-vidas curtas *in vivo*, requerendo tratamentos múltiplos para proporcionar eficácia ou sendo instáveis em fluidos biológicos ou não totalmente absorvidos a partir do tracto gastrointestinal por causa dos seus pesos moleculares relativamente elevados. Foram descritos diversos polímeros como uma matriz para a libertação de proteína. Os polímeros adequados incluem gelatina, colagénio, alginatos, dextrano. Os sistemas de distribuição preferidos incluem poli(ácido DL-láctico) (PLA), poli(lactida-co-glicolida) (PLG), poli(ácido glicólico) (PGA), poli(ϵ -caprolactona) (PCL), copolímeros de poli(ácido DL-láctico-co-glicólico) (PLGA) biodegradáveis. Outros sistemas preferidos incluem hidrogéis heterogéneos, tais como copolímeros multibloco poli(éter éster), contendo blocos repetitivos baseados em poli(etilenoglicol) (PEG) hidrofílico e poli(tereftalato de butileno) (PBT) hidrofóbico ou poli(etilenoglicol)-tereftalato/poli(tereftalato de butileno) (PEGT/PBT) (Sohier *et al.* Eur. J. Pharm and Biopharm, 2003, 55, 221-228). São preferidos os sistemas que proporcionam uma libertação sustentada durante 1 a 3 meses, tais como PLGA, PLA e PEGT/PBT.

É possível à composição imunogénica ou de vacina ser administrada uma única vez ou, de um modo preferido, ser administrada repetidamente, tantas vezes quanto necessário, por exemplo, entre 1 e 7 vezes, de um modo preferido entre 1 e 4 vezes, em intervalos entre cerca de 1 dia e cerca de 18 meses,

de um modo preferido um mês. Isto pode ser opcionalmente seguido por dosagem em intervalos regulares entre 1 e 12 meses durante um período até ao resto da vida do doente. Numa forma de realização preferida, o doente recebe o antigénio em formas diferentes num regime de "primeiro reforço". Assim, por exemplo, o antigénio, a proteína de fusão, é primeiro administrado como uma formulação de base adjuvante de proteína e depois subsequentemente administrado como uma vacina baseada em ADN. Este modo de administração é preferido. O adjuvante preferido é uma combinação de um oligonucleótido contendo CpG e um derivado de saponina, particularmente a combinação de CpG e QS21 como divulgada nos documentos WO 00/09159 e WO 00/62800. A incorporação de ADN nu pode ser aumentada por revestimento do ADN sobre esferas biodegradáveis, que são transportadas eficientemente para dentro das células. Alternativamente, o ADN pode ser transferido através de uma abordagem de bombardeamento de partículas, por exemplo, aceleração de partículas impelidas por gás com dispositivos, tais como aqueles fabricados pela Powderject Pharmaceuticals PLC (Oxford, Reino Unido) e Powderject Vaccines Inc. (Madison, WI) como aqui ensinado. Esta abordagem oferece uma abordagem de distribuição sem agulha em que uma formulação de pó seco de partículas microscópicas, tais como partículas de polinucleótido ou polipéptido, são aceleradas a alta velocidade dentro de um jacto de hélio gasoso produzido por um dispositivo agarrado manualmente, propulsionando as partículas num tecido alvo de interesse.

Noutra forma de realização preferida, a vacina baseada em ADN será administrada primeiro, seguida pela formulação de base adjuvante de proteína. Ainda outra forma de realização irá referir-se à transferência da construção de ADN através de vectores de transferência especializados, de um modo preferido através de um sistema viral, de um modo muito preferido através de sistemas baseados em adenovírus. Outros sistemas baseados em

vírus adequados de transferência de ADN incluem sistemas baseados em retronvírus, lentivírus, vírus adeno-associados, vírus herpes e vírus vaccinia.

Noutra forma de realização preferida, a formulação de base adjuvante de proteína e a vacina baseada em ADN podem ser co-administradas em sítios adjacentes ou sobrepostos. Dependendo da natureza da formulação de vacina de ADN, isto pode ser realizado por mistura da formulação adjuvante de proteína e ADN antes da administração ou por administração simultânea do ADN e formulação adjuvante da proteína.

O regime de tratamento será significativamente variado dependendo do tamanho e espécie do doente em causa, da quantidade de ácido nucleico de vacina e/ou composição de proteína administrada, da via de administração, da potência e dose de quaisquer compostos adjuvantes utilizados e outros factores que serão evidentes a um médico assistente especialista.

Dentro de aspectos adicionais, a presente invenção proporciona métodos para a estimulação de uma resposta imunológica num doente, de um modo preferido uma resposta de célula T num doente humano, compreendendo a administração de uma composição farmacêutica aqui descrita. O doente pode sofrer de cancro do pulmão ou do cólon ou cancro colorrectal ou cancro da mama, o que nesses casos os métodos proporcionam tratamento para a doença ou um doente considerado em risco para uma tal doença pode ser tratado profilaticamente.

Dentro de aspectos adicionais, a presente invenção proporciona métodos para inibição do desenvolvimento de um cancro num doente, compreendendo a administração a um doente uma composição farmacêutica como referida acima. O doente pode

sofrer de, por exemplo, sarcoma, cancro da próstata, ovário, bexiga, pulmão, cólon, colorrectal ou da mama, o que nesses casos os métodos proporcionam tratamento para a doença ou um doente considerado em risco para uma tal doença pode ser tratado profilaticamente.

A presente invenção proporciona ainda, dentro de outros aspectos, métodos para remoção de células tumorais de uma amostra biológica, compreendendo a colocação em contacto de uma amostra biológica com células T que reagem especificamente com um polipéptido da presente invenção, em que o passo de colocação em contacto é realizado sob condições e durante um tempo suficiente para permitir a remoção de células expressando a proteína da amostra.

Dentro de aspectos relacionados, são proporcionados métodos para a inibição do desenvolvimento de um cancro num doente, compreendendo a administração a um doente uma amostra biológica tratada como acima descrito.

São ainda proporcionados métodos, dentro doutros aspectos, para a estimulação e/ou expansão de células T específicas para um polipéptido da presente invenção, compreendendo a colocação em contacto da célula T com um ou mais de: (i) um polipéptido como acima descrito; (ii) um polinucleótido codificando esse polipéptido; e/ou (iii) uma célula apresentadora de antigénio que expressa esse polipéptido; sob condições e durante um tempo suficiente para permitir a estimulação e/ou expansão de células T. São também proporcionadas populações de células T isoladas compreendendo células T preparadas como acima descrito.

Dentro de aspectos adicionais, a presente invenção proporciona métodos para inibição do desenvolvimento de um cancro num doente, compreendendo a administração a um doente de

uma quantidade eficaz de uma população de células T, como acima descrito.

A presente invenção proporciona ainda métodos para inibição do desenvolvimento de um cancro num doente, compreendendo os passos de: (a) incubação de células T CD4+ e/ou CD8+ isoladas de um doente com um ou mais de: (i) um polipéptido aqui divulgado; (ii) um polinucleótido codificando esse polipéptido; e (iii) uma célula apresentadora de antigénio que expressa esse polipéptido; e (b) administração ao doente de uma quantidade eficaz das células T proliferadas e, desse modo, inibindo o desenvolvimento de um cancro no doente. As células proliferadas podem, mas não necessitam, ser clonadas antes da administração ao doente.

De acordo com uma outra forma de realização desta invenção, uma composição imunogénica aqui descrita é distribuída a um hospedeiro através de células apresentadoras de antigénio (APC), tais como células dendríticas, macrófagos, células B, monócitos e outras células que podem ser manipuladas para serem APC eficientes. Essas células podem, mas não necessitam, ser modificadas geneticamente para aumentar a capacidade de apresentação do antigénio, para melhorar a activação e/ou manutenção da resposta da célula T, para terem efeitos antitumorais *per se* e/ou para serem imunologicamente compatíveis com o receptor (*i. e.*, haplotipo HLA condizente). As APC podem ser geralmente isoladas a partir de qualquer uma variedade de fluidos e órgãos biológicos, incluindo tecidos tumorais e peritumorais, e podem ser células autólogas, alogénicas, singénicas ou xenogénicas.

Determinadas formas de realização preferidas da presente invenção utilizam células dendríticas ou suas progenitoras como células apresentadoras de antigénio. As células dendríticas são APC altamente potentes (Banchereau e Steinman, *Nature* 392:245-

251, 1998) e foi demonstrado que são eficazes como um adjuvante fisiológico para induzir imunidade antitumoral profilática 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 (estreladas *in situ*, com processos citoplasmáticos acentuados (dendrites) visíveis *in vitro*), na sua capacidade para incorporar, processar e apresentar antígenos com eficiência elevada e pela sua capacidade para activar respostas de células T naive. As células dendríticas podem, naturalmente, ser concebidas para expressar ligandos ou receptores específicos de superfície celular que não são habitualmente encontrados em células dendríticas *in vivo* ou *ex vivo*, e essas células dendríticas modificadas estão contempladas na presente invenção. Como uma alternativa às células dendríticas, podem ser utilizadas células dendríticas carregadas de antígeno excretadas em vesículas (denominadas exossomas) dentro de uma vacina (ver Zitvogel et al., *Nature Med.* 4:594-600, 1998).

As células dendríticas e progenitoras podem ser obtidas de sangue periférico, medula óssea, células infiltrantes de tumor, células infiltrantes de tecido peritumoral, nódulos linfáticos, baço, pele, sangue de cordão umbilical ou qualquer outro tecido ou fluido adequado. Por exemplo, as células dendríticas podem ser diferenciadas *ex vivo* por adição de 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, células positivas a CD34 recolhidas de sangue periférico, sangue de cordão umbilical ou medula óssea podem ser diferenciadas em células dendríticas através da 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 induz(em) diferenciação, maturação e proliferação de células dendríticas.

As células dendríticas são convenientemente categorizadas como células "imaturas" e "maduras", o que permite uma forma simples para discriminar entre dois fenótipos bem caracterizados. Contudo, esta nomenclatura não deve ser interpretada para excluir todas as fases intermediárias possíveis de diferenciação. As células dendríticas imaturas são caracterizadas como APC com uma capacidade elevada para incorporação e processamento de antigénio, que se correlaciona com a expressão elevada do receptor Fey e o receptor de manose. O fenótipo maduro é caracterizado tipicamente por uma expressão mais baixa destes marcadores, mas uma expressão elevada das moléculas de superfície celular responsáveis pela activação das células T, tais como MHC de classe I e classe II, moléculas de adesão (e. g., CD54 e CD11) e moléculas coestimuladoras (e. g., CD40, CD80, CD86 e 4-1 BB).

As APC podem geralmente ser transfectadas com um polinucleótido da invenção (ou porção ou outra variante do mesmo) de modo que o polipéptido codificado ou uma sua porção imunogénica, seja expresso na superfície celular. Essa transfecção pode ocorrer *ex vivo*, e uma composição farmacêutica compreendendo essas células transfectadas pode ser depois utilizada para objectivos terapêuticos, como aqui descritos. Alternativamente, um veículo de transferência génica que se direcione para uma célula dendrítica ou outra apresentadora de antigénio pode ser administrada a um doente, resultando em transfecção que ocorre *in vivo*. A transfecção *in vivo* e *ex vivo* de células dendríticas, por exemplo, pode geralmente ser realizada utilizando quaisquer métodos conhecidos na técnica, tais como aqueles descritos no documento WO 97/24447 ou a abordagem da pistola de genes descrita por Mahvi *et al.*, *Immunology and cell Biology* 75:456-460, 1997. O carregamento de antigénio de células dendríticas pode ser realizado por

incubação de células dendríticas ou células progenitoras com o polipéptido tumoral, ADN (nu ou dentro de um vector plasmídico) ou ARN; ou com bactéria recombinante expressando antigénio ou vírus (e. g., vectores de vaccinia, fowlpox, adenovírus ou lentivírus). Antes do carregamento, o polipéptido pode ser covalentemente conjugado a um parceiro imunológico que proporciona ajuda de célula T (e. g., uma molécula de veículo). Alternativamente, uma célula dendrítica pode ser pulsada com um parceiro imunológico não conjugado, separadamente ou na presença do polipéptido.

Definições

São também proporcionados pela invenção métodos para a análise de séries ou sequências de caracteres, particularmente sequências genéticas ou sequências codificadas por proteína. Os métodos preferidos de análise de sequências incluem, por exemplo, métodos de análise de homologia de sequências, tais como análise de identidade e semelhança, análise da estrutura de ADN, ARN e proteína, montagem de sequência, análise cladística, análise de motivos de sequência, determinação de grelha de leitura aberta, *base calling* de ácidos nucleicos, análise de utilização de codão, *base trimming* de ácidos nucleicos e análise de picos do cromatograma de sequenciação.

É proporcionado um método computacional para realização de identificação de homologia. Este método compreende os passos de: proporcionar uma primeira sequência de polinucleótidos compreendendo a sequência de um polinucleótido da invenção num meio de leitura por computador; e comparar a referida primeira sequência de polinucleótido com, pelo menos, uma segunda sequência de polinucleótidos ou polipeptídica para identificar a

homologia. É também proporcionado um método computacional para a realização de identificação de homologia, compreendendo o referido método os passos de: proporcionar uma primeira sequência polipeptídica compreendendo a sequência de um polipéptido da invenção num meio de leitura por computador; e comparar a referida primeira sequência polipeptídica com, pelo menos, uma segunda sequência de polinucleótidos ou polipeptídica para identificar a homologia.

A "identidade," como conhecida na técnica, é uma relação entre duas ou mais sequências polipeptídicas ou duas ou mais sequências de polinucleótidos, conforme o caso, como determinado por comparação das sequências. Na técnica, a "identidade" também significa o grau de parentesco de sequência entre sequências polipeptídicas ou de polinucleótidos, conforme o caso, como determinado pela igualdade entre séries dessas sequências. A "identidade" pode ser prontamente calculada através de métodos conhecidos, incluindo mas não limitados àqueles descritos em (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, Nova Iorque, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, Nova Iorque, 1993; *Computer Analysis of Sequence Data*, Parte I, Griffin, A.M., e Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heine, G., Academic Press, 1987; e *Sequence Analysis Primer*, Gribskov, M. e Devereux, J., eds., M Stockton Press, Nova Iorque, 1991; e Carillo, H., e Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Os métodos para determinar a identidade são concebidos para produzir a maior igualdade entre as sequências testadas. Além disso, os métodos para determinar a identidade são codificados em programas de computador publicamente disponíveis. Os métodos de programas computacionais para determinar a identidade entre duas sequências incluem, mas não estão

limitados, ao programa GAP no pacote de programa GCG (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN (Altschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990), e FASTA (Pearson e Lipman *Proc. Natl. Acad. Sci. USA* 85: 2444-2448 (1988)). A família BLAST de programas está disponível publicamente a partir de NCBI e outras fontes (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990)). O bem conhecido algoritmo de Smith Waterman também pode ser utilizado para determinar a identidade.

Os parâmetros para a comparação de sequências polipeptídicas incluem os seguintes:

Algoritmo: Needleman e Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

Matriz de comparação: BLOSSUM62 de Henikoff e Henikoff, *Proc. Natl. Acad. Sci. USA.* 89:10915-10919 (1992)

Penalidade de Abertura: 8

Penalidade de Comprimento de Abertura: 2

Um programa útil com estes parâmetros está publicamente disponível como o programa "gap" do Genetics Computer Group, Madison WI. Os parâmetros acima mencionados são os parâmetros por defeito para comparações peptídicas (conjuntamente sem penalizações para aberturas de extremidade).

Os parâmetros para a comparação de polinucleótidos incluem os seguintes:

Algoritmo: Needleman e Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

Matriz de comparação: igualdades = +10, falsa igualdade = 0

Penalidade de Abertura: 50

Penalidade de Comprimento de Abertura: 3

Disponível como: O programa "gap" do Genetics Computer Group, Madison WI. Estes são os parâmetros por defeito para comparações de ácidos nucleicos.

Um significado preferido para a "identidade" para polinucleótidos e polipéptidos, conforme o caso, é proporcionado em (1) e (2) abaixo.

(1) As formas de realização de polinucleótidos incluem ainda um polinucleótido isolado compreendendo uma sequência de polinucleótidos que tem, pelo menos, uma identidade de 50, 60, 70, 80, 85, 90, 95, 97 ou 100% com qualquer das sequências de referência de SEQ ID N°:9 a SEQ ID N°:16, em que a referida sequência de polinucleótidos pode ser idêntica a qualquer das sequências de referência de SEQ ID N°:9 a SEQ ID N°:16 ou pode incluir até um determinado número inteiro de alterações de nucleótidos comparativamente à sequência de referência, em que as referidas alterações são seleccionadas do grupo consistindo de, pelo menos, uma deleção, substituição, incluindo transição e transversão ou inserção de nucleótido, e em que as referidas alterações podem ocorrer nas posições terminais 5' ou 3' da sequência de nucleótidos de referência ou em qualquer local entre essas posições terminais, espalhadas individualmente entre os nucleótidos na sequência de referência ou em um ou mais grupos contíguos dentro da sequência de referência, e em que o referido número de alterações de nucleótidos é determinado por multiplicação do número total de nucleótidos em qualquer de SEQ ID N°:9 a SEQ ID N°:16 pelo número inteiro que define a percentagem de identidade dividido por 100 e depois subtraindo esse produto do referido número total de nucleótidos em qualquer de SEQ ID N°:9 a SEQ ID N°:16 ou:

$$n_n \leq x_n - (x_n \cdot y),$$

em que n_n é o número de alterações de nucleótidos, x_n é o número total de nucleótidos em qualquer de SEQ ID N°:9 a SEQ ID N°:16, y é 0,50 para 50%, 0,60 para 60%, 0,70 para 70%, 0,80 para 80%, 0,85 para 85%, 0,90 para 90%, 0,95 para 95%, 0,97 para 97% ou 1,00 para 100%, e $\cdot$ é o símbolo para o operador da multiplicação, e em que qualquer produto não inteiro de x_n e y é arredondado para baixo, para o número inteiro mais próximo antes de subtrair-lo de x_n . As alterações de sequências de polinucleótidos codificando os polipéptidos de qualquer uma de SEQ ID N°:1 a SEQ ID N°:8 podem criar mutações sem sentido, de falso sentido ou desvios de grelha nesta sequência codificante e alterando, desse modo, o polipéptido codificado pelo polinucleótido que segue essas alterações.

A título de exemplo, uma sequência de polinucleótidos da presente invenção pode ser idêntica a qualquer das sequências de referência de SEQ ID N°:9 a SEQ ID N°:16, isto é, pode ser 100% idêntica ou pode incluir até um determinado número inteiro de alterações de ácidos nucleicos comparativamente à sequência de referência de modo que a percentagem de identidade seja menos de 100% de identidade. Essas alterações são seleccionadas do grupo consistindo de, pelo menos, uma deleção, substituição, incluindo transição e transversão ou inserção de ácido nucleico, e em que as referidas alterações podem ocorrer nas posições terminais 5' ou 3' da sequência de polinucleótidos de referência ou em qualquer local entre essas posições terminais, espalhadas individualmente entre os ácidos nucleicos na sequência de referência ou em um ou mais grupos contíguos dentro da sequência de referência. O número de alterações de ácidos

nucleicos para uma determinada percentagem de identidade é determinado por multiplicação do número total de ácidos nucleicos em qualquer de SEQ ID N°:9 a SEQ ID N°:16 pelo número inteiro que define a percentagem de identidade dividido por 100 e depois subtraindo esse produto do referido número total de ácidos nucleicos em qualquer de SEQ ID N°:9 a SEQ ID N°:16 ou:

$$n_n \leq x_n - (x_n \cdot y),$$

em que n_n é o número de alterações de ácidos nucleicos, x_n é o número total de ácidos nucleicos em qualquer de SEQ ID N°:9 a SEQ ID N°:16, y é, por exemplo, 0,70 para 70%, 0,80 para 80%, 0,85 para 85% etc., $\cdot$ é o símbolo para o operador da multiplicação, e em que qualquer produto não inteiro de x_n e y é arredondado para baixo, para o número inteiro mais próximo antes de subtrair-lo de x_n .

(2) As formas de realização do polipéptido incluem ainda um polipéptido isolado compreendendo um polipéptido que tem, pelo menos, 50, 60, 70, 80, 85, 90, 95, 97 ou 100% de identidade com a sequência de referência de polipéptido de qualquer uma de SEQ ID N°:1 a SEQ ID N°:8, em que a referida sequência polipeptídica pode ser idêntica a qualquer sequência de referência de SEQ ID N°: a SEQ ID N°:8 ou pode incluir até um determinado número inteiro de alterações de aminoácidos comparativamente à sequência de referência, em que as referidas alterações são seleccionadas do grupo consistindo de, pelo menos, uma deleção, substituição, incluindo substituição conservadora e não conservadora ou inserção de aminoácido, e em que as referidas alterações podem ocorrer nas posições terminais amino ou carboxilo da sequência polipeptídica de referência ou em qualquer local entre aquelas posições terminais, espalhadas individualmente

entre os aminoácidos na sequência de referência ou em um ou mais grupos contíguo dentro da sequência de referência, e em que o referido número de alterações de aminoácidos é determinado por multiplicação do número total de aminoácidos em qualquer de SEQ ID N°:1 a SEQ ID N°:8 pelo número inteiro que define a percentagem de identidade dividido por 100 e depois subtraindo esse produto do referido número total de aminoácidos em qualquer de SEQ ID N°:1 a SEQ ID N°:8 ou:

$$n_a \leq x_a - (x_a \cdot y),$$

em que n_a é o número de alterações de aminoácidos, x_a é o número total de aminoácidos em SEQ ID N°:2, y é 0,50 para 50%, 0,60 para 60%, 0,70 para 70%, 0,80 para 80%, 0,85 para 85%, 0,90 para 90%, 0,95 para 95%, 0,97 para 97% ou 1,00 para 100%, e $\cdot$ é o símbolo para o operador da multiplicação, e em que qualquer produto não inteiro de x_a e y é arredondado para baixo, para o número inteiro mais próximo antes de subtrair-lo de x_a .

A título de exemplo, uma sequência polipeptídica da presente invenção pode ser idêntica à sequência de referência de qualquer uma de SEQ ID N°:1 a SEQ ID N°:8, isto é, pode ser 100% idêntica ou pode incluir até um determinado número inteiro de alterações de aminoácidos comparativamente à sequência de referência de modo que a percentagem de identidade seja menos de 100% de identidade. Essas alterações são seleccionadas do grupo consistindo de, pelo menos, uma deleção, substituição, incluindo substituição conservadora e não conservadora ou inserção de aminoácido, e em que as referidas alterações podem ocorrer nas posições terminais amino ou carboxilo da sequência polipeptídica de referência ou em qualquer local entre aquelas posições terminais, espalhadas individualmente entre

os aminoácidos na sequência de referência ou em um ou mais grupos contíguo dentro da sequência de referência. O número de alterações de aminoácidos para uma determinada % de identidade é determinado por multiplicação do número total de aminoácidos em qualquer de SEQ ID N°:1 a SEQ ID N°:8 pelo número inteiro que define a percentagem de identidade dividido por 100 e, depois, subtraindo esse produto do referido número total de aminoácidos em qualquer de SEQ ID N°:1 a SEQ ID N°:8 ou:

$$n_a \leq x_a - (x_a \cdot y),$$

em que n_a é o número de alterações de aminoácidos, x_a é o número total de aminoácidos em qualquer de SEQ ID N°:1 a SEQ ID N°:8, y é, por exemplo, 0,70 para 70%, 0,80 para 80%, 0,85 para 85% etc., e $\cdot$ é o símbolo para o operador da multiplicação, e em que qualquer produto não inteiro de x_a e y é arredondado para baixo, para o número inteiro mais próximo antes de subtrair-lo de x_a .

Legendas das figuras

Figura 1: Informação de sequência para C-LytA. Cada repetição foi definida na base de alinhamento múltiplo de sequência e previsão da estrutura secundária utilizando os seguintes programas de alinhamento: 1) matchBox (Depiereux E et al. (1992) Comput Applic Biosci 8:501-9); 2) ClustalW (Thompson JD et al. (1994) Nucl Acid Res 22:4673-80); 3) Block-Maker (Henikoff S et al. (1995) Gene 163:gc17-26)

Figura 2: CPC e Construções nativas (SEQ ID N° 27-36)

Figura 3: Estrutura esquemática da proteína de fusão CPC-p501 His expressa em *S. cerevisiae*

Figura 4: Estrutura primária da proteína de fusão CPC-p501 His (SEQ ID N° 41)

Figura 5: Sequência de nucleótidos de CPC-P501 His(pRIT15201) (SEQ ID N° 42)

Figura 6: Estratégia de clonagem para a produção do plasmídeo pRIT 15201

Figura 7: Mapa do plasmídeo pRIT15201

Figura 8: Expressão comparativa de CPC P501 e P501 na estirpe DC5 de *S. cerevisiae*

Figura 9: Produção de CPC-p501s HIS (Y1796) em pequena escala. A Fig. 9A representa a produtividade de antígeno como estimada por SDS-PAGE com coloração de prata; A Fig. 9B representa a produtividade de antígeno como estimada por transferência de Western.

Figura 10: Esquema de purificação de CPC-P501-His produzido por Y1796.

Figura 11: Perfil da proteína purificada de CPC P501 His (géis pré-fabricados de 4-12% de poliacrilamida Novex Nu-Page).

Figura 12: Sequência de P501 S de tamanho total nativa (SEQ ID N°:17)

Figura 13: Sequência da cassete de expressão de CPC-p501 S de JNW735 (SEQ ID N°:18)

Figura 14: Duas Sequências P501S de codão otimizado (SEQ ID N°:19-20)

Figura 15: Sequência 19 de codão otimizado re-manipulada (SEQ ID N°:21)

Figura 16: Sequência 20 de codão otimizado re-manipulada (SEQ ID N°:22)

Figura 17: A sequência inicial para a otimização de CPC (SEQ ID N°:23)

Figura 18: Sequências de CPC de codão otimizado representativas (SEQ ID N°:24-25)

Figura 19: Sequência de CPC de codão otimizado manipulada (SEQ ID N°:26)

Figura 20: Sequências e construções de candidato de fusão de P501S CPC (SEQ ID N° 37-40 e 45-48)

Figura 21: Análise de transferência de Western de células CHO após transfecção transitória com P501 S (JNW680), CPC-p501s (JNW735) e controlo com vector vazio.

Figura 22: Respostas de anticorpo anti-P501S após imunização no dia 0, 21 e 42 com pVAC-P501S (JNW680, murganhos B1-9) ou vector vazio (pVAC, murganhos A1-6). For recolhida uma pré-amostra de sangue no dia -1. Subsequentemente as amostras de sangue foram recolhidas no dia 28 e dia 49 (murganhos A1-3, B1-3) e dia 56 (murganhos A4-6, B4-9). Todos os soros foram

testados a uma diluição de 1/100. Os resultados para os murganhos imunizados com pVAC foram utilizados para calcular a média. São mostrados os resultados para os murganhos individuais imunizados com pVAC-P501 S. Como controlo positivo são incluídos soros de murganhos imunizados com Adeno-P501 S (Corixa Corp, diluídos a 1/100).

Figura 23: Rastreio da biblioteca peptídica utilizando murganhos C57BL/6 imunizados no dia 0, 21, 42 e 70 com pVAC-P501 S (JNW680). Todos os péptidos foram utilizados a uma concentração final de 50 µg/mL. Os péptidos 1-50 estão sobrepostos em 15-20 monómeros, obtidos da Corixa. Os péptidos 51-70 são previstos com 8-9 monómeros de epitopos Kb e DB e foram encomendados à Mimotopes (Reino Unido). As amostras 71-72 e 73-78 são controlos de DMSO e controlos sem péptido, respectivamente. O Gráfico A mostra as respostas IFN-γ enquanto que o Gráfico B mostra as respostas IL-2. Os péptidos seleccionados para utilização em ensaios imunológicos subsequentes são mostrados a preto.

Figura 24: Respostas celulares por ELISPOT no dia 77 após imunização PMID no dia 0, 21, 42, e 70 com pVAC-P501S (JNW680, B6-9) e pVAC vazio (A4-6). Os péptidos 18, 22 e 48 foram utilizados a 50 µg/mL. A proteína CPC-p501 S foi utilizada a 20 µg/mL. O Gráfico A mostra as respostas IFN-γ enquanto que o Gráfico B mostra as respostas IL-2.

Figura 25: Comparação de P501 S e CPC-p501 S. As respostas celulares foram medidas por IL-2 ELISPOT utilizando o péptido 22 (10 µg/mL) no dia 28. Os murganhos foram imunizados por PMID no dia 0 e 21 com pVAC vazio (controlo), pVAC-P501 S (JNW680) e CPC-p501 S (JNW735).

Figura 26: Resposta imunológica (linfoproliferação em células de baço) após imunização de proteína com CPC-p501 S.

Figura 27: Avaliação da resposta imunológica a construções CPC-p501 S diferentes. As respostas celulares foram medidas por IL-2 ELISPOT no dia 28. Os murganhos foram imunizados por PMID no dia 0 e 21 com p7313-ie vazio (controle), construções JNW735 e CPC-p501s (JNW770, 771 e 773)

Figura 28: Sequências de MUC-1 CPC (SEQ ID N° 49 e 50)

Figura 29: Sequências de ss-CPC-MUC-1 (SEQ ID N° 51 e 52)

A invenção vai ser adicionalmente descrita com referência aos seguintes exemplos:

EXEMPLO I: Preparação da estirpe de Levedura recombinante Y1796 expressando a Proteína de Fusão P501 contendo um C-LytA-P2-C-LytA (CPC) como parceiro de fusão

1. - Preparação da proteína

A estrutura da proteína de fusão C-P2-C-p501 (denominada alternativamente CPC-p501) para ser expressa em *S. cerevisiae* é representada na figura 3. Esta fusão contém a região do C-terminal do gene *LytA* (resíduos 187 a 306), na qual foi introduzido o fragmento P2 da toxina do tétano (resíduos 830-843). O fragmento P2 é colocado entre os resíduos 277 e 278 de C-Lyt-A. O fragmento de C-lytA contendo a inserção P2 é seguida por P501 (resíduos de aminoácidos 51 a 553) e pela extremidade His.

A estrutura primária da proteína de fusão resultante tem a sequência descrita na figura 4 e a sequência codificante correspondente à concepção de proteína acima está na figura 5.

2. - Estratégia de clonagem para a produção de um plasmídeo de levedura expressando a proteína de fusão CPC-p501 (51-553)-His

- O material inicial é o vector de levedura pRIT15068 (pedido de Patente UK 0015619.0).
- Este vector contém o promotor de levedura *Cup1*, a sequência codificante do sinal alfa prépro de levedura e a sequência codificante correspondente aos resíduos 55 a 553 de P501S seguida pela extremidade His.
- A estratégia de clonagem delineada na figura 6 inclui os seguintes passos:

a) O primeiro passo é a inserção da sequência P2 (de código otimizado para a expressão em levedura) em grelha, dentro da sequência codificante de C-lytA. A sequência codificante de C-lytA é contida pelo plasmídeo pRIT 14662 (documento PCT/EP99/00660). A inserção é realizada utilizando um adaptador formado por dois oligonucleótidos complementares denominados P21 e P22 para dentro do plasmídeo pRIT 14662 previamente aberto por NcoI.

A sequência de P21 e P22 é:

P21 5' catgcaatacatcaaggctaactctaagttcattggtatcactga
aggcgt 3'

P22 3' gttatgtagttccgattgagattcaagtaaccatagtgacttccg
cagtac 5'

Após ligação e transformação de *E. coli* e caracterização de transformante, é obtido o plasmídeo denominado pRIT15199.

b) O segundo passo é a preparação do fragmento de ADN C-lytA-P2-C-lytA por amplificação de PCR. A amplificação é realizada utilizando pRIT15199 como molde e os oligonucleótidos denominados C-LytANOTATG e C-LytA-aa55. Sendo a sequência de ambos os oligonucleótidos:

C-LytANOTATG

= 5' aaaaccatggcgccgcttacgtacattccgacggctcttatccaaaa
gacaag 3'

C-LytA-aa55

= 5' aaacatgtacatgaacttttctggcctgtctgccagtgttc 3'

O fragmento amplificado é tratado com as enzimas de restrição NcoI e AflIII para produzir as extremidades coesivas respectivas.

c) O passo seguinte é a ligação do fragmento acima com o vector pRIT15068 (o maior fragmento obtido após tratamento com NcoI) para produzir a sequência codificante completa da proteína de fusão. Após ligação e transformação de *E. coli*, é obtido o plasmídeo denominado pRIT15200. Neste plasmídeo, o sítio único NcoI restante contém o codificador ATG para o codão de iniciação.

d) No passo seguinte, o fragmento NcoI contendo o promotor CUP1 e uma porção de sequências plasmídicas 2 μ é preparado a partir do plasmídeo PRIT 15202. O plasmídeo pRIT 15202 é um derivado 2 μ de levedura contendo o promotor CUP1 com um sítio NcoI em ATG (sequência de ATG: AAACC ATG)

e) O fragmento NcoI isolado a partir de pRIT 15202 é ligado a pRIT15200, previamente aberto com NcoI, na orientação certa, de tal modo que o promotor pCUP1 está no lado 5' da sequência codificante. Isto resulta na produção de um plasmídeo de expressão final denominado pRIT15201 (ver figura 7).

3. - Preparação da estirpe Y1796 de levedura recombinante (RIX4440)

O plasmídeo pRIT 15201 é utilizado para transformar a estirpe DC5 de *S. cerevisiae* (ATCC 20820). Após selecção e caracterização dos transformantes de levedura contendo o plasmídeo pRIT 15201, é obtida uma estirpe de levedura recombinante denominada Y1796 expressando a proteína de fusão CPC-P501-His. A proteína, após redução e carboxiamidação, é isolada e purificada através de cromatografia de afinidade (IMAC) seguida por cromatografia de permuta aniónica (Q Sepharose FF).

Exemplo II

De um modo análogo, as construções de proteína como descritas na figura 2 podem ser expressas utilizando as sequências de ADN correspondentes aí mostradas. Em particular, a estirpe SC333 de levedura (construção 2) corresponde à estirpe Y1796 mas expressando P501₅₅₋₅₅₃ desprovido do parceiro de fusão CPC. A estirpe Y1800 de levedura (construção 3) corresponde à estirpe Y1796 mas compreende adicionalmente o sinal de sequência nativo para P501S (aa1-aa34), enquanto a estirpe Y1802 de levedura (construção 4) compreende a sequência de pré-sinal alfa a montante da sequência CPC-p501S. A estirpe Y1790 de levedura (construção 5) está a expressar uma construção de P501S desprovida de CPC e tendo a sequência do sinal pré-pro alfa.

Exemplo III. Preparação de CPC-P501 purificado

1. - Produção de CPC-P501S HIS (Y1796) em pequena escala

Para Y1796, em meio mínimo suplementado com histidina, a expressão é induzida na fase logarítmica por adição de CuSO_4 que varia de 100 a 500 μM , e a cultura é mantida a 30° . As células são recolhidas após 8 ou 24 h de indução. O cobre é adicionado imediatamente antes da utilização e previamente não misturado com o meio.

Para a análise de SDS-PAGE, a extracção das células de levedura é realizada em tampão fosfato de citrato pH 4,0 + 130 mM de NaCl. A extracção é realizada com esferas de vidro para uma pequena quantidade de células e com a prensa Francesa para uma quantidade mais elevada de células, e depois misturada com tampão de amostra e analisada por SDS-PAGE. Os resultados de análise comparativa em SDS-PAGE das diferentes construções são representados na figura 8 e resumidos na Tabela 2 abaixo. Como mostrado na Tabela 1 abaixo, o nível de expressão da cultura é muito maior para a estirpe Y1796 comparativamente ao nível de expressão da estirpe SC333 progenitora, uma estirpe que expressa o P501S-His correspondente desprovido do parceiro CPC. Do mesmo modo, a presença de uma sequência de sinal (alfa pré) não afecta os resultados discutidos acima: o nível de expressão da cultura é muito maior para a estirpe Y1802 comparativamente ao nível de expressão da correspondente estirpe Y1790, uma estirpe que expressa o P501S-His correspondente desprovido do parceiro CPC.

Tabela 2

Estirpe Recombi- nante	Plasmídeo	Promotor	Sequência de sinal	Parceiro de Fusão	Sequências P501 aa	Nível de expressão
SC333	Ma333	CUP1	-	-	55-553-His	- ND
Y1796	pRIT 15201	CUP 1	-	CPC	51-553-His	+++
Y1802	pRIT 15219	CUP 1	α pré	CPC	51-553-His	++++
Y1790	pRIT 15068	CUP 1	α prépro	-	55-553-His	+
CPC = clyta P2 clyta Nd = não detectável, mesmo em transferência de Western + = detectável em transferência de Western +++/++++ = detectável em transferência de Western e visível em géis corados com prata						

2. - Fermentação de Y1796 (RIX4440) em grande escala

São espalhados 100 μ L do inóculo de trabalho em meio sólido e cultivados durante aproximadamente 24 h a 30 °C. Esta pré-cultura sólida é, depois, utilizada para inocular uma pré-cultura líquida em balões de agitação.

Esta pré-cultura líquida é cultivada durante 20 h, a 30 °C e transferida para um fermentador de 20 L. A fermentação em *batch* inclui uma fase de crescimento de cerca de 44 h e uma fase de indução de cerca de 22 h.

A fonte de carbono (glucose) foi suplementada à cultura por alimentação contínua. A concentração residual de glucose foi mantida muito baixo (≤ 50 mg/L) de modo a minimizar a produção de etanol por fermentação. Isto foi realizado limitando o desenvolvimento do microrganismo através de taxa de alimentação de glucose limitada. No final da fase de crescimento, o promotor

CUP1 é induzido por adição de CuSO_4 de modo a produzir o antigénio.

A ausência de contaminações foi verificada por inoculação de 10^6 células em frasquinhos convencionais de TSB e THI suplementados com nistatina e incubados durante 14 dias a 20-25 °C e 30-35 °C, respectivamente. Não foi observado crescimento como esperado.

3. - Caracterização de antigénio e produtividade

Os homogenatos celulares foram preparados utilizando a prensa Francesa em amostras de fermentação recolhidas em tempos diferentes durante a fase de indução e analisadas por SDS-PAGE e Transferência de Western. Foi mostrado que a maior parte da proteína de interesse estava localizada na fracção insolúvel obtida do homogenato celular após centrifugação. As análises de SDS-PAGE e de Transferência de Western mostradas nas Figuras abaixo foram realizadas nos sedimentos obtidos após centrifugação destes homogenatos celulares.

As figuras 8A e B mostram uma cinética de produção de antigénio durante a fase de indução para a cultura PR0127. Parece que não ocorreu expressão do antigénio durante a fase de crescimento. A produtividade específica do antigénio parece aumentar a partir do início da fase de indução até 6 h e depois permaneceu completamente estável até ao final. Mas a produtividade volumétrica aumentou por um factor de 1,5 a 2 devido à acumulação de biomassa observada durante o mesmo período de tempo. A produtividade de antigénio foi estimada em cerca de 500 mg por litro de meio de fermentação por comparação da referência purificada de antigénio e extractos brutos em SDS-PAGE com coloração de prata (figura 9A) e análises de WB

utilizando um anticorpo anti-P501 S (um ascítico de murino dirigido contra P501S aa439-aa459 utilizado numa diluição de 1/1000) (figura 9B).

Exemplo IV. Purificação da proteína de fusão CPC-P501 (51-553)-His produzida por Y1796

Após ruptura das células, a proteína é associada com a fracção do sedimento. Foi introduzida uma carbamido-metilação da molécula no processo, de modo a lidar com a agregação oxidativa da molécula com ele própria ou com contaminantes proteicos da célula hospedeira através de ligação covalente com pontes dissulfureto. A utilização de detergentes foi também requerida para controlar o carácter hidrofóbico desta proteína (previstos 12 domínios transmembranares).

O protocolo de purificação, desenvolvido para a escala de 1 L de cultura com OD (densidade óptica) de 120, é descrito na figura 10. Todas as operações são realizadas à temperatura ambiente (t.a.).

De acordo com o ensaio de proteína DOC TCA BCA, o rendimento global da purificação é 30 - 70 mg de antigénio purificado/L de cultura a OD 120. O rendimento é ligado ao nível de expressão da cultura e é maior comparativamente ao rendimento de purificação da estirpe progenitora expressando P501 S-His não fundida. O ensaio de proteína é realizado como se segue: as proteínas são primeiro precipitadas utilizando TCA (ácido tricloroacético) na presença de DOC (desoxicolato) depois dissolvidas num meio alcalino na presença de SDS. As proteínas reagem depois com BCA (ácido bicinconínico) (Pierce) para formar um complexo roxo solúvel que apresenta uma absorvência elevada a 562 nm, que é proporcional à quantidade de proteínas presente na amostra. A

análise de SDS-PAGE de 3 quantidades purificadas (figura 11) não mostra diferença entre condições redutoras e não redutoras (ver pistas 2, 3 e 4 versus pistas 5, 6 e 7). O perfil consiste de uma principal banda de 70 kDa, um arrastamento de elevado peso molecular e ligeiras bandas de degradação. Todas as bandas são detectadas por um anticorpo monoclonal anti P501 S específico.

Exemplo V. Preparação de vacina utilizando a proteína CPC-P501S His

A proteína do Exemplo 3 ou 4 pode ser formulada numa vacina contendo QS21 e 3D-MPL numa emulsão de óleo em água.

1. - Preparação da Vacina:

O antigénio é produzido como mostrado no Exemplo 1 a 3, um C-LytA-P2-P501 S His. Como um adjuvante, a formulação compreende uma mistura de monofosforil-lípido A 3-O-acilado (3D-MPL) e QS21 numa emulsão de óleo/água. O sistema adjuvante SBAS2 foi previamente descrito no documento WO 95/17210.

3D-MPL: é um imunoestimulante derivado do lipopolissacárido (LPS) da bactéria Gram-negativa *Salmonella minnesota*. O MPL foi desacilado e carece de um grupo fosfato na unidade de lípido A. Este tratamento químico reduz dramaticamente a toxicidade enquanto preserva as propriedades imunoestimulantes (Ribi, 1986). A Ribi Immunochemistry produz e fornece MPL à SB-Biologicals. Experiências realizadas na Smith Kline Beecham Biologicals mostraram que 3D-MPL combinado com vários veículos intensifica fortemente a imunidade celular humoral e de tipo TH1.

QS21: é uma molécula natural de saponina extraída da casca da árvore Sul Americana, Quillaja saponaria Molina. Uma técnica de purificação desenvolvida para separar as saponinas individuais dos extractos brutos da casca, permitiram o isolamento da saponina particular, QS21, que é um glicósido de triterpeno demonstrando uma forte actividade adjuvante e uma toxicidade mais baixa comparativamente ao componente progenitor. Foi demonstrado que QS21 activa os CTL restritos da classe I de MHC até diversas subunidades de Ags, assim como estimula a proliferação linfocítica específica de AG (Kensil, 1992). A Aquila (anteriormente Cambridge Biotech Corporation) produz e fornece QS21 à SB-Biologicals. As experiências realizadas na SmithKline Beecham Biologicals demonstraram um claro efeito sinérgico de combinações de MPL e QS21 na indução de respostas imunológicas celulares humorais e de tipo TH1.

A emulsão de óleo/água é composta de uma fase orgânica feita de 2 óleos (um tocoferol e esqualeno), e uma fase aquosa de PBS contendo Tween 80 como emulsionante. A emulsão é compreendida de 5% de esqualeno, 5% de tocoferol e 0,4% de Tween 80 e tem um tamanho médio de partícula de 180 nm e é conhecida como SB62 (ver documento WO 95/17210). As experiências realizadas na SmithKline Beecham Biologicals provaram que a adição desta emulsão de óleo/água a 3D-MPL/QS21 (SBAS2) aumenta adicionalmente as propriedades imunoestimulantes dos últimos contra vários antígenos de subunidade.

2. - Preparação da emulsão SB62 (concentrada 2 vezes):

O tween 80 é dissolvido em soro fisiológico tamponado com fosfato (PBS) para produzir uma solução de 2% em PBS. Para proporcionar 100 mL de emulsão concentrada duas vezes, são misturados exaustivamente por vórtice 5 g de alfa DL tocoferol e 5 mL de esqualeno. São adicionados 90 mL de solução de PBS/Tween e misturados exaustivamente. A emulsão resultante é depois passada através de uma seringa e finalmente microfluidizada utilizando um aparelho M110S microfluidics. As gotas de óleo resultantes têm um tamanho de aproximadamente 180 nm.

3. - Formulações:

Uma formulação típica contendo 3D-MPL e QS21 numa emulsão de óleo/água é realizada como se segue: 20 µg - 25 µg de C-LytA P2-p501S são diluídos em PBS pH 6,8 10 vezes concentrado e H₂O antes da adição consecutiva de SB62 (50 µL), MPL (20 µg), QS21 (20 µg), opcionalmente compreendendo oligonucleótido CpG (100 µg) e 1 µg/mL de tiomersal como conservante. A quantidade de cada componente pode variar como necessário. Todas as incubações são realizadas à temperatura ambiente com agitação.

Exemplo VI. Sequências P501S de codão otimizado

1. - Preparação dos plasmídeos recombinantes de controle:

A sequência de tamanho total de P501S foi clonada em pVAC (Thomsen, Immunology, 1998; 95:510P105), produzindo o plasmídeo de expressão JNW680. A SEQ ID N°:17 representa a cassette de expressão humana de P501S no plasmídeo JNW680 e é ilustrada na

Figura 12. A sequência de proteína de SEQ ID N°:17 é mostrado em formato de letra única, sendo os codões de iniciação e terminação mostrados a cheio. A sequência de Kozak é denotada pelos símbolos cardinal. O índice de utilização de codão da sequência humana de P501 S (SEQ ID N°:17) é 0,618, como calculado pelo programa SynGene.

Programa SynGene

Basicamente, os codões são atribuídos utilizando um método estatístico para produzir um gene sintético possuindo uma frequência de codão mais próxima daquela encontrada naturalmente em genes altamente expressos de *E.coli* e humanos.

O SynGene é uma versão actualizada do programa de Visual Basic denominado Calcgene, escrito por R. S. Hale e G Thompson (Protein Expression and Purification Vol. 12 pp.185-188 (1998)). Para cada resíduo de aminoácido na sequência original, foi atribuído um codão baseado na probabilidade dele aparecer em genes altamente expressos de *E. coli*. Os detalhes do programa Calcgene, que opera sob Microsoft Windows 3.1, podem ser obtidos a partir dos autores. Porque o programa aplica um método estatístico para atribuir codões ao gene sintético, nem todos os codões resultantes são os mais frequentemente utilizados no organismo alvo. Mais exactamente, a proporção de codões frequentemente e infrequentemente utilizados do organismo alvo é reflectida na sequência sintética através da atribuição de codões nas proporções correctas. Contudo, como não existe uma regra firme para atribuir um codão particular a uma posição particular na sequência, cada vez que é o programa é executado irá produzir um gene sintético diferente - embora cada um tenha o mesmo perfil de utilização de codão e cada um codifique a mesma sequência de aminoácidos. Se o programa for executado

diversas vezes para uma dada sequência de aminoácidos e um dado organismo alvo, serão produzidas diversas sequências de nucleótidos diferentes que podem diferir no número, tipo e posição dos sítios de restrição, sinais de *splice* de intrão, etc., algumas das quais pode ser indesejáveis. O especialista será capaz de seleccionar uma sequência apropriada para utilização na expressão do polipéptido na base destas características.

Além disso, uma vez que os codões são aleatoriamente atribuídos numa base estatística, é possível (embora talvez improvável) que dois ou mais codões que são relativamente raras vezes utilizados no organismo alvo podem ser aglomerados em íntima proximidade. Acredita-se que esses aglomerados possam perturbar a maquinaria de tradução e resultar em taxas de expressão particularmente baixas, assim o algoritmo para a escolha dos codões no gene optimizado exclui quaisquer codões com um valor de RSCU menor que 0,2 para genes altamente expressos, de modo a prevenir quaisquer aglomerados de codões raros que sejam fortuitamente seleccionados. A distribuição dos codões restantes é depois alocada de acordo com as frequências para *E. coli* altamente expressos, para produzir uma distribuição global dentro do gene sintético que é típica desses genes (coeficiente = 0,85) e também para genes humanos altamente expressos (coeficiente = 0,50). Syngene (Peter Ertl, não publicado), uma versão actualizada do programa Calcgene, permite que a exclusão de codões raros seja opcional, e é também utilizado para alocar codões de acordo com o perfil de frequência de codão de genes humanos altamente expressos.

A sequência da cassette CPC-p501 S clonada a partir do vector pRIT15201 (ver Figura 7) em pVAC, produzindo assim o plasmídeo JNW735, é apresentada em SEQ ID N°:18 e é ilustrada na Figura 13. Esta sequência é idêntica à sequência pRIT15201 à

excepção da remoção do marcador His e a adição de uma sequência de Kozak (GCCACC) e sítios de enzimas de restrição apropriados. A sequência de aminoácidos de SEQ ID N°:18 é mostrada no formato de letra única, sendo os códons de iniciação e terminação mostrados a cheio. Os resíduos em caixas de texto são o epitopo auxiliar P2 do toxóide do tétano. Os resíduos sublinhados são o marcador de purificação Clyta. A sequência de Kozak é denotada através dos símbolos de cardinal.

2. - Produção dos plasmídeos recombinantes com sequências P501S de codão otimizado:

Embora o índice do coeficiente de codão (CI) da sequência nativa de P501S seja já alta (0,618), é possível aumentar o valor de CI ainda mais. Isto terá dois benefícios potenciais - melhorar a expressão e/ou imunogenicidade do antigénio e reduzir a possibilidade de recombinação entre o vector de P501 S e sequências genómicas.

Utilizando o programa Syngene, foi obtida uma selecção (SEQ ID N°:19 a SEQ ID N°:20) de sequências de codão otimizado (Figura 14). A Tabela 3 abaixo mostra uma comparação do índice do coeficiente de codão para a sequência inicial de P501 S e duas sequências de codão otimizado representativas, seleccionadas na base de um perfil adequado de sítio de enzima de restrição e de um bom índice de CI.

Tabela 3 - Comparação dos índices do coeficiente de codão de dois genes P501 S de codão otimizado

Sequência	Índice do coeficiente de codão (CI)
-----------	-------------------------------------

P501 S	0,618
SEQ ID N°:19	0,725
SEQ ID N°:20	0,755

3. Avaliação adicional das sequências de codão otimizado

Sequência SEQ ID N°:19

Embora a SEQ ID N°: 19 tenha um bom índice de CI (0,725), esta contém um duplete de codões raros na posição 202 e 203 de aminoácidos. Estes codões foram substituídos manualmente com codões mais frequentes através de alteração da sequência de ADN de TTGTTG para CTGCTG. Para facilitar a clonagem e expressão, foram adicionados sítios de enzima de restrição e uma sequência de Kozak. A sequência manipulada final (SEQ ID N°:21) é mostrada na Figura 15. O programa Syngene foi utilizado para fragmentar esta sequência em oligonucleótidos com uma sobreposição mínima de 19-20 bases. Deste modo, a Figura 15 mostra a SEQ ID N° 19 de P501S de codão otimizado remanipulada. Os sítios de enzima de restrição estão sublinhados, a sequência de Kozak está a cheio, a sequência de ADN remanipulada para remover um duplete de codões raros está numa caixa de texto.

Utilizando um protocolo de PCR de dois passos, os iniciadores com sobreposição, produzidos pelo programa Syngene foram primeiro montados utilizando um protocolo de Montagem por PCR (detalhado abaixo). A reacção de montagem produz uma população variada de fragmentos. O fragmento de tamanho total correcto foi recuperado/amplificado utilizando o protocolo de recuperação por PCR e os iniciadores terminais. O fragmento de PCR resultante foi excisado de um gel de agarose, purificado, digerido com NheI e XhoI e clonado em pVAC. Os clones positivos

foram identificados através de análise de enzima de restrição e confirmados através de sequenciação de cadeia dupla. Isto produz o plasmídeo JNW766 que, devido à natureza propensa a erros do processo de PCR, continha uma única mutação silenciosa (C para T na posição 360 da SEQ ID N°: 21).

1. Reacção de montagem - condições de PCR, protocolo genérico

Mistura de reacção (volume total = 50 µL):

- Tampão de reacção 1x (Pfx ou Proofstart)
- 1 µL de agrupamento de Oligos (mistura igual de todos os oligos de sobreposição)
- 0,5 mM de dNTPs
- ADN polimerase (Pfx ou Proofstart, 2,5-5 U)
- +/- 1 mM de MgSO₄
- +/- Solução intensificadora 1x (intensificador de Pfx ou tampão Q de Proofstart)

1. 94 °C durante 120 s (apenas Proofstart)
2. 94 °C durante 30 s
3. 40 °C durante 120 s
4. 72 °C durante 10 s
5. 94 °C durante 15 s
6. 40 °C durante 30 s
7. 72 °C durante 20 s + 3 s/ciclo
8. Ciclo até passo 5, 25 vezes
9. Manter a 4 °C

2. Reacção de recuperação - condições de PCR (protocolo genérico)

Mistura de reacção (volume total = 50 µL):

- Tampão de reacção 1x (Pfx ou Proofstart)
- 5-10 µL de mistura de reacção de montagem
- 0,3-0,75 mM de dNTPs
- 50 pmol de iniciador (iniciador do terminal 5', orientação de sentido)
- 50 pmol de iniciador (iniciador do terminal 3', orientação anti-sentido)
- ADN polimerase (Pfx ou Proofstart, 2,5-5 U)
- +/- 1 mM de MgSO₄
- +/- Solução intensificadora 1x (intensificador de Pfx ou tampão Q de Proofstart)

1. 94 °C 120 s (apenas Proofstart)

2. 94 °C 45 s

3. 60 °C 30 s

4. 72 °C 120 s

5. Ciclo até passo 2, 25 vezes

6. 72 °C 240 s

7. Manter a 4 °C

Sequência SEQ ID N°:20

Embora a SEQ ID N°: 20 tenha um muito bom índice de CI (0,755), verificou-se que continha um duplete de codões raros na posição 131 e 132 de aminoácidos. Estes codões foram substituídos manualmente com codões mais frequentes por alteração da sequência de ADN de TTGTTG para CTGCTG. Para facilitar a clonagem, foi removido um sítio BamHI interno através de mutação de G para C (ver o nucleótido duplamente sublinhado na Figura 16). Para facilitar a clonagem e expressão,

foram adicionados sítios de enzima de restrição e uma sequência de Kozak. A sequência manipulada final (SEQ ID N°:22) é mostrada na Figura 16. O programa Syngene foi utilizado para fragmentar esta sequência em oligonucleótidos com uma sobreposição mínima de 19-20 bases. A Figura 16 mostra assim a sequência 20 de P501S de codão otimizado remanipulada (SEQ ID N°:22). Os sítios de enzima de restrição estão sublinhados, a sequência de Kozak está a cheio, a sequência de ADN remanipulada para remover um duplete de códons raros está numa caixa de texto e uma mutação pontual silenciosa para remover um sítio BamHI está duplamente sublinhada.

Utilizando um protocolo de PCR de dois passos semelhante ao descrito acima, foi amplificado um fragmento de tamanho total de P501S e clonado em pVAC. Os clones positivos foram identificados através de análise de enzima de restrição e confirmados através de sequenciação de cadeia dupla. Isto produziu o plasmídeo JNW764. A sequência da cassette codificante de P501S é mostrada na Figura 16 (SEQ ID N°: 22).

Semelhança da sequência de ADN

As distâncias do par após alinhamento pelo método ClustalV (ponderado) são mostradas na Tabela 3 abaixo. A Tabela 4 abaixo mostra percentagem de semelhança entre a sequência humana de P501S inicial e as duas sequências de codão otimizado SEQ ID N°:21 e 22 seleccionadas para investigação adicional. Os dados confirmam que as sequências de ADN de codão otimizado são aproximadamente 80% semelhantes à sequência original de P501 S.

Tabela 4

SEQ ID N°:	% de semelhança com a sequência inicial de P501S
21	79,6
22	79,4

Exemplo VII. Sequências de CPC de codão otimizado

1. - Abordagem

Uma vez que a sequência original de CPC foi concebida originalmente para expressão óptima em levedura, esta secção descreve o processo de optimização de codão para expressão humana.

2. - Preparação de sequência

A sequência inicial para a optimização de CPC é mostrada na Figura 17 (SEQ ID N°: 23). Isto é inteiramente derivado do pRIT15201 e contém a sequência codificante inteira de CPC mais quatro aminoácidos de P501S para facilitar a clonagem a jusante. Utilizando o programa Syngene, foi obtida uma selecção de sequências de codão optimizado, das quais são mostradas sequências representativas na Figura 18 (SEQ ID N°: 24-25). A Tabela 5 abaixo mostra uma comparação do índice do coeficiente de codão para a sequência de CPC inicial e as duas sequências de codão optimizado representativas.

Tabela 5. Índices do coeficiente de codão para duas sequências optimizados de CPC

Sequência	Índice do coeficiente de codão (CI)
CPC Original = SEQ ID N°:23	0,506
SEQ ID N°:24	0,809
SEQ ID N°:25	0,800

Além da optimização de codão, todas as sequências foram também rastreadas para sítios de clonagem de enzima de restrição. Com base no valor mais elevado de CI e num perfil favorável de sítio da enzima de restrição, a SEQ ID N°: 24 foi seleccionada para construção. Para facilitar a clonagem e expressão, foram adicionados sítios de clonagem 5' e 3' e foi introduzida uma sequência de Kozak (GCCACC) a 5' do codão de iniciação ATG. Esta sequência manipulada é mostrada na Figura 19 (SEQ ID N°:26). Esta sequência inclui quatro aminoácidos de P501 S (numa caixa de texto), sítios de clonagem de enzima de restrição (NheI e XhoI, sublinhados), uma sequência de Kozak (a cheio), um codão de terminação (em itálico) e 4 pb de ADN irrelevante flanqueador para facilitar a clonagem.

O programa Syngene foi utilizado para fragmentar esta sequência em oligonucleótidos de 50-60 monómeros com uma sobreposição mínima de 18-20 bases. Utilizando um protocolo de PCR de dois passos semelhante ao acima descrito, o fragmento correcto foi recuperado/amplificado e clonado em pVAC. Os clones positivos foram identificados através de análise de enzima de restrição e a sequência verificada produzindo o vector JNW759.

4. - Semelhança de ADN

As distâncias do par após alinhamento pelo método ClustalV (ponderado) são mostradas na Tabela 6 abaixo. A Tabela mostra percentagem de semelhança ao nível de ADN entre a sequência inicial de CPC e a sequência de codão otimizado e confirma que as sequências de codão otimizado são aproximadamente 80% semelhantes à sequência original de CPC.

Tabela 6

Sequência SEQ ID N°:	% de semelhança com a sequência inicial de CPC
24	80,2
25	81,6

Exemplo VIII. Construção do candidato de fusão de P501S

Todos os candidatos mostrados no diagrama esquemático abaixo são de codão otimizado e construídos utilizando metodologias de PCR de sobreposição a partir dos plasmídeos JNW764 e JNW759 como moldes (SEQ ID N°: 22 e SEQ ID N°: 26, respectivamente), e clonados no vector de expressão p7313 ie.

Os quatro candidatos mostrados esquematicamente abaixo são baseados em CPC-p501S. O CPC-p501S de codão otimizado é a construção A. Os candidatos B, C, D também incluem a sequência codificando os 50 aminoácidos do N-terminal de P501S, posicionada no N-terminal de CPC-P501S (construção D), C-terminal de CPC-P501S (construção C) ou entre CPC e P501S (construção B). Uma representação esquemática das construções é apresentada na Figura 20. A sequência de nucleótidos e de

proteína para cada uma das quatro construções é mostrada em SEQ ID N°: 37-40 para as sequências de nucleótidos, e SEQ ID N° 45-48 para as sequências polipeptídicas correspondentes. Nas construções A, C e D, o codão sublinhado codifica, de um modo preferido, tirosina (quer TAC ou TAT) mas a sequência de nucleótidos pode ser alterada para codificar treonina (quer ACA, ACC, ACG ou ACT). Na construção B, o codão sublinhado codifica, de um modo preferido, treonina (quer ACA, ACC, ACG ou ACT), mas a sequência de nucleótidos pode ser alterada para codificar tirosina (quer TAC ou TAT). Em todas as construções, a sequência codificante é flanqueada por sítios de clonagem de enzima de restrição apropriados (nestes caso, NotI e BamHI), e uma sequência de Kozak imediatamente a montante do ATG de iniciação. A Tabela 7 abaixo mostra a identificação do plasmídeo para as construções detalhadas acima:

Tabela 7

Construção	Aminoácido no codão sublinhado	Sequência de codão	ID do Plasmídeo
A	Tirosina	TAC	JNW771
B	Treonina	ACA	JNW773
B	Tirosina	TAC	JNW770
C	Tirosina	TAC	JNW777
D	Tirosina	TAC	JNW769

As respostas celulares depois da imunização com p7313-ie (vector vazio), pVAC-P501 S (JNW735), JNW770, JNW771 e JNW773 foram avaliadas através de ELISPOT após uma imunização primária por PMID no dia 0 e três reforços no dia 21, 42 e 70. Os ensaios foram realizados 7 dias após o reforço. A Figura 27 mostra que

foram detectadas boas respostas de IL-2 ELISPOT em murganhos imunizados com JNW770, JNW771 e JNW773.

Exemplo IX. Experiências de imunogenicidade utilizando estudos de distribuição intradérmica mediada por partícula (PMID)

O P501 S de tamanho total, quando distribuído através de distribuição intradérmica mediada por partícula (PMID), produz boas respostas de anticorpo e celulares. Estes dados demonstram que a PMID é uma via muito eficaz de distribuição. Além disso, a comparação de P501 S e CPC-p501 S confirma que CPC-P501S induz uma resposta imunológica mais forte, como determinado por ELISPOT de péptido.

1. - Materiais e Métodos

1.1. Imunização cutânea por pistola de genes

O ADN plasmídico foi precipitado sobre esferas de ouro de 2 µm de diâmetro utilizando cloreto de cálcio e espermidina. As esferas carregadas foram revestidas em tubagem Tefzel como descrito (Eisenbraun *et al*, 1993; Pertmer *et al*, 1996). O bombardeamento de partículas foi realizado utilizando o sistema da transferência génica da Accell (documento PCT WO 95/19799). Para cada plasmídeo, foram imunizados murganhos C57BU6 fêmea nos dias 0, 21, 42 e 70. Cada administração consistiu em dois bombardeamentos com ADN/ouro, proporcionando uma dose total de aproximadamente 4-5 µg de plasmídeo.

1.2. Ensaio ELISPOT para respostas de célula T ao produto do gene de P501S

a) Preparação de esplenócitos

Os baços foram obtidos a partir de animais imunizados aos dias 7-14 após o reforço. Os baços foram processados através de moagem entre lâminas de vidro para produzir uma suspensão celular. Os glóbulos vermelhos foram lisados através de tratamento com cloreto de amônio e os restos foram removidos para deixar uma suspensão fina de esplenócitos. As células foram ressuspensas a uma concentração de 8×10^8 /mL em meio RPMI completo para utilização em ensaios ELISPOT.

b) Rastreio da biblioteca de péptidos

Uma biblioteca de péptidos cobrindo a maior parte da sequência de P501 S foi obtida da Corixa Corp. A biblioteca continha cinquenta péptidos de 15-20 monómeros sobrepondo em 4-11 péptidos de aminoácidos. Os péptidos são numerados 1-50. Além disso, foi utilizado um programa de previsão (H-G. Rammensee, et al.: Immunogenetics, 1999, 50: 213-219) (<http://syfpeithi.bmi-heidelberg.com/>) para prever epitopos putativos de Kb e Db a partir da sequência de P501 S. Os dez melhores epitopos para Kb e Db foram encomendados à Mimotopes (Reino Unido) e incluídos na biblioteca (péptidos 51-70). Para rastreio da biblioteca de péptidos, os péptidos foram utilizados numa concentração final de 50 µg/mL (aprox. 25-50 µM) em IFNγ e IL-2 ELISPOTS utilizando o protocolo descrito abaixo. Para IFNγ ELISPOTS, foi adicionado IL-2 aos ensaios a 10 ng/mL. Os esplenócitos utilizados para o rastreio foram recolhidos ao dia 84 de murganhos C57BU6 imunizados no dia 0, 21, 42 e 70. Foram identificados três péptidos a partir do rastreio da biblioteca - Péptidos 18 (HCRQAYSVYAFMISLGGLG), 22 (GLSAPSLSPHCCPCRARLAF) e

48 (VCLAAGITYVPPLLEVGV). Estes péptidos foram subsequentemente utilizados nos ensaios ELISPOT

c) Ensaio ELISPOT

As placas foram revestidas com 15 µg/mL (em PBS) de IFN γ de rato anti-murganho ou IL-2 de rato anti-murganho (Pharmingen). As placas foram revestidas de um dia para o outro a +4 °C. Antes da utilização, as placas foram lavadas três vezes com PBS. Os esplenócitos foram adicionados às placas a 4x10⁵ células/poço. Os péptidos identificados no rastreio da biblioteca foram novamente encomendados à Genemed Synthesis e utilizados a uma concentração final de 50 µg/mL. A proteína CPC-P501S (GSKBio) foi utilizada no ensaio a 20 µg/mL. Os ensaios ELISPOT foram realizados na presença de IL-2 (10 ng/mL), IL-7 (10 ng/mL) ou nenhuma citocina. O volume total em cada poço foi 200 µL. As placas contendo células estimuladas por péptido foram incubadas durante 16 horas numa incubadora a 37 °C humidificada.

e) Revelação de placas de ensaio ELISPOT.

As células foram removidas das placas por lavagem uma vez com água (com saturação de 10 minutos para assegurar lise das células) e três vezes com PBS. Foram adicionados IFN γ ou IL-2 de rato anti-murganho conjugados com biotina (Pharmingen) a 1 µg/mL em PBS. As placas foram incubadas com agitação durante 2 horas à temperatura ambiente. As placas foram depois lavadas três vezes com PBS antes da adição de fosfatase alcalina com estreptavidina (Caltag) numa diluição de 1/1000. Após três lavagens em PBS as manchas foram reveladas por incubação com substrato BCICP (Biorad) durante 15-45 minutos. O substrato foi removido por lavagem utilizando água e as placas foram deixadas a secar. As

manchas foram enumeradas utilizando um sistema de análise de imagem idealizado por Brian Hayes, Asthma Cell Biology Unit, GSK.

1.3. Ensaio de ELISA para anticorpos contra o produto do gene de P501 S

As amostras de soro foram obtidas dos animais através de venopunção nos dias -1, 28, 49 e 56, e ensaiadas para a presença de anticorpos anti-P501S. O ELISA foi realizado utilizando placas Nunc Maxisorp revestidas de um dia para o outro a 4 °C com 0,5 µg/mL de proteína CPC-P501S (GSKBio) em tampão de bicarbonato de sódio. Após lavagem com TBS-Tween (soro fisiológico tamponado com Tris, pH 7,4, contendo 0,05% de Tween 20) as placas foram bloqueadas com tampão de Bloqueamento (3% de BSA em tampão TBS-Tween) durante 2 h, à temperatura ambiente. Todos os soros foram incubados na diluição de 1:100 durante 1 hora, à t.a., em tampão de Bloqueamento. A ligação do anticorpo foi detectada utilizando imunoglobulinas de coelho anti-murganho conjugadas com HRP (P0260, Dako) na diluição de 1:2000 em tampão do Bloqueamento. As placas foram lavadas novamente e o conjugado ligado foi detectado utilizando reagentes corados Fast OPD (Sigma, Reino Unido). A reacção foi parada através da adição de ácido sulfúrico a 3 M e o produto OPD quantificado através da medição da absorvência a 490 nm.

1.4. Ensaios de transfecção transitória

A expressão de P501 S humano a partir de várias construções de ADN foi analisada através de transfecção transitória dos plasmídeos em células CHO (ovário de hamster Chinês) seguida por transferência de Western na proteína celular total. As transfecções transitórias foram realizadas com o reagente Transfectam (Promega) de acordo com as directrizes do fabricante. Resumidamente, foram inoculadas placas de cultura de tecidos de 24 poços com 5×10^4 de células CHO por poço em 1 mL de meio DMEM completo (DMEM, 10% de FCS, 2 mM de L-glutamina, 100 IU/mL de penicilina, 100 µg/mL de estreptomicina) e incubadas durante 16 horas a 37 °C. Foi adicionado 0,5 µg de ADN a 25 µL de NaCl a 0,3 M (suficiente para um poço) e foram adicionados 2 µL de Transfectam a 25 µL de Milli-Q. As soluções de ADN e Transfectam foram misturadas cuidadosamente e incubadas à temperatura ambiente durante 15 minutos. Durante este passo de incubação, as células foram lavadas uma vez em PBS e cobertas com 150 µL de meio sem soro (DMEM, 2 mM de L-glutamina). A solução de ADN-transfectam foi adicionada, gota a gota, às células, a placa agitada cuidadosamente e incubada a 37 °C durante 4-6 horas. Foram adicionados 500 µL de meio DMEM completo e as células incubadas durante mais 48-72 horas a 37 °C.

2. Análise de transferência de Western de células CHO transitoriamente transfectadas com plasmídeos de P501S

As células CHO transitoriamente transfectadas foram lavadas com PBS e tratadas com uma solução de Versene (1:5000)/0,025% de tripsina para transferir as células para suspensão. Após tripsinização, as células CHO foram sedimentadas e ressuspensas em 50 µL de PBS. Foi adicionado um volume igual de tampão de lise 2x NP40 e as células incubadas em gelo durante 30 minutos. Foram adicionados 100 µL de tampão de amostra 2x TRIS-Glicina

SDS (Invitrogen) contendo 50 mM de DTT e a solução foi aquecida a 95 °C durante 5 minutos. Foram aplicados 1-20 µL de amostra num gel de TRIS-Glicina a 4-20% de 1,5 mm (Invitrogen) e submetido a electroforese sob tensão constante (125 V) durante 90 minutos em tampão 1x TRIS-Glicina (Invitrogen). Foi utilizado um marcador de gama larga pré-corado (New England Biolabs, #P7708S) para determinar o tamanho das amostras. Após a electroforese, as amostras foram transferidas para membrana de PVDF Immobilon-P (Millipore), pré-molhada em metanol, utilizando um Xcell III Blot Module (Invitrogen), tampão de transferência 1x contendo 20% de metanol (Invitrogen) e uma tensão constante de 25 V durante 90 minutos. A membrana foi bloqueada de um dia para o outro a 4 °C em leite em pó magro contendo 3% de TBS-Tween (soro fisiológico tamponado com Tris, pH 7,4, contendo 0,05% de Tween 20) (Marvel). O anticorpo primário (10E3) foi diluído a 1:1000 e incubado com a membrana durante 1 hora à temperatura ambiente. Depois de lavagem exaustiva em TBS-Tween, o anticorpo secundário (imunoglobulinas de coelho anti-murganho conjugadas com HRP (#P0260, Dako)) foi diluído a 1:2000 em leite em pó magro contendo 3% de TBS-Tween e incubado com a membrana durante uma hora à temperatura ambiente. Depois de lavagem exaustiva, a membrana foi incubada com substrato Supersignal West Pico Chemiluminescent (Pierce) durante 5 minutos. O excesso de líquido foi removido e a membrana selada entre duas folhas de película aderente, e exposta a filme Hyperfilm ECL (Amersham-PharmaciaBiotech) durante 1-30 minutos.

3. Preparação da cassette de expressão de P501S humano de tamanho total

O ponto de partida para a construção de uma cassette de expressão de P501S foi o plasmídeo pcDNA3.1-P501 S (Corixa

Corp), que tem uma estrutura básica de pcDNA3.1 (Invitrogen) contendo uma cassette de ADNC de P501 S humano de tamanho total clonada entre os sítios EcoRI e NotI. Este vector é também denominado JNW673. A presença de P501S foi confirmada através de sequenciação fluorescente. A sequência da cassette de ADNC é dada pela sequência de NCBI/Genbank (número de acesso AY033593). O P501 S humano foi amplificado por PCR a partir de ADN molde JNW673, digerido com XbaI e SalI e clonado nos sítios NheI/XhoI de pVAC, produzindo o vector JNW680. A orientação correcta do fragmento relativamente ao promotor CMV foi confirmada através de PCR e por sequenciação de ADN. A sequência da cassette de expressão é mostrada na Figura 12 (SEQ ID N°: 17). Para construir uma cassette de expressão de CPC-p501 S, o CPC-p501 S foi amplificado por PCR a partir do vector pRIT15201 (ver Figura 7), digerido com XbaI e SalI e clonado nos sítios NheI e XhoI de pVAC, produzindo o plasmídeo JNW735. A orientação correcta foi confirmada através de PCR e sequenciação. A sequência da cassette de expressão de CPC-p501 S é mostrada na Figura 13 (SEQ ID N°:18).

4. Expressão de P501S humano a partir de plasmídeos JNW680 e JNW735

Os plasmídeos de expressão de P501S foram transitoriamente transfectados em células CHO e um lisado celular total preparado como descrito nos métodos. Uma transferência de Western de um lisado celular total identificou bandas únicas de aproximadamente 55 kDa e 62 kDa para amostras transfectadas com JNW680 e JNW735, respectivamente (Figura 21). Isto é consistente com os pesos moleculares previstos de 59,3 kDa e 63,3 kDa para P501 S e CPC-p501 S, respectivamente. A adição do marcador CPC não afecta adversamente a expressão de P501 S.

5. Resultados

5.1. Respostas de anticorpo a P501S humano após imunização de PMID

As respostas de anticorpo após imunização com pVAC (vector vazio) e pVAC-P501S (JNW680) foram avaliadas através de ELISA após uma imunização primária por PMID no dia 0 e três reforços no dia 21, dia 42 e dia 70. A Figura 22 mostra as respostas de anticorpo de soros recolhidos no dia -1, dia 28, dia 49 (murganhos A1-3, B1-3) e dia 56 (murganhos A4-6, B4-9). Embora existam algumas respostas não específicas ao vector vazio pVAC, foram observadas respostas específicas à construção P501 S em 5 de 9 murganhos.

5.2. Identificação dos novos epitopos de célula T a partir de P501S humano em murganhos C57BU6 por rastreio de uma biblioteca de péptidos de P501S

Após imunização com JNW680 (pVAC-P501S) por PMID no dia 0 e três reforços no dia 21, dia 42 e dia 70, os ensaios ELISPOT foram realizados no dia 84. Os péptidos da biblioteca de P501 S foram testados na concentração final de 50 µg/mL. A partir deste rastreio inicial, foram encontrados três péptidos que estimulam a secreção de IFN-γ e/ou IL-2. Péptidos 18, 22 e 48 (Figura 23). Estes péptidos foram utilizados em ensaios celulares subsequentes.

5.3. Respostas celulares a pVAC-P501S (JNW680) após imunização de PMID

As respostas celulares após imunização com pVAC (vector vazio) e pVAC-P501 S foram avaliadas por ELISPOT após imunização primária por PMID no dia 0 e três reforços no dia 21, 42 e 70. Os ensaios foram realizados 7 dias após o reforço. Foram utilizadas duas condições de ensaio diferentes: 1) os péptidos 18, 22 e 48 identificados no rastreio da biblioteca de péptidos utilizados à concentração final de 50 µg/mL e 2) a proteína CPC-P501S utilizada na concentração final de 20 µg/mL. A figura 24A mostra que embora não se tenham verificado respostas específicas de P501 S ao vector vazio (A4-6), a construção pVAC-P501 S induziu respostas específicas de IFN-γ aos Péptidos 18 e 22 em todos os murganhos (B6-9) enquanto que um murganho (B7) também mostrou uma resposta de IFN-γ ao Péptido 48. A figura 24B mostra que todos os murganhos apresentaram respostas específicas de IL-2 aos Péptidos 18, 22 e 48. Além disso, os murganhos imunizados com pVAC-P501 S (B6-9) também mostraram respostas de IL-2 moderadas a CPC-P501S, enquanto que os murganhos imunizados com vector vazio (A4-6) não apresentaram nenhuma resposta.

5.4. Comparação de respostas celulares a P501S e CPC-P501S após imunização PMID.

As respostas celulares após imunização com pVAC (vector vazio), pVAC-P501S (JNW680) e CPC-P501S (JNW735) foram avaliadas através de ELISPOT após uma imunização primária por PMID no dia 0 e reforços no dia 21 e 42. Os ensaios foram realizados 7 dias após o reforço. Foram utilizadas duas condições de ensaio diferentes: 1) os péptidos 18, 22 e 48 identificados no rastreio da biblioteca de péptidos utilizados à concentração final de 50 µg/mL e 2) a proteína CPC-P501S utilizada na concentração final de 20 µg/mL. A Figura 25 mostra que no dia 28, o CPC-p501 S induziu boas respostas de IL-2 a 10 µg/mL de péptido

22, enquanto não se verificaram respostas específicas de P501 S ao vector vazio ou a pVAC-P501 S. Estes resultados foram também observados utilizando a proteína CPC-p501 S para re-estimular os esplenócitos. No dia 49 (após 2º reforço), as respostas induzidas por P501 S e CPC-p501 S foram equivalentes. Estes dados sugerem que a adição do marcador CPC melhora a cinética e/ou magnitude da resposta a P501 S.

Exemplo IX. Experiências de imunogenicidade em estudos com murganhos utilizando Proteína P501S + adjuvantes

1. Concepção e formulação de adjuvante

A resposta imunológica induzida através de vacinação utilizando a proteína CPC-p501 S purificada recombinante formulada em adjuvantes é caracterizada em experiências realizadas em murganhos. Grupos de 5 a 10 murganhos C57BL6 fêmeas de oito semanas de idade são vacinados 2-6 vezes intramuscularmente, em intervalos de 2 semanas com 10 µg da proteína CPC-P501S formulada em sistemas adjuvantes diferentes. O volume administrado corresponde a 1/10 da dose humana (50 µL). A serologia (resposta total de Ig) e a resposta celular (linfoproliferação das células T e produção de citocinas) são analisadas em células do baço, 6-14 dias após a última vacinação utilizando protocolos convencionais como descritos em Gérard, C. *et al.*, 2001, Vaccine 19, 2583-2589.

São mostrados os dados de uma experiência representativa. Incluiu 5 grupos de oito murganhos C57BI/6 fêmeas que receberam 4 injeções intramusculares de CPC P501 (10 µg) + adjuvante (A, B, C) nos dias 0, 14, 28, 42. O Exemplo V proporciona um protocolo experimental de como realizar as formulações.

Resumidamente, as formulações adjuvantes são como se seguem (as quantidades são dadas para uma dose de 100 µL):

- **Adjuvante A:** QS21 (10 µg), MPL (10 µg) e CpG7909 (100 µg) preparado de acordo com o método divulgado no documento WO 00/62800;
- **Adjuvante B:** formulação de QS21 (20 µg), MPL (20 µg), CpG7909 (100 µg) e 50 µL de emulsão óleo-em-água SB62 (documento WO 95/17210);
- **Adjuvante C:** formulação de QS21 (10 µg), MPL (10 µg), CpG7909 (100 µg) e 10 µL de emulsão óleo-em-água SB62 (documento WO 99/12565).

2. Serologia

A resposta Ig total induzida através de vacinação foi medida por ELISA utilizando CPC-p501 ou RA12-P501 (C-term, que é uma forma truncada da proteína P501 correspondente ao C-terminal da proteína fundida no seu N-terminal, a uma proteína RA12 derivada de TB - Ra12 é derivada a partir do antígeno MTB32A descrito em Skeiky *et al.*, *Infection and Immun.* (1999) 67:3998-4007). As proteínas CPC-p501 S adjuvantadas dão uma boa resposta de anticorpo após vacinação.

3. Resposta celular

3.1. Linfoproliferação

7 Dias após a última vacina, a linfoproliferação foi realizada em células do baço individualmente. Foram plaqueadas 2,10e5 células do baço em quadruplicados, em microplaca de 96 poços, em meio RPMI contendo 1% de soro normal de murganho. Após

72 horas de re-estimulação com o imunogénio (CPC-p501) ou a proteína truncada (RA12 P501) a concentração diferente, foi adicionado 1 μ Ci de ^{3} H timidina (Amersham, 5 Ci/mL). Após 16 horas, as células foram recolhidas em placas de filtro. A radioactividade incorporada foi medida num contador β . Os resultados são expressos em CPM ou como índices de estimulação* (CPM geomean em culturas com antigénio / CPM geomean em culturas sem antigénio). A re-estimulação com ConA (2 μ g/mL) como controlo positivo foi incluída como controlo positivo.

Como mostrado na Figura 26, é observada uma linfoproliferação específica de P501 no baço de todos os grupos de murganhos que receberam a proteína adjuvantada após re-estimulação *in vitro* com o imunogénio ou outra proteína P501 produzida noutro sistema de expressão (*E. coli*), indicando que as células T foram iniciadas *in vivo* através da vacinação.

3.2. Produção de IFN γ medida através de coloração intracelular de células do baço

Células Dendríticas de Medula Óssea (BMDC) obtidas após cultura de PBL de murganho durante 7 dias na presença de GMCSF. 7 Dias após a última vacina, o baço ou PBL são recolhidos e é preparada uma suspensão celular. Foram incubadas 10e6 células (1 agrupamento por grupo) +/- 18 h com 10e5 BMDC pulsadas de um dia para o outro com 10 μ g/mL de proteína CPCp501 ou RA12. Após um tratamento com o anticorpo 2.4.G.2, as células do baço foram coradas com anticorpos anti CD4 e CD8 fluorescentes (anti CD4-APC e anti CDBPerCP). Após um passo de permeabilização e fixação, as células foram coradas com anticorpo anti IFN γ -FITC fluorescente.

Em murganhos vacinados com CPC P501 em adjuvante diferente, é mostrado que ambas as células T CD4 e CD8 produzem IFN γ em resposta a DC pulsado com o imunogénio e o p501 C-termo produzido em *E. coli* (como mostrado através de coloração intracelular do baço e PBL). Existe um aumento de 4-10X na % de células que produzem esta citocina nos grupos que receberam o CPC-p501 S adjuvantado comparativamente à proteína isolada, e é mostrado que entre 0,1 a 10% de células T CD4 ou CD8 produzem IFN γ .

Em conclusão, estes dados permitem concluir que a proteína CPC-p501 adjuvantada é imunogénica em murganhos. Tanto uma resposta humoral específica de P501 como celular incluindo produção de IFN γ por células T CD4 e CD8 podem ser detectadas após diversas vacinações intramusculares com CPC P501 em adjuvantes.

Exemplo X. Construções e sequências de CPC-MUC-1

A sequência de CPC é retirada da sequência de nucleótidos SEQ ID N° 28. A sequência MUC1 está disponível a partir da base de dados Genbank (número de acesso NM_002456).

1. Construção de MUC1-CPC

Devido à presença de uma sequência sinal em MUC1 que é clivada pós-tradução, o motivo CPC foi colocado no C-terminal. A sequência de ADN resultante de MUC1-CPC é representada na SEQ ID N° xx (figura 28A) e a sequência de proteína correspondente de MUC1-CPC em SEQ ID N° yy (figura 28B).

2. Construção de ss-CPC-MUC1

Devido à presença de uma sequência sinal em MUC1 que é clivada pós-tradução, a sequência sinal MUC1 foi substituída por uma sequência líder heteróloga (da cadeia pesada da imunoglobulina humana) e o motivo CPC foi introduzido entre a sequência líder heteróloga e a sequência MUC1, produzindo uma sequência denominada ss-CPC-MUC1 como representada na figura 29.

LISTAGEM DE SEQUÊNCIAS

<110> GlaxosmithKline Biologicals sa
Glaxo Group Ltd

<120> Composições imunogénicas

<130> B45311

<160> 52

<170> FastSEQ para Windows versão 4.0

<210> 1

<211> 15

<212> PRT

<213> Streptococcus pneumoniae

<400> 1

Gly	Trp	Gln	Lys	Asn	Asp	Thr	Gly	Tyr	Trp	Tyr	Val	His	Ser	Asp
1				5					10					15

<210> 2

<211> 21

<212> PRT

<213> Streptococcus pneumoniae

<400> 2

Gly	Ser	Tyr	Pro	Lys	Asp	Lys	Phe	Glu	Lys	Ile	Asn	Gly	Thr	Trp	Tyr
1				5					10					15	
Tyr	Phe	Asp	Ser	Ser											
				20											

<210> 3

<211> 22
 <212> PRT
 <213> Streptococcus pneumoniae
 <400> 3

Gly	Tyr	Met	Leu	Ala	Asp	Arg	Trp	Arg	Lys	His	Thr	Asp	Gly	Asn	Trp
1				5					10					15	
Tyr	Trp	Phe	Asp	Asn	Ser										
			20												

<210> 4
 <211> 20
 <212> PRT
 <213> Streptococcus pneumoniae
 <400> 4

Gly	Glu	Met	Ala	Thr	Gly	Trp	Lys	Lys	Ile	Ala	Asp	Lys	Trp	Tyr	Tyr
1				5					10					15	
Phe	Asn	Glu	Glu												
			20												

<210> 5
 <211> 21
 <212> PRT
 <213> Streptococcus pneumoniae
 <400> 5

Gly	Ala	Met	Lys	Thr	Gly	Trp	Val	Lys	Tyr	Lys	Asp	Thr	Trp	Tyr	Tyr
1				5					10					15	
Leu	Asp	Ala	Lys	Glu											
			20												

<210> 6
 <211> 23

<212> PRT

<213> Streptococcus pneumoniae

<400> 6

Gly	Ala	Met	Val	Ser	Asn	Ala	Phe	Ile	Gln	Ser	Ala	Asp	Gly	Thr	Gly
1				5					10					15	
Trp	Tyr	Tyr	Leu	Lys	Pro	Asp									
			20												

<210> 7

<211> 142

<212> PRT

<213> Streptococcus pneumoniae

<400> 7

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

<210> 8

<211> 112

<212> PRT

<213> Streptococcus pneumoniae

<400> 8

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

<210> 9

<211> 45

<212> ADN

<213> Streptococcus pneumoniae

<400> 9

ggctggcaga agaatgacac tggctactgg tacgtacatt cagac 45

<210> 10

<211> 63

<212> ADN

<213> Streptococcus pneumoniae

<400> 10

ggctcttattc	caaaagacaa	gtttgagaaa	atcaatggca	cttggtacta	ctttgacagt	60
tca						63

<210> 11

<211> 66

<212> ADN

<213> Streptococcus pneumoniae

<400> 11

```
ggctatatgc ttgcagaccg ctggaggaag cacacagacg gcaactggta ctggttcgac 60
aactca                                         66
```

<210> 12

<211> 60

<212> ADN

<213> Streptococcus pneumoniae

<400> 12

```
ggcgaaatgg ctacaggctg gaagaaaatc gctgataagt ggtactatatt
caacgaagaa          60
```

<210> 13

<211> 63

<212> ADN

<213> Streptococcus pneumoniae

<400> 13

```
ggtgccatga agacaggctg ggtcaagtac aaggacactt ggtactactt agacgctaaa 60
gaa                                         63
```

<210> 14

<211> 69

<212> ADN

<213> Streptococcus pneumoniae

<400> 14

ggcgccatgg tatcaaatgc ctttatccag tcagcggacg gaacaggctg gtactacctc 60
aaaccagac 69

<210> 15
<211> 429
<212> ADN
<213> Streptococcus pneumoniae

<400> 15

ggctggcaga agaatgacac tggctactgg tacgtacatt cagacggctc ttatccaaaa 60
gacaagtttg agaaaatcaa tggcacttgg tactactttg acagttcagg ctatatgctt 120
gcagaccgct ggaggaagca cacagacggc aactggtact ggttcgacaa ctcaggcgaa 180
atggctacag gctggaagaa aatcgctgat aagtgggtact atttcaacga agaagggtgcc 240
atgaagacag gctgggtcaa gtacaaggac acttgggtact acttagacgc taaagaaggc 300
gccatggtat caaatgcctt tatccagtca gcggacggaa caggctggta ctacctcaa 360
ccagacggaa cactggcaga caggccagaa ttcacagtag agccagatgg cttgattaca 420
gtaaaataa 429

<210> 16
<211> 336
<212> ADN
<213> Streptococcus pneumoniae

<400> 16

tacgtacatt ccgacggctc ttatccaaaa gacaagtttg agaaaatcaa tggcacttgg 60
tactactttg acagttcagg ctatatgctt gcagaccgct ggaggaagca cacagacggc 120
aactggtact ggttcgacaa ctcaggcgaa atggctacag gctggaagaa aatcgctgat 180
aagtgggtact atttcaacga agaagggtgcc atgaagacag gctgggtcaa gtacaaggac 240
acttgggtact acttagacgc taaagaaggc gccatggtat caaatgcctt tatccagtca 300
gcggacggaa caggctggta ctacctcaa ccagac 336

<210> 17
<211> 1674
<212> ADN
<213> Homo sapiens

<400> 17

```
gccaccatgg tccagaggct gtgggtgagc cgcttgcctgc ggcaaccgaa agcccagctc 60
ttgctgggtca acctgctaac ctttggcctg gaggtgtgtt tggccgcagg catcacctat 120
gtgccgcctc tgctgctgga agtgggggta gaggagaagt tcatgaccat ggtgctgggc 180
attggtccag tgctgggcct ggtctgtgtc ccgctcctag gctcagccag tgaccactgg 240
cgtggacgct atggccgccg ccggcccttc atctgggcac tgctcttggg catcctgctg 300
agcctctttc tcatcccaag ggccggctgg ctagcagggc tgctgtgccc ggatcccagg 360
cccctggagc tggcactgct catcctgggc gtggggctgc tggacttctg tggccagggtg 420
tgcttcactc cactggaggc cctgctctct gacctcttcc gggacccgga ccactgtcgc 480
caggcctact ctgtctatgc cttcatgatc agtcttgggg gctgcctggg ctacctcctg 540
cctgccattg actgggacac cagtgccttg gccccctacc tgggcaccca ggaggagtgc 600
ctctttggcc tgctcaccct catcttcttc acctgcctag cagccacact gctgggtggc 660
gaggaggcag cgctgggccc caccgagcca gcagaagggc tgcgggcccc ctcttgtctg 720
ccccactgct gtccatgccg ggcccgttg gctttccgga acctgggccc cctgcttccc 780
cggctgcacc agctgtgctg ccgcatgccc cgcacctgc gccggctctt cgtggctgag 840
ctgtgcagct ggatggcact catgaccttc acgctgtttt acacggattt cgtgggagag 900
gggctgtacc agggcgtgcc cagagctgag ccgggcaccg agggccggag acactatgat 960
gaaggcgctt ggatgggcag cctggggctg ttcttgagc gcgccatctc cctggtcttc 1020
tctctggtca tggaccggct ggtgcagcga ttccggcactc gagcagctta tttggccagt 1080
tgggcagctt tccctgtggc tgccgggtgc acatgcctgt ccacagtggt ggccgtggtg 1140
acagcttcag ccgccctcac cgggttcacc ttctcagccc tgcagatcct gccctacaca 1200
ctggcctccc tctaccaccg ggagaagcag gtgttcctgc ccaaataccg aggggacact 1260
ggagggtgcta gcagtgagga cagcctgatg accagcttcc tgcaggccc taagcctgga 1320
gctcccttcc ctaatggaca cgtgggtgct ggaggcagtg gcctgctccc acctccacc 1380
gcgctctgcg gggcctctgc ctgtgatgtc tccgtacgtg tgggtggtgg tgagccacc 1440
gaggccaggg tggttccggg ccggggcatc tgcttgacc tcgccatcct ggatagtgcc 1500
ttctgctgtg ccaggtggc cccatccctg tttatgggct ccattgtcca gctcagccag 1560
tctgtcactg cctatatggt gtctgccgca ggcctgggtc tggtcgccat ttactttgct 1620
acacaggtag tatttgacaa gagcgacttg gccaaatact cagcgtaggt cgag 1674
```

<210> 18

<211> 1947

<212> ADN

<213> Sequência Artificial

<220>

<223> Gene híbrido entre C-LytA de St. pneum., epitopo P2 de auxiliar T e P501S humano.

<400> 18

```

gccaccatgg cggccgctta cgtacattcc gacggctctt atccaaaaga caagtttgag 60
aaaatcaatg gcacttggtg ctactttgac agttcaggct atatgcttgc agaccgctgg 120
aggaagcaca cagacggcaa ctggtactgg ttcgacaact caggcgaaat ggctacaggg 180
tggaagaaaa tcgctgataa gtggtactat ttcaacgaag aaggtgccat gaagacaggg 240
tgggtcaagt acaaggacac ttggtactac ttagacgcta aagaaggcgc catgcaatac 300
atcaaggcta actctaagtt cattgggtatc actgaaggcg tcatgggtatc aaatgccttt 360
atccagtcag cggacggaac aggctggtag tacctcaaac cagacggaac actggcagac 420
aggccagaaa agttcatgta catggtgctg ggcattgggtc cagtgtctggg cctggtctgt 480
gtcccgtctc taggctcagc cagtgaccac tggcgtggac gctatggccg ccgccggccc 540
ttcatctggg cactgtcctt gggcatcctg ctgagcctct ttctcatccc aagggccggc 600
tggctagcag ggctgtctgt cccggatccc aggccctgg agctggcact gctcatcctg 660
ggcgtggggc tgctggactt ctgtggccag gtgtgttca ctccactgga ggccctgtct 720
tctgacctct tccgggaccc ggaccactgt cgccaggcct actctgtcta tgccttcatg 780
atcagtcctg ggggctgcct gggctacctc ctgctgcca ttgactggga caccagtgcc 840
ctggccccc acctgggcac ccaggaggag tgcctctttg gcctgtcac cctcatcttc 900
ctcacctgcg tagcagccac actgctgggtg gctgaggagg cagecgtggg cccaccgag 960
ccagcagaag ggctgtcggc cccctccttg tgcgccact gctgtccatg ccgggcccgc 1020
ttggctttcc ggaacctggg cgccctgctt ccccgctgc accagctgtg ctgccgatg 1080
ccccgcacc tgcgccggct ctctgtgggt gagctgtgca gctggatggc actcatgacc 1140
ttcacgtgtg ttacacgga ttctgtgggc gaggggctgt accagggcgt gccagagct 1200
gagccgggca ccgagggccg gagacactat gatgaaggcg ttcggatggg cagcctgggg 1260
ctgttcctgc agtgcccat ctccctggtc ttctctctgg tcatggaccg gctggtgcag 1320
cgattcggca ctcgagcagt ctatttggcc agtgtggcag ctttccctgt ggctgccggt 1380
gccacatgcc tgtcccacag tgtggccgtg gtgacagctt cagccgccct caccgggttc 1440
accttctcag cctgcagat cctgccctac aactggcct ccctctacca ccgggagaag 1500
atgaccagct tcctgccagg ccctaagcct ggagctccct tccctaattg acacgtgggt 1620
gctggaggca gtggcctgct cccacctcca cccgcgtct gcggggcctc tgcctgtgat 1680
gtctccgtac gtgtgggtgt ggggtgagccc accgagcca ggggtggttc gggccggggc 1740
atctgcctgg acctcgccat cctggatagt gccttctctg tgtcccaggt ggccccatcc 1800
ctgtttatgg gctccattgt ccagctcagc cagtctgtca ctgcctatat ggtgtctgcc 1860
gcaggcctgg gtctggtcgc catttacttt gctacacagg tagtatttga caagagcgac 1920
ttggccaaat actcagcgta ggtcgag 1947

```

<210> 19

<211> 1662

<212> ADN

<213> Sequência Artificial

<220>

<223> P501S humano de codão otimizado

<400> 19

```

atggtgcagc ggctctgggt gagccgcctc ctgcggcatc gcaaggccca gtccttctgt 60
gtgaatctgc tcacattcgg cctggagggtg tgcctggccg ccggcatcac ctacgtgccc 120
ccccctctgc tggagggtggg agtcgaggag aagttcatga ccatgggtgt gggcattggg 180
cccgctcctgg gcctcgtgtg cgtgcctctc ctcggcagcg ctcccgacca ttggcgcggc 240
cggtatggcc gcaggagacc ctcatctctg gctctgagtc tcggcatcct gctgagcctg 300
ttcctgatcc ctcgggccgg ctggctggcc gggctgctgt gcccgatcc tcggccccctg 360
gagctggccc tgctgacctt cggcgtgggc ctgctggact tctgcggcca ggtgtgcttc 420
acgccccctg aggcactgct gagcgacctg ttccgggacc ccgaccattg ccgccaggcg 480
tacagcgtgt acgccttcat gatctccctg ggaggctgcc tgggctacct gctccccgcc 540
atcgattggg acaccagcgc actcgcccc tctctcgga caccaggagg atgctgttc 600
ggattgttga cgtcatctt cctcacgtgc gtccgggcca ccctgttggg ggccgaggag 660
gccgcctctg ggcccaccga gccggccgag ggactgagcg ccccgagcct gagtccacac 720
tgctgccctt gccggggccc cctggccttc cgtaatctgg gcgcctcct gctcggctc 780
catcagctgt gttgcagaat gcctaggacg ctgcggcgcc tgttcgtcgc tgagttgtgc 840
tcctggatgg ctctcatgac ctccaccctg tttatacgg acttcgtcgg ggagggcctg 900
taccaggggg tgccgcgcgc cgagcccggg acagaggcgc gccgccacta cgacgaggga 960
gtgcgtatgg gctccctggg cctcttcttg cagtgcgcca tcagtctggt tttctctctg 1020
gtcatggaca ggctggtgca gcgcttcgga acccgggcgg tgtacctggc gagcgtggcc 1080
gccttccccg tggctgccgg cgccacctgc ctctctcact cgggtggcgt ggtcaccgcc 1140
agcgccgccc tgaccgggtt caccctctct gccctgcaga ttctgcctta caccctggcc 1200
agcctgtacc atcgcgagaa acaggtgttt ctccccaaat acagaggcga caccgggggc 1260
gcctccagcg aggacagcct catgacctcc ttctgcctg gcccgaagcc cggcgccccct 1320
ttccccaaac ggacagtggt cgccggcggg agtgggctcc tgcccccccc tcctgcgctg 1380
tgcggggcca gcgcctgcga cgtgagcgtg cgcgtggtgg tgggcgagcc caccgaggcc 1440
cgcgtggtgc cgggcagagg catttgtctg gacctggcca tcctcgactc cgcttctctc 1500
ctcagccagg tggccccgtc cctcttcatg ggctctatcg tccagctgtc tcagagcgtc 1560
accgcttaca tgggtgtccgc tgctggactg ggcttgggtg ctatttattt cgccaccag 1620
tggggtgttc acaagagcga cctggccaaa tactccgcct ga 1662

```

<210> 20

<211> 1662

<212> ADN

<213> Sequência Artificial

<220>

<223> P501S humano de códon otimizado

<400> 20

```

atggtgcagc ggctgtgggt gtcccggtg ctgcgccata gaaaggccca gttgctgctg 60
gtgaacctgc tgactttcgg actggagggtg tgcctggctg ccgggatcac gtacgtgccc 120
ccccctctgc tggagggtggg cgtggaggag aagttcatga caatgggtgt gggcatcggc 180
cccgctcctgg gcctcgtgtg tgtgccctc ctcgggagtg cgtccgatca ttggcggggc 240
cgctacggcc gccgcagacc gttcatctgg gccctgagcc tggggatcct gctctctctc 300
ttcctgatcc cccggggccg ctggctggcc ggcttctgt gtcccgacc ccgcctctct 360

```

```

gagctggccc tcctgacccg gggcgtgggc ttgttggaact tctgcggcca ggtgtgtttc 420
actcccttgg aggtctctgct ctccgacctc ttccgcgacc ccgaccactg taggcaggct 480
tacagcgtgt acgccttcat gatcagtctg gggggatgcc tgggctatct gctgcccgtc 540
atcgaactggg acaccagcgc cctggccccc tacctgggga ctcaggagga gtgcctgttc 600
ggcctgctca ccttgatctt cctgacgtgc gtcgcgcgca ccctgctggt ggccgaggag 660
gcggccctgg ggccaccga gccgcgcgag ggctgagcg ctcccagcct gagccccat 720
tgctgcccgt gcagggctag gctcgccttc aggaatctgg gcgctttgct gcccgcctg 780
catcagctgt gctgtcgcat gcctcgacc ctgcgcgcgc tgctcgtcgc tgagctctgt 840
tcctggatgg ccctgatgac gttcaccctc ttctacaccg acttcgtggg ggagggcctg 900
taccagggcg tgcccagggc cgagcccgcc accgaggcta ggccgatta cgacgagggc 960
gtcaggatgg gctctctggg cctcttccctg cagtgcgcca tcagtctggt gttctctctg 1020
gtgatggacc ggctgggtgca gcgcttcggc acccgggccg tgtacctcgc ctctgtggcg 1080
gctttccccc tcgcgcgcgg cgcgacctgc ctgtctcatt ctgtcgccgt ggtgaccgcc 1140
agcgcgcgcc tgaccggctt caccctcagt gcgctccaga ttctgcccta caccctggcg 1200
tctctgtacc atcgcgagaa gcagggtgtt ctgcccaggt acccggggga cacaggggga 1260
gcttctctctg aggacagcct gatgaccagc ttcttgcccg gcccgaagcc gggggccctc 1320
ttcccacaac gccatgtcgg ggcgggcggc agcggcctgc tccctcccc cccgcctcctg 1380
tgcggcgcta gtgctgcga cgtgagcgtg cgggtgggtg tgggggagcc caccgaggct 1440
agggcgtgct ctggccgggg gatctgcctg gacctggcca tctcgaactc cgccttctctg 1500
ctctcccagg tggcgccag cctgttcatt ggcagtatcg tgcagctgag ccagagcgtg 1560
accgcctaca tggtagcgc gcgcggcctg ggggtgggtg ccatctactt tgccaccag 1620
gtcgtgttcg acaagagcga tctcgccaag tatagcgct ga 1662

```

<210> 21

<211> 1688

<212> ADN

<213> Sequência Artificial

<220>

<223> P501S humano de códon otimizado

<400> 21

```

gacggctagc gccaccatgg tgcagcggct ctgggtgagc cgcctcctgc ggcacgcaa 60
ggcccagctc ctgctgggtga atctgctcac attcggcctg gaggtgtgcc tggccgccc 120
catcacctac gtgccccccc tcctgctgga ggtgggagtc gaggagaagt tcatgaccat 180
ggtgctgggc attgggcccg tcctgggcct cgtgtgcgtg cctctcctcg gcagcgcttc 240
cgaccattgg cgcggccggt atggccgcag gagacccttc atctgggctc tgagtctcgg 300
cactctgctg agcctgttcc tgatccctcg ggccggctgg ctggccgggc tgetgtgccc 360
cgatcctcgg cccctggagc tggccctgct gatcctcggc gtgggcctgc tggacttctg 420
cggccagggtg tgettcacgc cctggaggc actgctgagc gacctgttcc gggacccga 480
ccattgccgc caggcgtaac gcgtgtacgc cttcatgate tccctgggag gctgcctggg 540
ctacctgctc cccgcctcgt attgggacac cagcgcactc gcccctatc tcggaacaca 600
ggagggaatgc ctgttcggac tgctgacgct catcttcctc acgtgcgtcg cggccacct 660

```

```

gttggtggcc gaggagggcc ccctggggcc caccgagccg gccgagggac tgagcgcccc 720
gagcctgagt ccacactgct gcccttgccg ggcccgcctg gccttccgta atctgggcgc 780
cctcctgcct cggtcccatc agctgtgttg cagaatgcct aggacgctgc ggcgctgtt 840
cgtcgctgag ttgtgtcctt ggatggctct catgaccttc accctgtttt atacggactt 900
cgtcggggag ggcctgtacc agggggtgcc gcgcgccgag cccgggacag aggcgcgccg 960
ccactacgac gagggagtgc gtatgggctc cctgggcctc ttcttgagct gcgccatcag 1020
tctggttttc tctctggtca tggacaggct ggtgcagcgc ttcggaaccc gggcggtgta 1080
cctggcgagc gtggccgcct tccccgtggc tggcgccgcc acctgcctct ctactcggc 1140
ggcctggtc accgccagcg ccgccctgac cgggttcacc ttctctgcc tgcagattct 1200
gccttacacc ctggccagcc tgtaccatcg cgagaaacag gtgtttctcc ccaagtacag 1260
aggcgacacc gggggcgctt ccagcgagga cagcctcatg acctccttcc tgcctggccc 1320
caagcccggc gcccctttcc ccaacgggca cgtggcgccc ggcgggagtg ggctcctgcc 1380
ccccctcct gcgtgtgctg gggccagcgc ctgcgacgtg agcgtgcgcg tgggtgggtgg 1440
cgagcccacc gagggccgcg tggtgccggg cagaggcatt tgtctggacc tggccatcct 1500
cgactccgcc ttctctctca gccagggtgg cccgtccctc ttcatgggct ctatcgtcca 1560
gctgtctcag agcgtcaccg cttacatggt gtccgctgct ggactgggct tgggtggctat 1620
ttatttcgcc acccagggtg tggttcgacaa gagcgacctg gccaaatact ccgcctgact 1680
cgaggcag                                     1688

```

<210> 22

<211> 1688

<212> ADN

<213> Sequência Artificial

<220>

<223> P501S humano de códon otimizado

<400> 22

```

gacggctagc gccaccatgg tgcagcggct gtgggtgtcc cggctgctgc gccatagaaa 60
ggcccagttg ctgctggtga acctgctgac ttccggactg gaggtgtgcc tggctgccgg 120
gatcacgtac gtgccccccc tgctgctgga ggtgggcgtg gaggagaagt tcatgacaat 180
ggtgctgggc atcgcccccg tcctgggcct cgtgtgtgtg cccctcctcg ggagtgcgtc 240
cgatcattgg cggggccgct acggccgcgc cagaccgttc atctgggccc tgagcctggg 300
catcctgctc tctctcttcc tgatcccccg ggccggctgg ctggccggcc tggctgtgtc 360
cgacccccgc cctctggagc tggccctcct gatectgggc gtgggcctgc tggacttctg 420
cggccaggtg tgtttcactc ccctggaggc tctgctctcc gacctcttcc gcgacccga 480
ccactgtagg caggcttaca gcgtgtacgc cttcatgac agtctggggg gatgcctggg 540
ctatctgctg cccgctatcg actgggacac cagcgccctg gcccctacc tggggactca 600
ggaggagtgc ctgttcggcc tgctcacctt gatcttctcg acgtgcgtcg ccgccaccct 660
gctggtggcc gaggagggcg ccctggggcc caccgagccc gccgagggcc tgagcgctcc 720
cagcctgagc cccattgct gcccgtcag ggctaggctc gccttcagga atctgggcgc 780
tttgctgccc cgcctgcac agctgtgctg tcgcatgcct cgcaccctgc gccgcctgtt 840
cgtcgctgag ctctgttctt ggatggccct gatgacgttc acctcttct acaccgactt 900
cgtgggggag ggcctgtacc agggcgtgcc cagggccgag cccggcaccg aggctaggcg 960

```

```

ccattacgac gagggcgtea ggatgggctc tctgggcctc ttcctgcagt gcgccatcag 1020
tctggtgttc tctctggtga tggaccggct ggtgcagcgc ttcggcaccg gggccgtgta 1080
cctcgectct gtggcggtt tccccgtcgc cgccggcgcg acctgcctgt ctcatctctgt 1140
cgccgtggtg accgccagcg ccgccctgac cggcttcacc ttcagtgcgc tccagattct 1200
gccctacacc ctggcgctct tgtaccatcg cgagaagcag gtgttcctgc ccaagtaccg 1260
cggggacaca gggggagctt cctctgagga cagcctgatg accagcttct tgccccggcc 1320
caagccgggg gccccttcc ccaacggcca tgcggggcg ggcggcagcg gcctgctccc 1380
tcccccccc gccctgtgcg gcgctagtgc ctgcgacgtg agcgtgcggg tgggtggtggg 1440
ggagcccacc gaggctaggg tcgtgcctgg ccgggggatc tgcttgacc tggccatcct 1500
cgactccgcc ttcctgctct ccaggtggc gccagcctg ttcattggga gtatcgtgca 1560
gctgagccag agcgtgaccg cctacatggt gagcgccgc ggcctggggt tgggtggccat 1620
ctactttgcc acccaggtcg tggtcgacaa gagcgatctc gccaaagtata gcgcctgact 1680
cgaggcag

```

<210> 23

<211> 435

<212> ADN

<213> Sequência Artificial

<220>

<223> Gene híbrido entre C-LytA de St. pneum., epitopo P2 de auxiliar T e uma pequena porção da extremidade 5' de P501S humano

<400> 23

```

atggcgggcg cttacgtaca ttccgacggc tcttatccaa aagacaagtt tgagaaaatc 60
aatggcactt ggtactactt tgacagttca ggctatatgc ttgcagaccg ctggaggaag 120
cacacagacg gcaactggta ctggttcgac aactcaggcg aaatggctac aggctggaag 180
aaaatcgctg ataagtggta ctatttcaac gaagaagggtg ccatgaagac aggctgggtc 240
aagtacaagg aacttggta ctacttagac gctaaagaag gcgccatgca atacatcaag 300
gctaactcta agttcattgg tatcactgaa ggcgtcatgg tatcaaagtc ctttatccag 360
tcagcggacg gaacaggctg gtactacctc aaaccagacg gaacactggc agacaggcca 420
gaaaagttca tgtac

```

<210> 24

<211> 435

<212> ADN

<213> Sequência Artificial

<220>

<223> Gene híbrido entre C-LytA de St. pneum., epitopo P2 de auxiliar T e uma pequena porção da extremidade 5' de P501S humano - codão otimizado

<400> 24

```
atggccgccg cctacgtgca tagcgacggg agctacccca aggacaagtt cgagaagatc 60
aacgggacat ggtactactt cgactcctcc ggctacatgc tcgccgaccg ctggcggaag 120
cacaccgacg gcaactggta ctggttcgat aactcgggag agatggccac cggctggaag 180
aagatcgccg acaagtggta ctatttcaac gaggagggcg ccatgaagac cggctgggtg 240
aagtataagg acacctggta ctacctcgac gccaaaggagg gcgccatgca gtatatcaag 300
gccaacagca agttcatcgg catcaccgag ggagtgatgg tcagcaacgc ctttatccag 360
agcgccgacg gcaccggatg gtactacttg aagccggacg gcaccctcgc ggatcggccc 420
gagaagttca tgtac                                     435
```

<210> 25

<211> 435

<212> ADN

<213> Sequência Artificial

<220>

<223> Gene híbrido entre C-LytA de St. pneum., epitopo P2 de auxiliar T e uma pequena porção da extremidade 5' de P501S humano - codão otimizado

<400> 25

```
atggccgccg cctacgtgca cagcgacggg tcctacccaa aggacaagtt cgagaagatc 60
aacggcacgt ggtactatct cgacagcagc ggctacatgc tcgccgatcg ctggcgcaag 120
cacaccgacg ggaactggta ctggttcgac aactctggcg agatggctac ggggtggaag 180
aagatcgccg acaagtggta ctacttcaac gaggagggcg ccatgaagac cgggtgggtg 240
aagtacaagg acacctggta ctacctggac gctaaggagg gcgccatgca gtacatcaag 300
gccaaactcga agttcatcgg gatcaccgag ggcgtgatgg tcagtaacgc tttcatccag 360
agcgcgacg gcacaggctg gtattacctg aagcccgatg gcaccctggc ggacagacct 420
gagaaattca tgtac                                     435
```

<210> 26

<211> 464

<212> ADN

<213> Sequência Artificial

<220>

<223> Gene híbrido entre C-LytA de St. pneum., epitopo P2 de auxiliar T e uma pequena porção da extremidade 5' de P501S humano - codão otimizado

<400> 26

```
gacggctagc gccaccatgg ccgccccta cgtgcatagc gacgggagct accccaagga 60
caagttcgag aagatcaacg ggacatggta ctacttcgac tcctccggt acatgctcgc 120
cgaccgctgg cggaagcaca ccgacggcaa ctggtactgg ttcgataact cgggagagat 180
ggccaccggc tggaagaaga tcgcggaaca gtggtactat ttcaacgagg agggcgccat 240
gaagaccggc tgggtgaagt ataaggacac ctggtactac ctcgacgcca aggagggcgc 300
catgcagtat atcaaggcca acagcaagtt catcggcata accgaggag tgatggtcag 360
caacgccttt atccagagcg ccgacggcac cggatggtag tacttgaagc cggacggcac 420
cctcgcggat cggcccgaga agttcatgta ctgactcgag gcag 464
```

<210> 27

<211> 652

<212> PRT

<213> Sequência Artificial

<220>

<223> Proteína híbrida entre C-LytA de St. pneum., epitopo P2 de auxiliar T e aminoácidos 51-553 de P501S humano

<400> 27

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

Lys	Trp	Tyr	Tyr	Phe	Asn	Glu	Glu	Gly	Ala	Met	Lys	Thr	Gly	Trp	Val
65					70					75					80
Lys	Tyr	Lys	Asp	Thr	Trp	Tyr	Tyr	Leu	Asp	Ala	Lys	Glu	Gly	Ala	Met
			85						90					95	
Gln	Tyr	Ile	Lys	Ala	Asn	Ser	Lys	Phe	Ile	Gly	Ile	Thr	Glu	Gly	Val
			100					105					110		
Met	Val	Ser	Asn	Ala	Phe	Ile	Gln	Ser	Ala	Asp	Gly	Thr	Gly	Trp	Tyr
		115					120					125			
Tyr	Leu	Lys	Pro	Asp	Gly	Thr	Leu	Ala	Asp	Arg	Pro	Glu	Lys	Phe	Met
	130					135					140				
Tyr	Met	Val	Leu	Gly	Ile	Gly	Pro	Val	Leu	Gly	Leu	Val	Cys	Val	Pro
145					150					155					160
Leu	Leu	Gly	Ser	Ala	Ser	Asp	His	Trp	Arg	Gly	Arg	Tyr	Gly	Arg	Arg
			165						170					175	
Arg	Pro	Phe	Ile	Trp	Ala	Leu	Ser	Leu	Gly	Ile	Leu	Leu	Ser	Leu	Phe
		180						185					190		
Leu	Ile	Pro	Arg	Ala	Gly	Trp	Leu	Ala	Gly	Leu	Leu	Cys	Pro	Asp	Pro
	195						200					205			
Arg	Pro	Leu	Glu	Leu	Ala	Leu	Leu	Ile	Leu	Gly	Val	Gly	Leu	Leu	Asp
	210					215					220				
Phe	Cys	Gly	Gln	Val	Cys	Phe	Thr	Pro	Leu	Glu	Ala	Leu	Leu	Ser	Asp
225					230					235					240
Leu	Phe	Arg	Asp	Pro	Asp	His	Cys	Arg	Gln	Ala	Tyr	Ser	Val	Tyr	Ala
			245						250					255	
Phe	Met	Ile	Ser	Leu	Gly	Gly	Cys	Leu	Gly	Tyr	Leu	Leu	Pro	Ala	Ile
		260						265					270		
Asp	Trp	Asp	Thr	Ser	Ala	Leu	Ala	Pro	Tyr	Leu	Gly	Thr	Gln	Glu	Glu
		275					280					285			
Cys	Leu	Phe	Gly	Leu	Leu	Thr	Leu	Ile	Phe	Leu	Thr	Cys	Val	Ala	Ala
	290					295					300				
Thr	Leu	Leu	Val	Ala	Glu	Glu	Ala	Ala	Leu	Gly	Pro	Thr	Glu	Pro	Ala
305					310					315					320
Glu	Gly	Leu	Ser	Ala	Pro	Ser	Leu	Ser	Pro	His	Cys	Cys	Pro	Cys	Arg
			325						330					335	
Ala	Arg	Leu	Ala	Phe	Arg	Asn	Leu	Gly	Ala	Leu	Leu	Pro	Arg	Leu	His
		340						345					350		
Gln	Leu	Cys	Cys	Arg	Met	Pro	Arg	Thr	Leu	Arg	Arg	Leu	Phe	Val	Ala
	355						360					365			
Glu	Leu	Cys	Ser	Trp	Met	Ala	Leu	Met	Thr	Phe	Thr	Leu	Phe	Tyr	Thr
	370					375					380				
Asp	Phe	Val	Gly	Glu	Gly	Leu	Tyr	Gln	Gly	Val	Pro	Arg	Ala	Glu	Pro
385					390					395					400
Gly	Thr	Glu	Ala	Arg	His	Tyr	Asp	Glu	Gly	Val	Arg	Met	Gly	Ser	
			405					410					415		
Leu	Gly	Leu	Phe	Leu	Gln	Cys	Ala	Ile	Ser	Leu	Val	Phe	Ser	Leu	Val
		420						425				430			
Met	Asp	Arg	Leu	Val	Gln	Arg	Phe	Gly	Thr	Arg	Ala	Val	Tyr	Leu	Ala
	435						440					445			
Ser	Val	Ala	Ala	Phe	Pro	Val	Ala	Ala	Gly	Ala	Thr	Cys	Leu	Ser	His
	450					455					460				
Ser	Val	Ala	Val	Val	Thr	Ala	Ser	Ala	Ala	Leu	Thr	Gly	Phe	Thr	Phe
465					470					475					480
Ser	Ala	Leu	Gln	Ile	Leu	Pro	Tyr	Thr	Leu	Ala	Ser	Leu	Tyr	His	Arg
			485						490					495	

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

<210> 28

<211> 1959

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando a proteína híbrida entre C-LytA de St. pneum., epitopo P2 de auxiliar T e aminoácidos 51-553 de P501S humano

<400> 28

```

atggcgggccg cttacgtaca ttccgacggc tcttatccaa aagacaagtt tgagaaaatc 60
aatggcactt ggtactactt tgacagttca ggctatatgc ttgcagaccg ctggaggaag 120
cacacagacg gcaactggta ctggttcgac aactcagggc aaatggctac aggctggaag 180
aaaatcgctg ataagtggta ctatttcaac gaagaagggtg ccatgaagac aggctgggtc 240
aagtacaagg acacttggta ctacttagac gctaaagaag gcgccatgca atacatcaag 300
gctaactcta agttcattgg tatcactgaa ggcgtcatgg tatcaaatgc ctttatccag 360
tcagcggacg gaacaggctg gtactacctc aaaccagacg gaacactggc agacaggcca 420
gaaaagttca tgtacatggt gctgggcatt ggtccagtgc tgggcctggt ctgtgtcccg 480
ctcctaggct cagccagtga cactggcgt ggacgctatg gccgccgccg gcccttcac 540
tgggcactgt ccttgggcat cctgctgagc ctctttctca tcccaagggc cggctggcta 600

```

```

gcagggctgc tgtgcccgga tcccaggccc ctggagctgg cactgctcat cctgggcgtg 660
gggctgctgg acttctgtgg ccaggtgtgc ttcactccac tggaggccct gctctctgac 720
ctcttccggg acccggacca ctgtcgccag gcctactctg tctatgcctt catgatcagt 780
cttgggggct gcctgggcta cctcctgcct gccattgact gggacaccag tgccctggcc 840
ccctacctgg gcacccagga ggagtgcctc tttggcctgc tcaccctcat cttcctcacc 900
tgcgtagcag ccacactgct ggtggctgag gaggcagcgc tgggccccac cgagccagca 960
gaagggtgt cgccccctc cttgtcgccc cactgctgtc catgccgggc ccgcttggct 1020
ttccggaacc tgggcgccct gcttccccgg ctgcaccagc tgtgctgccg catgccccgc 1080
accctgcgcc ggctcttcgt ggctgagctg tgcagctgga tggcactcat gaccttcacg 1140
ctgttttaca cggatttcgt gggcgagggg ctgtaccagg gcgtgccag agctgagccg 1200
ggcaccgagg cccggagaca ctatgatgaa ggcgttcgga tgggcagcct ggggctgttc 1260
ctgcagtgcg ccctctccct ggtcttctct ctggtcatgg accggctggt gcagcgattc 1320
ggcactcgag cagtctattt ggccagtgtg gcagctttcc ctgtggctgc cggtgccaca 1380
tgctgtctcc acagtgtggc cgtggtgaca gcttcagccg ccctcaccgg gtccaccttc 1440
tcagccctgc agatcctgcc ctacacactg gcctccctct accaccggga gaagcagggtg 1500
ttcctgcccc aataccgagg ggacactgga ggtgctagca gtgaggacag cctgatgacc 1560
agcttctctg caggccctaa gcctggagct cccttcccta atggacacgt gggtgctgga 1620
ggcagtggcc tgctcccacc tccaccgcg ctctgcgggg cctctgcctg tgatgtctcc 1680
gtacgtgtgg tgggtgggta gcccaccgag gccagggtgg ttccggggccg gggcatctgc 1740
ctggacctcg ccctcctgga tagtgccttc ctgctgtccc aggtggcccc atccctgttt 1800
atgggctcca ttgtccagct cagccagtct gtcactgcct atatggtgtc tgccgcagggc 1860
ctgggtctgg tcgccattta ctttgctaca caggtagtat ttgacaagag cgacttggcc 1920
aaatactcag cgggtggaca ccataccat caccattaa 1959

```

<210> 29

<211> 507

<212> PRT

<213> Sequência Artificial

<220>

<223> P501S humano (aminoácidos 55-553) fundido a 6 resíduos de histidina

<400> 29

```

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

```

Cys	Gly	Gln	Val	Cys	Phe	Thr	Pro	Leu	Glu	Ala	Leu	Leu	Ser	Asp	Leu	
				85					90					95		
Phe	Arg	Asp	Pro	Asp	His	Cys	Arg	Gln	Ala	Tyr	Ser	Val	Tyr	Ala	Phe	
			100					105					110			
Met	Ile	Ser	Leu	Gly	Gly	Cys	Leu	Gly	Tyr	Leu	Leu	Pro	Ala	Ile	Asp	
		115					120					125				
Trp	Asp	Thr	Ser	Ala	Leu	Ala	Pro	Tyr	Leu	Gly	Thr	Gln	Glu	Glu	Cys	
	130					135					140					
Leu	Phe	Gly	Leu	Leu	Thr	Leu	Ile	Phe	Leu	Thr	Cys	Val	Ala	Ala	Thr	
145					150					155					160	
Leu	Leu	Val	Ala	Glu	Glu	Ala	Ala	Leu	Gly	Pro	Thr	Glu	Pro	Ala	Glu	
				165					170					175		
Gly	Leu	Ser	Ala	Pro	Ser	Leu	Ser	Pro	His	Cys	Cys	Pro	Cys	Arg	Ala	
			180					185					190			
Arg	Leu	Ala	Phe	Arg	Asn	Leu	Gly	Ala	Leu	Leu	Pro	Arg	Leu	His	Gln	
	195						200					205				
Leu	Cys	Cys	Arg	Met	Pro	Arg	Thr	Leu	Arg	Arg	Leu	Phe	Val	Ala	Glu	
	210					215					220					
Leu	Cys	Ser	Trp	Met	Ala	Leu	Met	Thr	Phe	Thr	Leu	Phe	Tyr	Thr	Asp	
225					230					235					240	
Phe	Val	Gly	Glu	Gly	Leu	Tyr	Gln	Gly	Val	Pro	Arg	Ala	Glu	Pro	Gly	
				245				250						255		
Thr	Glu	Ala	Arg	Arg	His	Tyr	Asp	Glu	Gly	Val	Arg	Met	Gly	Ser	Leu	
			260					265					270			
Gly	Leu	Phe	Leu	Gln	Cys	Ala	Ile	Ser	Leu	Val	Phe	Ser	Leu	Val	Met	
		275					280					285				
Asp	Arg	Leu	Val	Gln	Arg	Phe	Gly	Thr	Arg	Ala	Val	Tyr	Leu	Ala	Ser	
	290					295					300					
Val	Ala	Ala	Phe	Pro	Val	Ala	Ala	Gly	Ala	Thr	Cys	Leu	Ser	His	Ser	
305					310					315					320	
Val	Ala	Val	Val	Thr	Ala	Ser	Ala	Ala	Leu	Thr	Gly	Phe	Thr	Phe	Ser	
				325					330					335		
Ala	Leu	Gln	Ile	Leu	Pro	Tyr	Thr	Leu	Ala	Ser	Leu	Tyr	His	Arg	Glu	
		340						345					350			
Lys	Gln	Val	Phe	Leu	Pro	Lys	Tyr	Arg	Gly	Asp	Thr	Gly	Gly	Ala	Ser	
	355						360					365				
Ser	Glu	Asp	Ser	Leu	Met	Thr	Ser	Phe	Leu	Pro	Gly	Pro	Lys	Pro	Gly	
	370					375					380					
Ala	Pro	Phe	Pro	Asn	Gly	His	Val	Gly	Ala	Gly	Gly	Ser	Gly	Leu	Leu	
385					390					395					400	
Pro	Pro	Pro	Pro	Ala	Leu	Cys	Gly	Ala	Ser	Ala	Cys	Asp	Val	Ser	Val	
				405					410					415		
Arg	Val	Val	Val	Gly	Glu	Pro	Thr	Glu	Ala	Arg	Val	Val	Pro	Gly	Arg	
			420					425					430			
Gly	Ile	Cys	Leu	Asp	Leu	Ala	Ile	Leu	Asp	Ser	Ala	Phe	Leu	Leu	Ser	
		435					440					445				
Gln	Val	Ala	Pro	Ser	Leu	Phe	Met	Gly	Ser	Ile	Val	Gln	Leu	Ser	Gln	
	450					455					460					
Ser	Val	Thr	Ala	Tyr	Met	Val	Ser	Ala	Ala	Gly	Leu	Gly	Leu	Val	Ala	
465					470					475					480	
Ile	Tyr	Phe	Ala	Thr	Gln	Val	Val	Phe	Asp	Lys	Ser	Asp	Leu	Ala	Lys	
			485					490						495		
Tyr	Ser	Ala	Gly	Gly	His	His	His	His	His	His						
			500					505								

<210> 30

<211> 1524

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando P501S humano (aminoácidos 55-553) fundido a 6 resíduos de histidina

<400> 30

```
atggtgctgg gcatagggtcc agtgctgggc ctggtctgtg tcccgtcctc aggtcagcc 60
agtgaccact ggcgtggacg ctatggcgcg cgccggccct tcatctgggc actgtccttg 120
ggcaccctgc tgagcctctt tctcatccca agggccggct ggctagcagg gctgctgtgc 180
ccgatccca gggccctgga gctggcactg ctcatcctgg gcgtggggct gctggacttc 240
tgtggccagg tgtgttcac tccactggag gccctgctct ctgacctctt ccgggacctg 300
gaccactgtc gccaggccta ctctgtctat gccttcatga tcagtcttgg gggctgctg 360
ggctacctcc tgctgccaat tgactgggac accagtggc tgggcccta cctgggcacc 420
caggaggagt gcctctttgg cctgctcacc ctcatcttcc tcacctgct agcagccaca 480
ctgctggtgg ctgaggaggc agcgtgggc cccaccgagc cagcagaagg gctgtcggcc 540
ccctccttgt cgccccactg ctgtccatgc cgggcccgtg tggtttccg gaacctgggc 600
gccctgcttc cccggctgca ccagctgtgc tgccgcatgc cccgcaccct gcgccggctc 660
ttcgtggctg agctgtgcag ctggatggca ctcatgacct tcacgctgtt ttacacggat 720
ttcgtggcg aggggctgta ccagggcgtg cccagagctg agccgggcac cgaggcccgg 780
agacactatg atgaaggcgt tcggatgggc agcctggggc tgttcctgca gtgcgccatc 840
tccctggtct tctctctggt catggaccgg ctggtgcagc gattcggcac tcgagcagtc 900
tatttgccca gtgtggcagc tttccctgtg gctgccggtg ccacatgcct gtcccacagt 960
gtggccgtgg tgacagcttc agccgccctc accgggttca ccttctcagc cctgcagatc 1020
ctgccctaca cactggcctc cctctaccac cgggagaagc aggtgttctc gcccaaatac 1080
cgaggggaca ctggagggtg tagcagttag gacagcctga tgaccagctt cctgccaggc 1140
cctaagcctg gagctccctt ccctaattgga cacgtgggtg ctggaggcag tggcctgctc 1200
ccacctccac ccgcgtctg cggggcctct gcctgtgatg tctccgtacg tgtggtggtg 1260
ggtgagccca ccgaggccag ggtggttccg ggccggggca tctgcctgga cctcgccatc 1320
ctggatagtg ccttctctgt gtcccagggt gccccatccc tgtttatggg ctccattgtc 1380
cagctcagcc agtctgtcac tgctatatg gtgtctgccg caggcctggg tctggtcgcc 1440
atttactttg ctacacaggt agtatttgac aagagcgact tggccaaata ctacgcgggt 1500
ggacaccatc accatcacca ttaa 1524
```

<210> 31

<211> 685

<212> PRT

<213> Sequência Artificial

<220>

<223> P501S humano (aminoácidos 1-34 fundidos a 55-553) fundido
a 6 resíduos de histidina

<400> 31

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

Ala	Glu	Gly	Leu	Ser	Ala	Pro	Ser	Leu	Ser	Pro	His	Cys	Cys	Pro	Cys
		355					360					365			
Arg	Ala	Arg	Leu	Ala	Phe	Arg	Asn	Leu	Gly	Ala	Leu	Leu	Pro	Arg	Leu
		370				375					380				
His	Gln	Leu	Cys	Cys	Arg	Met	Pro	Arg	Thr	Leu	Arg	Arg	Leu	Phe	Val
385					390					395					400
Ala	Glu	Leu	Cys	Ser	Trp	Met	Ala	Leu	Met	Thr	Phe	Thr	Leu	Phe	Tyr
			405						410					415	
Thr	Asp	Phe	Val	Gly	Glu	Gly	Leu	Tyr	Gln	Gly	Val	Pro	Arg	Ala	Glu
			420					425					430		
Pro	Gly	Thr	Glu	Ala	Arg	Arg	His	Tyr	Asp	Glu	Gly	Val	Arg	Met	Gly
		435					440					445			
Ser	Leu	Gly	Leu	Phe	Leu	Gln	Cys	Ala	Ile	Ser	Leu	Val	Phe	Ser	Leu
		450				455						460			
Val	Met	Asp	Arg	Leu	Val	Gln	Arg	Phe	Gly	Thr	Arg	Ala	Val	Tyr	Leu
465					470					475					480
Ala	Ser	Val	Ala	Ala	Phe	Pro	Val	Ala	Ala	Gly	Ala	Thr	Cys	Leu	Ser
			485						490					495	
His	Ser	Val	Ala	Val	Val	Thr	Ala	Ser	Ala	Ala	Leu	Thr	Gly	Phe	Thr
		500						505					510		
Phe	Ser	Ala	Leu	Gln	Ile	Leu	Pro	Tyr	Thr	Leu	Ala	Ser	Leu	Tyr	His
		515					520					525			
Arg	Glu	Lys	Gln	Val	Phe	Leu	Pro	Lys	Tyr	Arg	Gly	Asp	Thr	Gly	Gly
	530					535					540				
Ala	Ser	Ser	Glu	Asp	Ser	Leu	Met	Thr	Ser	Phe	Leu	Pro	Gly	Pro	Lys
545					550					555					560
Pro	Gly	Ala	Pro	Phe	Pro	Asn	Gly	His	Val	Gly	Ala	Gly	Gly	Ser	Gly
			565						570					575	
Leu	Leu	Pro	Pro	Pro	Pro	Ala	Leu	Cys	Gly	Ala	Ser	Ala	Cys	Asp	Val
		580						585					590		
Ser	Val	Arg	Val	Val	Val	Gly	Glu	Pro	Thr	Glu	Ala	Arg	Val	Val	Pro
		595					600					605			
Gly	Arg	Gly	Ile	Cys	Leu	Asp	Leu	Ala	Ile	Leu	Asp	Ser	Ala	Phe	Leu
	610					615					620				
Leu	Ser	Gln	Val	Ala	Pro	Ser	Leu	Phe	Met	Gly	Ser	Ile	Val	Gln	Leu
625					630					635					640
Ser	Gln	Ser	Val	Thr	Ala	Tyr	Met	Val	Ser	Ala	Ala	Gly	Leu	Gly	Leu
			645						650					655	
Val	Ala	Ile	Tyr	Phe	Ala	Thr	Gln	Val	Val	Phe	Asp	Lys	Ser	Asp	Leu
		660						665					670		
Ala	Lys	Tyr	Ser	Ala	Gly	Gly	His	His	His	His	His	His			
		675					680					685			

<210> 32

<211> 2058

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando P501S humano (aminoácidos 1-34 fundidos a 55-553) fundido a 6 resíduos de histidina

<400> 32

```
atggcgggccg tgcagaggct atgggtatcg agactgctaa gacaccgcaa agctcagttg 60
ttgttggtta acttggtgac cttcgggctg gaagtctgtt tggcgggccgc ttacgtacat 120
tccgacgggct cttatccaaa agacaagttt gagaaaatca atggcacttg gtactacttt 180
gacagttcag gctatatgct tgcagaccgc tggaggaagc acacagacgg caactggtac 240
tggttcgaca actcaggcga aatggctaca ggctggaaga aaatcgctga taagtggtag 300
tatttcaacg aagaagggtc catgaagaca ggctgggtca agtacaagga cacttggtag 360
tacttagacg cttaaagaagg cgccatgcaa tacatcaagg ctaactctaa gttcattggt 420
atcactgaag gcgtcatggt atcaaatgcc tttatccagt cagcggacgg aacaggctgg 480
tactacctca aaccagacgg aacactggca gacaggccag aaaagttcat gtacatggtg 540
ctgggcattg gtccagtgtc gggcctggtc tgtgtcccgc tcctaggctc agccagtgc 600
cactggcggtg gacgttatgg ccgccgccgg cccttcctct gggcactgtc cttgggcac 660
ctgttgagcc tctttctcat cccaaggggc ggctggctag cagggtgtgt gtgcccggat 720
cccaggcccc tggagctggc actgctcatc ctgggcgtgg ggctgctgga cttctgtggc 780
caggtgtgtc tccactccact ggaggccctg ctctctgacc tcttccggga cccggaccac 840
tgtcgccagg cctactctgt ctatgccttc atgatcagtc ttgggggctg cctgggctac 900
ctcctgcctg ccattgactg ggacaccagt gccctggccc cctacctggg caccagggag 960
gagtgcctct ttggcctgtc caccctcatc ttctcacct gcgtagcagc cacactgctg 1020
gtggctgagg aggcagcgtc gggccccacc gagccagcag aagggtgtgc ggccccctcc 1080
ttgtcgcccc actgtgtgct atgccggggc cgcttggtct tccggaacct gggcgccctg 1140
cttccccggc tgcaccagct gtgtgtgcgc atgccccgca ccctgcgcgc gctcttcgtg 1200
gctgagctgt gcagctggat ggcactcatg accttcacgc tgttttacac ggatttcgtg 1260
ggcgagggggc tgtaccaggc cgtgcccaga gctgagccgg gcaccgaggc ccggagacac 1320
tatgatgaag gcgttcggat gggcagcctg gggctgttcc tgcagtgcgc catctccctg 1380
gtcttctctc tggctcatgga ccggctggtg cagcgattcg gactcagagc agtctatttg 1440
gccagtgtgg cagctttccc tgtggctgcc ggtgccacat gcctgtccca cagtgtggcc 1500
gtggtgacag cttcagccgc cctcaccggg ttcacctctc cagccctgca gatcctgcc 1560
tacacactgg cctccctcta ccaccgggag aagcagggtg tcctgccccaa ataccgaggg 1620
gacactggag gtgctagcag tgaggacagc ctgatgacca gcttcctgcc aggccctaag 1680
cctggagctc ccttccctaa tggacaagtg ggtgctggag gcagtggcct gctcccacct 1740
ccaccgcgc tctgcggggc ctctgcctgt gatgtctccg tacgtgtggt ggtgggtgag 1800
cccaccgagg ccagggtggt tccgggccgg ggcattctgc tggacctcgc catcctggat 1860
agtgccttcc tgctgtccca ggtggcccca tccctgttta tgggtccat tgtccagctc 1920
agccagtctg tcaactgccta tatggtgtct gccgcaggcc tgggtctggt cgccatttac 1980
tttgctacac aggtagtatt tgacaagagc gacttggcca aatactcagc ggggtggacac 2040
catcaccatc accattaa 2058
```

<210> 33

<211> 671

<212> PRT

<213> Sequência Artificial

<220>

<223> Porção C-LytA de St. pneum. fundida ao epitopo P2 de auxiliar T fundido a P501S humano (aminoácidos 55-553) fundido a 6 resíduos de histidina a jusante de sequência sinal alfa pré-pró de levedura

<400> 33

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

				325					330				335				
Glu	Pro	Ala	Glu	Gly	Leu	Ser	Ala	Pro	Ser	Leu	Ser	Pro	His	Cys	Cys		
			340					345					350				
Pro	Cys	Arg	Ala	Arg	Leu	Ala	Phe	Arg	Asn	Leu	Gly	Ala	Leu	Leu	Pro		
		355					360					365					
Arg	Leu	His	Gln	Leu	Cys	Cys	Arg	Met	Pro	Arg	Thr	Leu	Arg	Arg	Leu		
	370					375					380						
Phe	Val	Ala	Glu	Leu	Cys	Ser	Trp	Met	Ala	Leu	Met	Thr	Phe	Thr	Leu		
385					390					395					400		
Phe	Tyr	Thr	Asp	Phe	Val	Gly	Glu	Gly	Leu	Tyr	Gln	Gly	Val	Pro	Arg		
			405					410						415			
Ala	Glu	Pro	Gly	Thr	Glu	Ala	Arg	Arg	His	Tyr	Asp	Glu	Gly	Val	Arg		
		420					425					430					
Met	Gly	Ser	Leu	Gly	Leu	Phe	Leu	Gln	Cys	Ala	Ile	Ser	Leu	Val	Phe		
	435					440					445						
Ser	Leu	Val	Met	Asp	Arg	Leu	Val	Gln	Arg	Phe	Gly	Thr	Arg	Ala	Val		
	450				455						460						
Tyr	Leu	Ala	Ser	Val	Ala	Ala	Phe	Pro	Val	Ala	Ala	Gly	Ala	Thr	Cys		
465				470					475					480			
Leu	Ser	His	Ser	Val	Ala	Val	Val	Thr	Ala	Ser	Ala	Ala	Leu	Thr	Gly		
			485					490					495				
Phe	Thr	Phe	Ser	Ala	Leu	Gln	Ile	Leu	Pro	Tyr	Thr	Leu	Ala	Ser	Leu		
		500					505					510					
Tyr	His	Arg	Glu	Lys	Gln	Val	Phe	Leu	Pro	Lys	Tyr	Arg	Gly	Asp	Thr		
	515					520					525						
Gly	Gly	Ala	Ser	Ser	Glu	Asp	Ser	Leu	Met	Thr	Ser	Phe	Leu	Pro	Gly		
	530				535						540						
Pro	Lys	Pro	Gly	Ala	Pro	Phe	Pro	Asn	Gly	His	Val	Gly	Ala	Gly	Gly		
545				550				555						560			
Ser	Gly	Leu	Leu	Pro	Pro	Pro	Pro	Ala	Leu	Cys	Gly	Ala	Ser	Ala	Cys		
		565						570				575					
Asp	Val	Ser	Val	Arg	Val	Val	Val	Gly	Glu	Pro	Thr	Glu	Ala	Arg	Val		
		580					585					590					
Val	Pro	Gly	Arg	Gly	Ile	Cys	Leu	Asp	Leu	Ala	Ile	Leu	Asp	Ser	Ala		
	595					600						605					
Phe	Leu	Leu	Ser	Gln	Val	Ala	Pro	Ser	Leu	Phe	Met	Gly	Ser	Ile	Val		
	610				615						620						
Gln	Leu	Ser	Gln	Ser	Val	Thr	Ala	Tyr	Met	Val	Ser	Ala	Ala	Gly	Leu		
625					630				635					640			
Gly	Leu	Val	Ala	Ile	Tyr	Phe	Ala	Thr	Gln	Val	Val	Phe	Asp	Lys	Ser		
		645					650					655					
Asp	Leu	Ala	Lys	Tyr	Ser	Ala	Gly	Gly	His	His	His	His	His	His			
		660					665					670					

<210> 34

<211> 2477

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando a porção C-LyTA de St. pneum. fundido a epitopo P2 de auxiliar T fundido a P501S humano (aminoácidos 55-553) fundido a 6 resíduos de histidina a jusante de sequência sinal alfa pré-pró de levedura

<400> 34

```
tacgtacatt ccgacggctc ttatccaaaa gacaagtttg agaaaatcaa tggcacttgg 60
tactactttg acagttcagg ctatatgctt gcagaccgct ggaggaagca cacagacggc 120
aactgggtact ggttcgacaa ctcaggcgaa atgggtacag gctggaagaa aatcgctgat 180
aagtgggtact atttcaacga agaaggtgcc atgaagacag gctgggtcaa gtacaaggac 240
acttgggtact acttagacgc taaagaaggc gccatgcaat acatcaaggc taactctaag 300
ttcattggta tcactgaagg cgtcattgta tcaaatgcct ttatccagtc agcggacgga 360
acaggetggg actacctcaa accagacgga acactggcag acaggccaga aatggcggcc 420
agatttcctt caatttttac tgcagtttta ttgcagcat cctccgcatt agcggccgct 480
tacgtacatt ccgacggctc ttatccaaaa gacaagtttg agaaaatcaa tggcacttgg 540
tactactttg acagttcagg ctatatgctt gcagaccgct ggaggaagca cacagacggc 600
aactgggtact ggttcgacaa ctcaggcgaa atgggtacag gctggaagaa aatcgctgat 660
aagtgggtact atttcaacga agaaggtgcc atgaagacag gctgggtcaa gtacaaggac 720
acttgggtact acttagacgc taaagaaggc gccatgcaat acatcaaggc taactctaag 780
ttcattggta tcactgaagg cgtcattgta tcaaatgcct ttatccagtc agcggacgga 840
acaggetggg actacctcaa accagacgga acactggcag acaggccaga agctgggtatt 900
acttacgttc caccattggt gttggaagtt ggtgttgaag aaaagtcat gtacatggtg 960
ctgggcattg gtccagtgtt gggcctggtc tgtgtccgc tcctaggctc agccagtgtg 1020
cactggcggt gacgtatgtt ccgccgcgg cccttcatct gggcactgtc cttgggcatt 1080
ctgctgagcc tctttctcat cccaagggcc ggctggctag cagggtgtgt gtgcccggat 1140
cccaggcccc tggagctggc actgtctatc ctgggcgttg ggctgtgtga cttctgtggc 1200
caggtgtgtt tcaactccat ggagggcctg ctctctgacc tcttccggga cccggaccac 1260
tgtcgccagg cctactctgt ctatgcttca tgatcagtct tgggggtgtc ctgggtacc 1320
tctgctgtgc cattgactgg gacaccagt cctggcccc ctacctgggc acccaggagg 1380
agtgcctctt tggcctgtc accctcatct tctcacctg cgtagcagcc acactgctgt 1440
tggctgagga ggcagcgtg ggcgccacc agccagcaga agggctgtc gccccctct 1500
tgtgccccca ctgctgtcca tgccggggcc gcttggcttt ccggaacctg ggcgccccct 1560
ttccccggct gcaccagtgt tgctgccgca tgccccgcac cctgcgccgg ctcttcgtgg 1620
ctgagctgtg cagctggatg gcaactcatg ccttcacgct gttttacacg gatttcgtgg 1680
gagagggggt gtaccagggc gtgcccagag ctgagccggg caccgaggcc cggagacact 1740
atgatgaagg cgttcggatg ggcagcctgg ggctgttct gcagtgcgcc atctccctgg 1800
tcttctctct ggtcatggac cggtcgtgtc agcgattcgg cactcgagca gtctatttgg 1860
ccagtgtggc agctttccct gtggctgccg gtgccacatg cctgtccac agtgtggccg 1920
tggtgacagc ttcagccgcc ctcaccgggt tcacctctc agccctgcag atcctgccct 1980
acacactggc ctccctctac caccgggaga agcaggtgtt cctgccccaa taccgagggg 2040
acactggagg tgctagcagt gaggacagcc tgatgaccag cttcctgcca ggccctaagc 2100
ctggagctcc cttccctaag ggacacgtgg gtgctggagg cagtggcctg ctcccacctc 2160
caccgcgct ctgccccggc tctgcctgtg atgtctccgt acgtgtggtg gtgggtgagc 2220
ccaccgaggc cagggtggtt ccgggcccgg gcactctgct ggacctgcgc atcctggata 2280
gtgcttctct gctgtcccag gtggccccat cctgtttat gggctccatt gtccagctca 2340
gccagtctgt cactgcctat atggtgtctg ccgaggccct gggctctggt gccatttact 2400
ttgctacaca ggtagtattt gacaagagcg acttggccaa atactcagcg ggtggacacc 2460
atccatca ccatataa 2477
```

<210> 35

<211> 595

<212> PRT

<213> Sequência Artificial

<220>

<223> P501S humano (aminoácidos 55-553) fundido a 6 resíduos de histidina a jusante de sequência sinal alfa pré-pró de levedura

<400> 35

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

Ala	Leu	Leu	Pro	Arg	Leu	His	Gln	Leu	Cys	Cys	Arg	Met	Pro	Arg	Thr
290					295						300				
Leu	Arg	Arg	Leu	Phe	Val	Ala	Glu	Leu	Cys	Ser	Trp	Met	Ala	Leu	Met
305				310					315						320
Thr	Phe	Thr	Leu	Phe	Tyr	Thr	Asp	Phe	Val	Gly	Glu	Gly	Leu	Tyr	Gln
			325					330					335		
Gly	Val	Pro	Arg	Ala	Glu	Pro	Gly	Thr	Glu	Ala	Arg	Arg	His	Tyr	Asp
		340					345						350		
Glu	Gly	Val	Arg	Met	Gly	Ser	Leu	Gly	Leu	Phe	Leu	Gln	Cys	Ala	Ile
	355					360					365				
Ser	Leu	Val	Phe	Ser	Leu	Val	Met	Asp	Arg	Leu	Val	Gln	Arg	Phe	Gly
370					375					380					
Thr	Arg	Ala	Val	Tyr	Leu	Ala	Ser	Val	Ala	Ala	Phe	Pro	Val	Ala	Ala
385				390					395						400
Gly	Ala	Thr	Cys	Leu	Ser	His	Ser	Val	Ala	Val	Val	Thr	Ala	Ser	Ala
			405					410					415		
Ala	Leu	Thr	Gly	Phe	Thr	Phe	Ser	Ala	Leu	Gln	Ile	Leu	Pro	Tyr	Thr
	420						425					430			
Leu	Ala	Ser	Leu	Tyr	His	Arg	Glu	Lys	Gln	Val	Phe	Leu	Pro	Lys	Tyr
	435					440					445				
Arg	Gly	Asp	Thr	Gly	Gly	Ala	Ser	Ser	Glu	Asp	Ser	Leu	Met	Thr	Ser
450					455					460					
Phe	Leu	Pro	Gly	Pro	Lys	Pro	Gly	Ala	Pro	Phe	Pro	Asn	Gly	His	Val
465				470					475						480
Gly	Ala	Gly	Gly	Ser	Gly	Leu	Leu	Pro	Pro	Pro	Pro	Ala	Leu	Cys	Gly
			485					490					495		
Ala	Ser	Ala	Cys	Asp	Val	Ser	Val	Arg	Val	Val	Val	Gly	Glu	Pro	Thr
	500						505					510			
Glu	Ala	Arg	Val	Val	Pro	Gly	Arg	Gly	Ile	Cys	Leu	Asp	Leu	Ala	Ile
	515					520					525				
Leu	Asp	Ser	Ala	Phe	Leu	Leu	Ser	Gln	Val	Ala	Pro	Ser	Leu	Phe	Met
530					535					540					
Gly	Ser	Ile	Val	Gln	Leu	Ser	Gln	Ser	Val	Thr	Ala	Tyr	Met	Val	Ser
545				550					555					560	
Ala	Ala	Gly	Leu	Gly	Leu	Val	Ala	Ile	Tyr	Phe	Ala	Thr	Gln	Val	Val
			565				570						575		
Phe	Asp	Lys	Ser	Asp	Leu	Ala	Lys	Tyr	Ser	Ala	Gly	Gly	His	His	His
		580					585						590		
His	His	His													
	595														

<210> 36

<211> 1788

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando P501S humano (aminoácidos 55-553) fundido a 6 resíduos de histidina a jusante de sequência sinal alfa pré-pró de levedura

<400> 36

```
atgagtttcc tcaattttac tgcagtttta ttccgcagcat cctccgcatt agctgctcca 60
gtcaacacta caacagaaga tgaaacggca caaattccgg ctgaagctgt catcggttac 120
tcagatttag aaggggattt cgatgttgct gttttgccat ttccaacag cacaataac 180
gggttattgt ttataaatac tactattgcc agcattgctg ctaaaagaaga aggggtatct 240
ctcgagaaaa gagaggctga agccatgggtg ctgggcattg gtccagtgtg gggcctggtc 300
tgtgtcccgc tcctaggctc agccagtga cactggcgtg gacgctatgg cgcgcgcgg 360
cccttcattc gggcactgtc cttgggcata ctgctgagcc tctttctcat cccaagggcc 420
ggctggctag cagggtgtgt gtgcccggat ccaggcccc tggagctggc actgctcatc 480
ctgggcgtgg ggctgctgga cttctgtggc cagggtgtgt tccactccact ggaggccctg 540
ctctctgacc tcttccggga cccggaccac tgcgccagg cctactctgt ctatgccttc 600
atgatcagtc ttgggggctg cctgggctac ctccctgctg ccattgactg ggacaccagt 660
gccctggccc cctacctggg caccagagg gagtgccctc ttggcctgct caccctcatc 720
ttcctcacct gcgtagcagc cacactgtgt gtggctgagg aggcagcgtg gggccccacc 780
gagccagcag aagggtgtgt gggcccctcc ttgtcgcccc actgctgtcc atgccgggcc 840
cgcttggtct tccggaacct gggcgccctg cttccccggc tgcaccagct gtgctgccgc 900
atgccccgca ccctgcgccg gctcttcgtg gctgagctgt gcagctggat ggcactcatg 960
accttcacgc tgttttacac ggatttcgtg ggcgaggggc tgtaccaggg cgtgcccca 1020
gctgagccgg gcaccgaggc ccggagacac tatgatgaag gcgttcggat gggcagcctg 1080
gggctgttcc tgcagtgcgc catctccctg gtcttctctc tggcatgga cgggctgggt 1140
cagcgattcg gcactcgagc agtctatttg gccagtgtgg cagctttccc tgtggctgcc 1200
ggtgccaat gcctgtccca cagtgtggcc gtggtgacag cttcagccgc cctcaccggg 1260
ttcaccttct cagccctgca gacccgtccc tacacactgg cctccctcta ccaccgggag 1320
aagcaggtgt tccgtcccaa ataccgaggg gacactggag gtgctagcag tgaggacagc 1380
ctgatgacca gcttcctgcc aggccctaag cctggagctc cttccctaa tggacacgtg 1440
gggtgtggag gcagtggcct gctcccacct ccaccgcgc tctgcggggc ctctgcctgt 1500
gatgtctccg tacgtgtggg ggtgggtgag ccaccgagg ccagggtggg tccgggccgg 1560
ggcatctgcc tggacctcgc catcctggat agtgcccttc tgctgtcca ggtggcccca 1620
tcctgttta tgggctccat tgtccagctc agccagtctg tccactgccta tatggtgtct 1680
gccgcaggcc tgggtctggg cgcattttac tttgctacac aggtagtatt tgacaagagc 1740
gacttggcca aatactcagc ggggtggacac catcaccatc accattaa 1788
```

<210> 37

<211> 1955

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando P501S humano de códon otimizado
(aminoácidos 51-553) fundido a C-LytA de St.pneum. epitopo P2
auxiliar C-Lyta

<400> 37

```
gcggccgcgc caccatggcc gccgcctacg tgcatagcga cgggagctac cccaaggaca 60
agttcgagaa gatcaacggg acatgggtact acttcgactc ctccggctac atgctcgccg 120
accgctggcg gaagcacacc gacggcaact ggtactggtt cgataactcg ggagagatgg 180
ccaccggctg gaagaagatc gcggacaagt ggtactatct caacgaggag ggcgccatga 240
agaccggctg ggtgaagtat aaggacacct ggtactacct cgacgccaaag gagggcgcca 300
tgcagtatat caaggccaac agcaagtcca tcggcatcac cgagggagtg atggtcagca 360
acgcctttat ccagagcgcc gacggcaccg gatggtacta cttgaagccg gacggcacc 420
tcgcgatcg gcccgagaag ttcattgtaca tgggtgctggg catcgcccc gtccctggggc 480
tcgtgtgtgt cccctcctc gggagtgcgt ccgactctg gcggggccgc tacggccgc 540
gcagaccgtt catctgggccc ctgagcctgg gcactctgt ctctctcttc ctgatecccc 600
gggcccggctg gctggccggc ctgctgtgtc ccgaccccc cctctggag ctggccctcc 660
tgatcctggg cgtgggcctg ctggacttct gcggccaggt gtgtttcact cccctggagg 720
ctctgtcttc cgacctcttc cgcgaccccc accactgtag gcaggcttac agcgtgtacg 780
ccttcatgat cagtctgggg ggatgcctgg gctatctgt gcccgctatc gactgggaca 840
ccagcgccct ggccccctac ctggggactc aggaggagt cctgttcggc ctgctcacct 900
tgatcttctt gacgtgcgtc gccgccaccc tgcctggggc cgaggaggcg gccctggggc 960
ccaccgagcc cgcgcagggc ctgagcgctc ccagcctgag ccccatctgc tgcccggtga 1020
gggctaggct cgccttcagg aatctggggc ctttgctgcc ccgcctgcat cagctgtgct 1080
gtcgcatgcc tcgcacctg cgcgcctgt tcgtcgctga gctctgttc tggtatggccc 1140
tgatgacgtt caccctcttc tacaccgact tcgtggggga gggcctgtac cagggcggtc 1200
ccagggccga gcccggcacc gaggctaggc gccattacga cgagggcgtc aggatgggct 1260
ctctgggcct ctccctgcag tgcgccatca gtctggtgt ctctctggtg atggaccggc 1320
tgggtgcagc ctccggcacc cgggcccgtg acctcgctc tgtggcggct tccccgctc 1380
ccgcggcgcc gacctgctg tctcattctg tcgcctggt gaccgccagc gccgcctga 1440
ccggtctcac ctccagtgct ctccagattc tgccctacac cctggcgtct ctgtaccatc 1500
gcgagaagca ggtgttctc cccaagtacc gcggggacac agggggagct tcctctgagg 1560
acagcctgat gaccagcttc ttgcccgccc ccaagccggg ggcccccttc cccaacggcc 1620
atgtcggggc gggcggcagc ggcctgctcc ctccccccc cgcctgtgc ggcgctagt 1680
cctgcgacgt gacggtgcgg gtggtggtg gggagccac cgaggctagg gtcgtgctg 1740
gccgggggat ctgcctggac ctggccatcc tcgactccgc cttcctgctc tcccaggtg 1800
cgcccagcct gttcatgggc agtatcgtgc agctgagcca gagcgtgacc gcctacatgg 1860
tgagcgccgc cggcctgggg ttggtggcca tctactttgc caccaggtc gtgttcgaca 1920
agagcgatct cgccaagtat agcgcctgag gatcc 1955
```

<210> 38

<211> 2045

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando P501S humano de códon otimizado
(aminoácidos 1-553) fundido a C-LyTA de St. pneum. epitopo P2
auxiliar C-Lyta

<400> 38

```
gcggccgcgc caccatggcc gccgcctacg tgcatagcga cgggagctac cccaaggaca 60
agttcgagaa gatcaacggg acatgggtact acttcgactc ctccggctac atgctcgccg 120
accgctggcg gaagcacacc gacggcaact ggtactggtt cgataactcg ggagagatgg 180
ccaccggctg gaagaagatc gccgacaagt ggtactattt caacgaggag ggcgccatga 240
agaccggctg ggtgaagtat aaggacacct ggtactacct cgacgccaag gagggcgcca 300
tgcagtatat caaggccaac agcaagttca tcggcatcac cgagggagtg atggtcagca 360
acgcctttat ccagagcgcc gacggcaccc gatggtacta cttgaagccg gacggcaccc 420
tcgcggatcg gcccagagatg gtgcagcggc tgtgggtgtc ccggtgtgtg cgccatagaa 480
aggcccagtt gctgctggtg aacctgctga ctttcggact ggaggtgtgc ctggctgccg 540
tgggtgctggg catcgcccc gtccctgggccc tcgtgtgtgt gcccctcttc gggagtgcgt 600
ccgatcattg gcggggccgc tacggccgcc gcagaccgtt catctgggcc ctgagcctgg 660
gcatectgct ctctctcttc ctgaccccc gggccggctg gctggccggc ctgctgtgtc 720
ccgacccccg cctcttgagg ctggccctcc tgatcctggg cgtgggacct ctggacttct 780
gcggccaggt gtgtttcact cccctggagg ctctgtcttc cgacctcttc cgcgaccccc 840
accactgtag gcaggcttac agcgtgtacg ccttcagatg cagtctgggg ggatgcctgg 900
gctatctgct gcccgtatc gactgggaca ccagcgccct ggccccctac ctggggactc 960
aggaggagtg cctggttcggc ctgctcacct tgactcttcc gacgtgcgtc gccggcacc 1020
tgctggtggc cgaggaggcg gccctggggc ccaccgagcc cgccgagggc ctgagcgctc 1080
ccagcctgag ccccatgtgc tgcccgtgca gggctaggct cgccctcagg aatctgggcg 1140
ctttgctgcc ccgcctgcat cagctgtgct gtcgcatgcc tcgcaccctg cgccgcctgt 1200
tcgtcgctga gctctgttcc tggatggccc tgatgacgtt caccctcttc tacaccgact 1260
tcgtggggga gggcctgtac cagggcggtc ccagggccga gcccggcacc gaggctaggc 1320
gccattacga cgaggcgctc aggatgggct ctctgggcct ctctctgcag tgcgccatca 1380
gtctggtgtt ctctctgggt atggaccggc tgggtgcagc ctctggcacc cgggccgtgt 1440
acctcgctc tgtggcggtt ttccccgtcg ccgcccggcg gacctgcctg tctcattctg 1500
tcgcccgtgt gaccgccagc gccgccctga ccggcttcac cttcagtgcg ctccagattc 1560
tgccctacac cctggcgctc ctgtaccatc gcgagaagca ggtgttcctg cccaagtacc 1620
gcggggacac agggggagct tcctctgagg acagcctgat gaccagcttc ttgcccggcc 1680
ccaagccggg ggcccccttc cccaacggcc atgtcggggc gggcggcagc ggcctgctcc 1740
ctcccccccc cgccctgtgc ggcgtagtgc cctgcgacgt gagcgtgcgg gtggtggtgg 1800
gggagccccc cgaggctagg gtcgtgcctg gccgggggat ctgcctggac ctggccatcc 1860
tcgactccgc cttcctgtct tcccaggtgg cgcccagcct gttcatgggc agtatcgtgc 1920
agctgagcca gagcgtgacc gcctacatgg tgagcgccgc cggcctgggg ttggtggcca 1980
tctactttgc caccaggtc gtgttcgaca agagcgatct cgccaagtat agcgctgag 2040
gatcc 2045
```

<210> 39

<211> 2105

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando C-LytA de St. pneum. epitopo P2 auxiliar
C-Lyta fundido a P501S humano (aminoácidos 51-553) fundido a
P501S humano (aminoácidos 1-50) - codão otimizado

<400> 39

```
gcgggcgcgc caccatggcc gccgcctacg tgcatagcga cgggagctac cccaaggaca 60
agttcgagaa gatcaacggg acatgggtact acttcgactc ctccggctac atgctcgccg 120
accgctggcg gaagcacacc gacggcaact ggtactgggt cgataactcg ggagagatgg 180
ccaccggctg gaagaagatc gcggacaagt ggtactattt caacgaggag ggcgccatga 240
agaccggctg ggtgaagtat aaggacacct ggtactacct cgacgccaaag gagggcgcca 300
tgcagtatat caaggccaac agcaagttca tcggcatcac cgagggagtg atggtcagca 360
acgcctttat ccagagcgcc gacggcaccg gatggtacta cttgaagccg gacggcacc 420
tcgcggatcg gcccagagaag ttcatgtaca tgggtgctgg catcggccccc gtcctgggcc 480
tcgtgtgtgt gcccctctc gggagtgcgt ccgatcattg gcggggccgc tacggccgcc 540
gcagaccgtt catctgggcc ctgagcctgg gcacccctgt ctctctcttc ctgatcccc 600
gggcccggctg gctggccggc ctgctgtgtc ccgacccccg cctctgggag ctggccctcc 660
tgatcctggg cgtgggcctg ctggacttct gcggccagggt gtgtttcact cccctggagg 720
ctctgtctct cgacctcttc cgcgaccccc accactgtag gcaggcttac agcgtgtacg 780
ccttcatgat gactctgggg ggatgcctgg gctatctgct gcccgctatc gactgggaca 840
ccagcgccct gggcccctac ctggggactc aggaggagtg cctgttcggc ctgctcacct 900
tgatcttctt gacgtgcgtc gccgccaccc tgctggtggc cgaggaggcg gccctggggc 960
ccaccgagcc cgccgagggc ctgagcgctc ccagcctgag ccccatctgc tgcccggtga 1020
gggctaggct cgcccttcagg aatctgggcg ctttgcctgc ccgcctgcat cagctgtgct 1080
gtcgcatgcc tcgcacctg cgccgcctgt tcgtcgctga gctctgttcc tggatggccc 1140
tgatgacgtt caccctcttc tacaccgact tcgtggggga gggcctgtac cagggcgtgc 1200
ccaggggcca gcccggcacc gaggttaggc gccattacga cgagggcgtc aggatgggct 1260
ctctgggcct ctctctgcag tgcgccatca gtctggtgtt ctctctggtg atggaccggc 1320
tggtgcagcg cttcggcacc cgggcctgtt acctgcctc tgtggcggtt tccccgtcg 1380
ccgcccgcgc gacctgcctg tctcattctg tcgccgtggg gaccgccagc gccgccctga 1440
ccggttcac cttcagtgcg ctccagattc tgccctacac cctggcgtct ctgtaccatc 1500
gcgagaagca ggtgttctct cccaagtacc gcggggacac agggggagct tcctctgagg 1560
acagcctgat gaccagcttc ttgcccggcc ccaagccggg gggcccttcc cccaacggcc 1620
atgtcggggc gggcggcagc ggcctgctcc ctccccccc cgccctgtgc ggcctagtgc 1680
cctgcgacgt gacgtgcgg gtggtggtgg gggagccac cgaggctagg gtcgtgcctg 1740
gccgggggat ctgcctggac ctggccatcc tcgactccgc ctctctgctc tcccaggtgg 1800
cgccagcct gttcatgggc agtatcgtgc agctgagcca gagcgtgacc gcctacatgg 1860
tgagcgccgc cggcctgggg ttggtggcca tctactttgc caccaggtc gtgttcgaca 1920
agagcgatct cgccaagtat agcgccatgg tgcagcggct gtgggtgtcc cggtgctgc 1980
gccatagaaa ggcccagttg ctgctggtga acctgctgac tttcggactg gaggtgtgcc 2040
tggtgcccgg gatcacgtac gtgccccccc tgctgctgga ggtgggcgtg gaggagtgg 2100
gatcc 2105
```

<210> 40

<211> 2105

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando P501S humano (aminoácidos 1-50) fundido a C-LytA de St. pneum. epitopo P2 auxiliar C-Lyta fundido a P501S humano (aminoácidos 51-553) - codão otimizado

<400> 40

```
gcggccgcgc caccatgggtg cagcggctgt ggggtgtcccg gctgctgcgc catagaaagg 60
cccagttgct gctggtgaac ctgctgactt tgggactgga ggtgtgcctg gctgccggga 120
tcacgtacgt gccccccctg ctgctggagg tgggcgtgga ggagatggcc gccgcctacg 180
tgcatacgca cgggagctac cccaaggaca agttcgagaa gatcaacggg acatgggtact 240
acttcgactc ctccggctac atgctcgccg accgctggcg gaagcacacc gacggcaact 300
ggtactgggt cgataactcg ggagagatgg ccaccggctg gaagaagatc gcggacaagt 360
ggtactatct caacgaggag ggcgccatga agaccggctg ggtgaagtat aaggacacct 420
ggtactacct cgacgccaag gagggcgcca tgcagtatat caaggccaac agcaagttca 480
tcggcatcac cgagggagtg atggtcagca acgcctttat ccagagcgcc gacggcaccg 540
gatggtacta cttgaagccg gacggcaccg tcgcggtatc gcccgagaag ttcattgtaca 600
tggtgctggg catcgccccc gtctctggcc tcgtgtgtgt gcccctcttc gggagtgcgt 660
ccgatcattg gcggggccgc tacggccgcc gcagaccgtt catctgggcc ctgagcctgg 720
gcatcctgct ctctctcttc ctgatcccc gggccggctg gctggccggc ctgctgtgtc 780
ccgacccccg cctcttgag ctggccctcc tgatcctggg cgtgggcctg ctggacttct 840
cgggccaggt gtgttttact cccctggagg ctctgctctc cgacctcttc cgcgaccccc 900
accactgtag gcaggcttac agcgtgtacg ccttcatgat cagtctgggg ggatgcctgg 960
gctatctgct gccgctatc gactgggaca ccagcgccct ggccccctac ctggggactc 1020
aggaggagtg cctgttcggc ctgctcacct tgatcttctt gacgtgcgtc gccgccaccg 1080
tgctggtggc cgaggaggcg gccctggggc ccaccgagcc cgccgagggc ctgagcgctc 1140
ccagcctgag ccccatctgc tgcccggtga gggctaggct cgccctcagg aatctggggc 1200
ctttgctgcc ccgcctgcat cagctgtgct gtcgcatgcc tcgcacctg cgcgcctgt 1260
tcgtcgctga gctctgttcc tggatggccc tgatgacgtt caccctcttc tacaccgact 1320
tcgtggggga gggcctgtac cagggcgtgc ccaggggcga gccgggcacc gaggctaggc 1380
gccattacga cgagggcgtc aggatgggct ctctgggcct ctctctgcag tgcgccatca 1440
gtctggtgtt ctctctggtg atggaccggc tgggtgcagc cttcggcacc cgggccgtgt 1500
acctcgccctc tgtggcggtt ttcccctcgc ccgcggcgcc gacctgcctg tctcattctg 1560
tcgccgtggt gaccgccagc gccgccctga ccggttcac cttcagtgcg ctccagattc 1620
tgccctacac cctggcgtct ctgtaccatc gcgagaagca ggtgttctct cccaagtacc 1680
gcggggacac agggggagct tcctctgagg acagcctgat gaccagcttc ttgcccggcc 1740
ccaagccggg ggcccccttc cccaacggcc atgtcggggc gggcggcagc ggcctgctcc 1800
ctcccccccc cgccctgtgc ggcgctagt cctgcgacgt gacgtgcgg gtggtggtg 1860
gggagccac cgaggctagg gtcgtgcctg gccgggggat ctgctggac ctggccatcc 1920
tcgactccgc cttctgtctc tcccagggtg cggccagcct gttcatgggc agtatcgtgc 1980
agctgagcca gagcgtgacc gcctacatgg tgagcgccgc cggcctgggg ttggtggcca 2040
tctactttgc caccaggtc gtgttcgaca agagcgatct cgccaagtat agcgccctgag 2100
gatcc 2105
```

<210> 41

<211> 652

<212> PRT

<213> Sequência Artificial

<220>

<223> C-LytA de St. pneum. epitopo P2 auxiliar C-Lyta fundido a P501S humano

<400> 41

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

Cys	Leu	Phe	Gly	Leu	Leu	Thr	Leu	Ile	Phe	Leu	Thr	Cys	Val	Ala	Ala	290	295	300
Thr	Leu	Leu	Val	Ala	Glu	Glu	Ala	Ala	Leu	Gly	Pro	Thr	Glu	Pro	Ala	305	310	315
Glu	Gly	Leu	Ser	Ala	Pro	Ser	Leu	Ser	Pro	His	Cys	Cys	Pro	Cys	Arg	325	330	335
Ala	Arg	Leu	Ala	Phe	Arg	Asn	Leu	Gly	Ala	Leu	Leu	Pro	Arg	Leu	His	340	345	350
Gln	Leu	Cys	Cys	Arg	Met	Pro	Arg	Thr	Leu	Arg	Arg	Leu	Phe	Val	Ala	355	360	365
Glu	Leu	Cys	Ser	Trp	Met	Ala	Leu	Met	Thr	Phe	Thr	Leu	Phe	Tyr	Thr	370	375	380
Asp	Phe	Val	Gly	Glu	Gly	Leu	Tyr	Gln	Gly	Val	Pro	Arg	Ala	Glu	Pro	385	390	395
Gly	Thr	Glu	Ala	Arg	Arg	His	Tyr	Asp	Glu	Gly	Val	Arg	Met	Gly	Ser	405	410	415
Leu	Gly	Leu	Phe	Leu	Gln	Cys	Ala	Ile	Ser	Leu	Val	Phe	Ser	Leu	Val	420	425	430
Met	Asp	Arg	Leu	Val	Gln	Arg	Phe	Gly	Thr	Arg	Ala	Val	Tyr	Leu	Ala	435	440	445
Ser	Val	Ala	Ala	Phe	Pro	Val	Ala	Ala	Gly	Ala	Thr	Cys	Leu	Ser	His	450	455	460
Ser	Val	Ala	Val	Val	Thr	Ala	Ser	Ala	Ala	Leu	Thr	Gly	Phe	Thr	Phe	465	470	475
Ser	Ala	Leu	Gln	Ile	Leu	Pro	Tyr	Thr	Leu	Ala	Ser	Leu	Tyr	His	Arg	485	490	495
Glu	Lys	Gln	Val	Phe	Leu	Pro	Lys	Tyr	Arg	Gly	Asp	Thr	Gly	Gly	Ala	500	505	510
Ser	Ser	Glu	Asp	Ser	Leu	Met	Thr	Ser	Phe	Leu	Pro	Gly	Pro	Lys	Pro	515	520	525
Gly	Ala	Pro	Phe	Pro	Asn	Gly	His	Val	Gly	Ala	Gly	Gly	Ser	Gly	Leu	530	535	540
Leu	Pro	Pro	Pro	Pro	Ala	Leu	Cys	Gly	Ala	Ser	Ala	Cys	Asp	Val	Ser	545	550	555
Val	Arg	Val	Val	Val	Gly	Glu	Pro	Thr	Glu	Ala	Arg	Val	Val	Pro	Gly	565	570	575
Arg	Gly	Ile	Cys	Leu	Asp	Leu	Ala	Ile	Leu	Asp	Ser	Ala	Phe	Leu	Leu	580	585	590
Ser	Gln	Val	Ala	Pro	Ser	Leu	Phe	Met	Gly	Ser	Ile	Val	Gln	Leu	Ser	595	600	605
Gln	Ser	Val	Thr	Ala	Tyr	Met	Val	Ser	Ala	Ala	Gly	Leu	Gly	Leu	Val	610	615	620
Ala	Ile	Tyr	Phe	Ala	Thr	Gln	Val	Val	Phe	Asp	Lys	Ser	Asp	Leu	Ala	625	630	635
Lys	Tyr	Ser	Ala	Gly	His	His	His	His	His	His	His	His	His	His	His	645	650	

<210> 42

<211> 1959

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando C-LytA de St. pneum. epitopo P2 auxiliar
C-Lyta fundido a P501S humano (mais marcador his)

<400> 42

```
atggcgggccg cttacgtaca ttccgacggc tcttatccaa aagacaagtt tgagaaaatc 60
aatggcactt ggtactactt tgacagttca ggctatatgc ttgcagaccg ctggaggaag 120
cacacagacg gcaactggta ctggttcgac aactcaggcg aaatggctac aggctggaag 180
aaaatcgctg ataagtggta ctatttcaac gaagaagggtg ccatgaagac aggctgggtc 240
aagtacaagg acacttggta ctacttagac gctaaagaag gcgcatgca atacatcaag 300
gctaactcta agttcattgg tatcactgaa ggcgtcatgg tatcaaatgc ctttatccag 360
tcagcggacg gaacaggetg gtactacctc aaaccagacg gaacactggc agacaggcca 420
gaaaagttca tgtacatggt gctgggcatt ggtccagtgc tgggcctggt ctgtgtcccg 480
ctcctagget cagccagtga ccactggcgt ggacgctatg gccgcccgcg gcccttcac 540
tgggcactgt ccttgggcat cctgctgagc ctctttctca tcccaagggc cggctggcta 600
gcagggctgc tgtgcccga tcccaggccc ctggagctgg cactgctcat cctgggcgtg 660
gggctgctgg acttctgtgg ccagggtgtg ttcactccac tggaggccct gctctctgac 720
ctcttccggg acccggaacca ctgtcgccag gcctactctg tctatgcctt catgatcagt 780
cttgggggct gccctgggcta cctcctgcct gccattgact gggacaccag tgccctggcc 840
ccctacctgg gcacccagga ggagtgcctc tttggcctgc tcaccctcat cttcctcacc 900
tgcgtagcag ccacactgct ggtggctgag gaggcagcgc tgggccccac cgagccagca 960
gaagggctgt cggccccctc cttgtcgccc cactgctgtc catgccgggc ccgcttggt 1020
ttccggaacc tgggcgcctt gcttccccgg ctgcaccagc tgtgtgcccg catgccccgc 1080
accctgcgcc ggctcttcgt ggctgagctg tgcagctgga tggcactcat gaccttcacg 1140
ctgttttaca cggatttcgt gggcgagggg ctgtaccagg gcgtgcccag agctgagccg 1200
ggcaccgagg cccggagaca ctatgatgaa ggcgttcgga tgggcagcct ggggctgttc 1260
ctgcagtgcg ccatctccct ggtcttctct ctggtcatgg accggctggt gcagcgattc 1320
ggcactcgag cagtctatct ggccagtgtg gcagctttcc ctgtggctgc cggtgccaca 1380
tgccgtgccc acagtgtggc cgtgggtgaca gcttcagccg ccctcaccgg gttcaccttc 1440
tcagccctgc agatccctgc ctacacactg gcctccctct accaccgga gaagcaggtg 1500
ttcctgcccc aataccgagg ggacactgga ggtgctagca gtgaggacag cctgatgacc 1560
agcttcctgc caggccctaa gcctggagct cccttcccta atggacacgt ggggtgctgga 1620
ggcagtgggc tgcctccacc tccaccgcg ctctgcgggg cctctgcctg tgatgtctcc 1680
gtacgtgtgg tgggtgggtga gcccaccgag gccagggtgg ttccgggccc gggcatctgc 1740
ctggacctcg ccatcctgga tagtgcttc ctgctgtccc aggtggcccc atccctgttt 1800
atgggctcca ttgtccagct cagccagtct gtcactgcct atatggtgtc tgccgcaggc 1860
ctgggtctgg tcgccattta ctttgctaca caggtagtat ttgacaagag cgacttggcc 1920
aaatactcag cgggtggaca ccatcaccat caccattaa 1959
```

<210> 43

<211> 553

<212> PRT

<213> Homo sapiens

<400> 43

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

```

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

```

<210> 44

<211> 644

<212> PRT

<213> Sequência Artificial

<220>

<223> C-LytA de St. pneum. epitopo P2 auxiliar C-Lyta fundido a P501S humano

<400> 44

```

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

```

Lys	Trp	Tyr	Tyr	Phe	Asn	Glu	Glu	Gly	Ala	Met	Lys	Thr	Gly	Trp	Val
65					70					75					80
Lys	Tyr	Lys	Asp	Thr	Trp	Tyr	Tyr	Leu	Asp	Ala	Lys	Glu	Gly	Ala	Met
			85						90					95	
Gln	Tyr	Ile	Lys	Ala	Asn	Ser	Lys	Phe	Ile	Gly	Ile	Thr	Glu	Gly	Val
			100					105					110		
Met	Val	Ser	Asn	Ala	Phe	Ile	Gln	Ser	Ala	Asp	Gly	Thr	Gly	Trp	Tyr
		115					120					125			
Tyr	Leu	Lys	Pro	Asp	Gly	Thr	Leu	Ala	Asp	Arg	Pro	Glu	Lys	Phe	Met
	130					135					140				
Tyr	Met	Val	Leu	Gly	Ile	Gly	Pro	Val	Leu	Gly	Leu	Val	Cys	Val	Pro
145					150					155					160
Leu	Leu	Gly	Ser	Ala	Ser	Asp	His	Trp	Arg	Gly	Arg	Tyr	Gly	Arg	Arg
			165						170					175	
Arg	Pro	Phe	Ile	Trp	Ala	Leu	Ser	Leu	Gly	Ile	Leu	Leu	Ser	Leu	Phe
			180					185					190		
Leu	Ile	Pro	Arg	Ala	Gly	Trp	Leu	Ala	Gly	Leu	Leu	Cys	Pro	Asp	Pro
		195					200					205			
Arg	Pro	Leu	Glu	Leu	Ala	Leu	Leu	Ile	Leu	Gly	Val	Gly	Leu	Leu	Asp
	210					215					220				
Phe	Cys	Gly	Gln	Val	Cys	Phe	Thr	Pro	Leu	Glu	Ala	Leu	Leu	Ser	Asp
225					230					235					240
Leu	Phe	Arg	Asp	Pro	Asp	His	Cys	Arg	Gln	Ala	Tyr	Ser	Val	Tyr	Ala
			245						250					255	
Phe	Met	Ile	Ser	Leu	Gly	Gly	Cys	Leu	Gly	Tyr	Leu	Leu	Pro	Ala	Ile
			260					265					270		
Asp	Trp	Asp	Thr	Ser	Ala	Leu	Ala	Pro	Tyr	Leu	Gly	Thr	Gln	Glu	Glu
		275					280						285		
Cys	Leu	Phe	Gly	Leu	Leu	Thr	Leu	Ile	Phe	Leu	Thr	Cys	Val	Ala	Ala
	290					295					300				
Thr	Leu	Leu	Val	Ala	Glu	Glu	Ala	Ala	Leu	Gly	Pro	Thr	Glu	Pro	Ala
305					310					315					320
Glu	Gly	Leu	Ser	Ala	Pro	Ser	Leu	Ser	Pro	His	Cys	Cys	Pro	Cys	Arg
			325						330					335	
Ala	Arg	Leu	Ala	Phe	Arg	Asn	Leu	Gly	Ala	Leu	Leu	Pro	Arg	Leu	His
			340					345					350		
Gln	Leu	Cys	Cys	Arg	Met	Pro	Arg	Thr	Leu	Arg	Arg	Leu	Phe	Val	Ala
	355					360						365			
Glu	Leu	Cys	Ser	Trp	Met	Ala	Leu	Met	Thr	Phe	Thr	Leu	Phe	Tyr	Thr
	370					375					380				
Asp	Phe	Val	Gly	Glu	Gly	Leu	Tyr	Gln	Gly	Val	Pro	Arg	Ala	Glu	Pro
385					390					395					400
Gly	Thr	Glu	Ala	Arg	Arg	His	Tyr	Asp	Glu	Gly	Val	Arg	Met	Gly	Ser
			405						410					415	
Leu	Gly	Leu	Phe	Leu	Gln	Cys	Ala	Ile	Ser	Leu	Val	Phe	Ser	Leu	Val
			420					425					430		
Met	Asp	Arg	Leu	Val	Gln	Arg	Phe	Gly	Thr	Arg	Ala	Val	Tyr	Leu	Ala
	435						440					445			
Ser	Val	Ala	Ala	Phe	Pro	Val	Ala	Ala	Gly	Ala	Thr	Cys	Leu	Ser	His
	450					455					460				
Ser	Val	Ala	Val	Val	Thr	Ala	Ser	Ala	Ala	Leu	Thr	Gly	Phe	Thr	Phe
465					470					475					480
Ser	Ala	Leu	Gln	Ile	Leu	Pro	Tyr	Thr	Leu	Ala	Ser	Leu	Tyr	His	Arg
			485						490					495	

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

<210> 45

<211> 644

<212> PRT

<213> Sequência Artificial

<220>

<223> proteína híbrida de codão otimizado entre C-LytA de St. pneum. epitopo P2 auxiliar C-Lyta fundido a P501S humano aminoácidos 51-553)

<400> 45

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

151

Gly	Ala	Pro	Phe	Pro	Asn	Gly	His	Val	Gly	Ala	Gly	Gly	Ser	Gly	Leu
530						535					540				
Leu	Pro	Pro	Pro	Pro	Ala	Leu	Cys	Gly	Ala	Ser	Ala	Cys	Asp	Val	Ser
545					550					555					560
Val	Arg	Val	Val	Val	Gly	Glu	Pro	Thr	Glu	Ala	Arg	Val	Val	Pro	Gly
				565					570					575	
Arg	Gly	Ile	Cys	Leu	Asp	Leu	Ala	Ile	Leu	Asp	Ser	Ala	Phe	Leu	Leu
			580					585					590		
Ser	Gln	Val	Ala	Pro	Ser	Leu	Phe	Met	Gly	Ser	Ile	Val	Gln	Leu	Ser
		595					600					605			
Gln	Ser	Val	Thr	Ala	Tyr	Met	Val	Ser	Ala	Ala	Gly	Leu	Gly	Leu	Val
	610					615					620				
Ala	Ile	Tyr	Phe	Ala	Thr	Gln	Val	Val	Phe	Asp	Lys	Ser	Asp	Leu	Ala
625					630					635					640
Lys	Tyr	Ser	Ala												

<210> 46

<211> 694

<212> PRT

<213> Sequência Artificial

<220>

<223> C-LytA de St. pneum. epitopo P2 auxiliar C-Lyta fundido a P501S humano (aminoácidos 1-553)- codão otimizado

<400> 46

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

145					150					155				160
Leu	Val	Asn	Leu	Leu	Thr	Phe	Gly	Leu	Glu	Val	Cys	Leu	Ala	Ala Gly
					165					170				175
Ile	Thr	Tyr	Val	Pro	Pro	Leu	Leu	Leu	Glu	Val	Gly	Val	Glu	Glu Lys
			180						185					190
Phe	Met	Thr	Met	Val	Leu	Gly	Ile	Gly	Pro	Val	Leu	Gly	Leu	Val Cys
		195					200					205		
Val	Pro	Leu	Leu	Gly	Ser	Ala	Ser	Asp	His	Trp	Arg	Gly	Arg	Tyr Gly
	210					215					220			
Arg	Arg	Arg	Pro	Phe	Ile	Trp	Ala	Leu	Ser	Leu	Gly	Ile	Leu	Leu Ser
	225				230					235				240
Leu	Phe	Leu	Ile	Pro	Arg	Ala	Gly	Trp	Leu	Ala	Gly	Leu	Leu	Cys Pro
			245						250					255
Asp	Pro	Arg	Pro	Leu	Glu	Leu	Ala	Leu	Leu	Ile	Leu	Gly	Val	Gly Leu
			260					265					270	
Leu	Asp	Phe	Cys	Gly	Gln	Val	Cys	Phe	Thr	Pro	Leu	Glu	Ala	Leu Leu
	275						280					285		
Ser	Asp	Leu	Phe	Arg	Asp	Pro	Asp	His	Cys	Arg	Gln	Ala	Tyr	Ser Val
	290					295					300			
Tyr	Ala	Phe	Met	Ile	Ser	Leu	Gly	Gly	Cys	Leu	Gly	Tyr	Leu	Leu Pro
	305				310					315				320
Ala	Ile	Asp	Trp	Asp	Thr	Ser	Ala	Leu	Ala	Pro	Tyr	Leu	Gly	Thr Gln
			325						330					335
Glu	Glu	Cys	Leu	Phe	Gly	Leu	Leu	Thr	Leu	Ile	Phe	Leu	Thr	Cys Val
			340					345					350	
Ala	Ala	Thr	Leu	Leu	Val	Ala	Glu	Glu	Ala	Ala	Leu	Gly	Pro	Thr Glu
		355					360					365		
Pro	Ala	Glu	Gly	Leu	Ser	Ala	Pro	Ser	Leu	Ser	Pro	His	Cys	Cys Pro
	370					375					380			
Cys	Arg	Ala	Arg	Leu	Ala	Phe	Arg	Asn	Leu	Gly	Ala	Leu	Leu	Pro Arg
	385				390					395				400
Leu	His	Gln	Leu	Cys	Cys	Arg	Met	Pro	Arg	Thr	Leu	Arg	Arg	Leu Phe
			405						410					415
Val	Ala	Glu	Leu	Cys	Ser	Trp	Met	Ala	Leu	Met	Thr	Phe	Thr	Leu Phe
			420					425					430	
Tyr	Thr	Asp	Phe	Val	Gly	Glu	Gly	Leu	Tyr	Gln	Gly	Val	Pro	Arg Ala
	435						440					445		
Glu	Pro	Gly	Thr	Glu	Ala	Arg	Arg	His	Tyr	Asp	Glu	Gly	Val	Arg Met
	450					455					460			
Gly	Ser	Leu	Gly	Leu	Phe	Leu	Gln	Cys	Ala	Ile	Ser	Leu	Val	Phe Ser
	465				470				475					480
Leu	Val	Met	Asp	Arg	Leu	Val	Gln	Arg	Phe	Gly	Thr	Arg	Ala	Val Tyr
			485						490					495
Leu	Ala	Ser	Val	Ala	Ala	Phe	Pro	Val	Ala	Ala	Gly	Ala	Thr	Cys Leu
			500					505					510	
Ser	His	Ser	Val	Ala	Val	Val	Thr	Ala	Ser	Ala	Ala	Leu	Thr	Gly Phe
	515						520					525		
Thr	Phe	Ser	Ala	Leu	Gln	Ile	Leu	Pro	Tyr	Thr	Leu	Ala	Ser	Leu Tyr
	530					535						540		
His	Arg	Glu	Lys	Gln	Val	Phe	Leu	Pro	Lys	Tyr	Arg	Gly	Asp	Thr Gly
	545				550					555				560
Gly	Ala	Ser	Ser	Glu	Asp	Ser	Leu	Met	Thr	Ser	Phe	Leu	Pro	Gly Pro
			565						570					575
Lys	Pro	Gly	Ala	Pro	Phe	Pro	Asn	Gly	His	Val	Gly	Ala	Gly	Gly Ser
			580					585					590	
Gly	Leu	Leu	Pro	Pro	Pro	Pro	Ala	Leu	Cys	Gly	Ala	Ser	Ala	Cys Asp
	595						600					605		
Val	Ser	Val	Arg	Val	Val	Val	Gly	Glu	Pro	Thr	Glu	Ala	Arg	Val Val
	610					615					620			
Pro	Gly	Arg	Gly	Ile	Cys	Leu	Asp	Leu	Ala	Ile	Leu	Asp	Ser	Ala Phe
	625				630					635				640

```

Leu Leu Ser Gln Val Ala Pro Ser Leu Phe Met Gly Ser Ile Val Gln
      645                      650                      655
Leu Ser Gln Ser Val Thr Ala Tyr Met Val Ser Ala Ala Gly Leu Gly
      660                      665                      670
Leu Val Ala Ile Tyr Phe Ala Thr Gln Val Val Phe Asp Lys Ser Asp
      675                      680                      685
Leu Ala Lys Tyr Ser Ala
      690

```

<210> 47

<211> 694

<212> PRT

<213> Sequência Artificial

<220>

<223> C-LyTA de St. pneum. epitopo P2 auxiliar C-LyTA fundido a P501S humano (aminoácidos 51-553) fundido a P501S humano (aminoácidos 1-50) - codão otimizado

<400> 47

```

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

```

Arg Pro Phe Ile Trp Ala Leu Ser Leu Gly Ile Leu Leu Ser Leu Phe
 180 185 190
 Leu Ile Pro Arg Ala Gly Trp Leu Ala Gly Leu Leu Cys Pro Asp Pro
 195 200 205
 Arg Pro Leu Glu Leu Ala Leu Leu Ile Leu Gly Val Gly Leu Leu Asp
 210 215 220
 Phe Cys Gly Gln Val Cys Phe Thr Pro Leu Glu Ala Leu Leu Ser Asp
 225 230 235 240
 Leu Phe Arg Asp Pro Asp His Cys Arg Gln Ala Tyr Ser Val Tyr Ala
 245 250 255
 Phe Met Ile Ser Leu Gly Gly Cys Leu Gly Tyr Leu Leu Pro Ala Ile
 260 265 270
 Asp Trp Asp Thr Ser Ala Leu Ala Pro Tyr Leu Gly Thr Gln Glu Glu
 275 280 285
 Cys Leu Phe Gly Leu Leu Thr Leu Ile Phe Leu Thr Cys Val Ala Ala
 290 295 300
 Thr Leu Leu Val Ala Glu Glu Ala Ala Leu Gly Pro Thr Glu Pro Ala
 305 310 315 320
 Glu Gly Leu Ser Ala Pro Ser Leu Ser Pro His Cys Cys Pro Cys Arg
 325 330 335
 Ala Arg Leu Ala Phe Arg Asn Leu Gly Ala Leu Leu Pro Arg Leu His
 340 345 350
 Gln Leu Cys Cys Arg Met Pro Arg Thr Leu Arg Arg Leu Phe Val Ala
 355 360 365
 Glu Leu Cys Ser Trp Met Ala Leu Met Thr Phe Thr Leu Phe Tyr Thr
 370 375 380
 Asp Phe Val Gly Glu Gly Leu Tyr Gln Gly Val Pro Arg Ala Glu Pro
 385 390 395 400
 Gly Thr Glu Ala Arg Arg His Tyr Asp Glu Gly Val Arg Met Gly Ser
 405 410 415
 Leu Gly Leu Phe Leu Gln Cys Ala Ile Ser Leu Val Phe Ser Leu Val
 420 425 430
 Met Asp Arg Leu Val Gln Arg Phe Gly Thr Arg Ala Val Tyr Leu Ala
 435 440 445
 Ser Val Ala Ala Phe Pro Val Ala Ala Gly Ala Thr Cys Leu Ser His
 450 455 460
 Ser Val Ala Val Val Thr Ala Ser Ala Ala Leu Thr Gly Phe Thr Phe
 465 470 475 480
 Ser Ala Leu Gln Ile Leu Pro Tyr Thr Leu Ala Ser Leu Tyr His Arg
 485 490 495
 Glu Lys Gln Val Phe Leu Pro Lys Tyr Arg Gly Asp Thr Gly Gly Ala
 500 505 510
 Ser Ser Glu Asp Ser Leu Met Thr Ser Phe Leu Pro Gly Pro Lys Pro
 515 520 525
 Gly Ala Pro Phe Pro Asn Gly His Val Gly Ala Gly Gly Ser Gly Leu
 530 535 540
 Leu Pro Pro Pro Pro Ala Leu Cys Gly Ala Ser Ala Cys Asp Val Ser
 545 550 555 560
 Val Arg Val Val Val Gly Glu Pro Thr Glu Ala Arg Val Val Pro Gly
 565 570 575
 Arg Gly Ile Cys Leu Asp Leu Ala Ile Leu Asp Ser Ala Phe Leu Leu
 580 585 590
 Ser Gln Val Ala Pro Ser Leu Phe Met Gly Ser Ile Val Gln Leu Ser
 595 600 605

```

Gln Ser Val Thr Ala Tyr Met Val Ser Ala Ala Gly Leu Gly Leu Val
 610          615          620
Ala Ile Tyr Phe Ala Thr Gln Val Val Phe Asp Lys Ser Asp Leu Ala
625          630          635          640
Lys Tyr Ser Ala Met Val Gln Arg Leu Trp Val Ser Arg Leu Leu Arg
          645          650          655
His Arg Lys Ala Gln Leu Leu Leu Val Asn Leu Leu Thr Phe Gly Leu
          660          665          670
Glu Val Cys Leu Ala Ala Gly Ile Thr Tyr Val Pro Pro Leu Leu Leu
          675          680          685
Glu Val Gly Val Glu Glu
          690

```

<210> 48

<211> 694

<212> PRT

<213> Sequência Artificial

<220>

<223> P501S humano (aminoácidos 1-50) fundido a C-LytA de St. pneum. epitopo P2 auxiliar C-Lyta fundido a P501S humano (aminoácidos 51-553) - codão otimizado

<400> 48

```

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

```

Gly	Val	Met	Val	Ser	Asn	Ala	Phe	Ile	Gln	Ser	Ala	Asp	Gly	Thr	Gly	165	170	175
Trp	Tyr	Tyr	Leu	Lys	Pro	Asp	Gly	Thr	Leu	Ala	Asp	Arg	Pro	Glu	Lys	180	185	190
Phe	Met	Tyr	Met	Val	Leu	Gly	Ile	Gly	Pro	Val	Leu	Gly	Leu	Val	Cys	195	200	205
Val	Pro	Leu	Leu	Gly	Ser	Ala	Ser	Asp	His	Trp	Arg	Gly	Arg	Tyr	Gly	210	215	220
Arg	Arg	Arg	Pro	Phe	Ile	Trp	Ala	Leu	Ser	Leu	Gly	Ile	Leu	Leu	Ser	225	230	235
Leu	Phe	Leu	Ile	Pro	Arg	Ala	Gly	Trp	Leu	Ala	Gly	Leu	Leu	Cys	Pro	245	250	255
Asp	Pro	Arg	Pro	Leu	Glu	Leu	Ala	Leu	Leu	Ile	Leu	Gly	Val	Gly	Leu	260	265	270
Leu	Asp	Phe	Cys	Gly	Gln	Val	Cys	Phe	Thr	Pro	Leu	Glu	Ala	Leu	Leu	275	280	285
Ser	Asp	Leu	Phe	Arg	Asp	Pro	Asp	His	Cys	Arg	Gln	Ala	Tyr	Ser	Val	290	295	300
Tyr	Ala	Phe	Met	Ile	Ser	Leu	Gly	Gly	Cys	Leu	Gly	Tyr	Leu	Leu	Pro	305	310	315
Ala	Ile	Asp	Trp	Asp	Thr	Ser	Ala	Leu	Ala	Pro	Tyr	Leu	Gly	Thr	Gln	325	330	335
Glu	Glu	Cys	Leu	Phe	Gly	Leu	Leu	Thr	Leu	Ile	Phe	Leu	Thr	Cys	Val	340	345	350
Ala	Ala	Thr	Leu	Leu	Val	Ala	Glu	Glu	Ala	Ala	Leu	Gly	Pro	Thr	Glu	355	360	365
Pro	Ala	Glu	Gly	Leu	Ser	Ala	Pro	Ser	Leu	Ser	Pro	His	Cys	Cys	Pro	370	375	380
Cys	Arg	Ala	Arg	Leu	Ala	Phe	Arg	Asn	Leu	Gly	Ala	Leu	Leu	Pro	Arg	385	390	395
Leu	His	Gln	Leu	Cys	Cys	Arg	Met	Pro	Arg	Thr	Leu	Arg	Arg	Leu	Phe	405	410	415
Val	Ala	Glu	Leu	Cys	Ser	Trp	Met	Ala	Leu	Met	Thr	Phe	Thr	Leu	Phe	420	425	430
Tyr	Thr	Asp	Phe	Val	Gly	Glu	Gly	Leu	Tyr	Gln	Gly	Val	Pro	Arg	Ala	435	440	445
Glu	Pro	Gly	Thr	Glu	Ala	Arg	Arg	His	Tyr	Asp	Glu	Gly	Val	Arg	Met	450	455	460
Gly	Ser	Leu	Gly	Leu	Phe	Leu	Gln	Cys	Ala	Ile	Ser	Leu	Val	Phe	Ser	465	470	475
Leu	Val	Met	Asp	Arg	Leu	Val	Gln	Arg	Phe	Gly	Thr	Arg	Ala	Val	Tyr	485	490	495
Leu	Ala	Ser	Val	Ala	Ala	Phe	Pro	Val	Ala	Ala	Gly	Ala	Thr	Cys	Leu	500	505	510
Ser	His	Ser	Val	Ala	Val	Val	Thr	Ala	Ser	Ala	Ala	Leu	Thr	Gly	Phe	515	520	525
Thr	Phe	Ser	Ala	Leu	Gln	Ile	Leu	Pro	Tyr	Thr	Leu	Ala	Ser	Leu	Tyr	530	535	540
His	Arg	Glu	Lys	Gln	Val	Phe	Leu	Pro	Lys	Tyr	Arg	Gly	Asp	Thr	Gly	545	550	555
Gly	Ala	Ser	Ser	Glu	Asp	Ser	Leu	Met	Thr	Ser	Phe	Leu	Pro	Gly	Pro	565	570	575

Lys	Pro	Gly	Ala	Pro	Phe	Pro	Asn	Gly	His	Val	Gly	Ala	Gly	Gly	Ser
			580					585					590		
Gly	Leu	Leu	Pro	Pro	Pro	Pro	Ala	Leu	Cys	Gly	Ala	Ser	Ala	Cys	Asp
		595					600					605			
Val	Ser	Val	Arg	Val	Val	Val	Gly	Glu	Pro	Thr	Glu	Ala	Arg	Val	Val
	610					615					620				
Pro	Gly	Arg	Gly	Ile	Cys	Leu	Asp	Leu	Ala	Ile	Leu	Asp	Ser	Ala	Phe
625					630					635					640
Leu	Leu	Ser	Gln	Val	Ala	Pro	Ser	Leu	Phe	Met	Gly	Ser	Ile	Val	Gln
			645						650					655	
Leu	Ser	Gln	Ser	Val	Thr	Ala	Tyr	Met	Val	Ser	Ala	Ala	Gly	Leu	Gly
		660						665					670		
Leu	Val	Ala	Ile	Tyr	Phe	Ala	Thr	Gln	Val	Val	Phe	Asp	Lys	Ser	Asp
		675					680					685			
Leu	Ala	Lys	Tyr	Ser	Ala										
	690														

<210> 49

<211> 1971

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando MUC-1 Humano fundido a C-LyTA de St. pneum. epitopo P2 auxiliar C-Lyta

<400> 49

```

atgacaccgg gcaccagtc tcctttcttc ctgctgctgc tcctcacagt gcttacagtt 60
gttacagggt ctggtcatgc aagctctacc ccaggtggag aaaaggagac ttcggctacc 120
cagagaaggt cagtgccag ctctactgag aagaatgctg tgagtatgac cagcagcgt 180
ctctccagcc acagcccggg ttcaggctcc tccaccactc agggacagga tgtcactctg 240
gccccggcca cggaaccagc ttcaggttca gctgccacct ggggacagga tgtcacctcg 300
gtcccagtc cagggccagc cctgggctcc accaccccg cagcccacga tgtcacctca 360
gccccggaca acaagccagc cccgggctcc accgcccccc cagcccacgg tgtcacctcg 420
gccccggaca ccaggccgccc cccgggctcc accgcccccc cagcccacgg tgtcacctcg 480
gccccggaca ccaggccgccc cccgggctcc accgcccccc cagcccacgg tgtcacctcg 540
gccccggaca ccaggccggc cccgggctcc accgcccccc cagcccacgg tgtcacctcg 600
gccccggaca acaggccgccc cttggcgctc accgcccccc cagtccacaa tgtcacctcg 660
gcctcagggt ctgcatcagg ctccagcttct actctgggtg acaacggcac ctctgccagg 720
gctaccacaa cccagccag caagagcact ccattctcaa tcccagcca ccactctgat 780
actcctacca cccttgccag ccatagcacc aagactgatg ccagtagcac tcaccatagc 840
acggtacctc ctctcacctc ctccaatcac agcacttctc cccagttgtc tactggggtc 900
tctttctttt tcctgtcttt tcacatttca aacctccagt ttaattcctc tctggaagat 960
cccagcaccc actactacca agagctgcag agagacattt ctgaaatgtt tttgcagatt 1020
tataaacaag ggggttttct gggcctctcc aatattaagt tcaggccagg atctgtggtg 1080
gtacaattga ctctggcctt ccgagaaggt accatcaatg tccacgacgt ggagacacag 1140

```

```

ttcaatcagt ataaaacgga agcagcctct cgatataacc tgacgatctc agacgtcagc 1200
gtgagtgatg tgccatttcc tttctctgcc cagtctgggg ctgggggtgcc aggctggggc 1260
atcgcgctgc tgggtgctggc ctgtgttctg gttgcgctgg ccattgtcta tctcattgcc 1320
ttggctgtct gtcagtgccg ccgaaagaac tacgggcagc tggacatctt tccagcccgg 1380
gataacctacc atcctatgag cgagtacccc acctaccaca cccatgggcg ctatgtgccc 1440
cctagcagta ccgatcgtag cccctatgag aagggttctg caggtaatgg tggcagcagc 1500
ctctcttaca caaaccagc agtggcagcc acttctgcca acttgatggc ggccgcttac 1560
gtacattccg acggctctta tccaaaagac aagtttgaga aaatcaatgg cacttggtac 1620
tactttgaca gttcaggcta tatgcttgca gaccgctgga ggaagcacac agacggcaac 1680
tggtactggt tcgacaaactc aggcgaaatg gctacaggct ggaagaaaat cgctgataag 1740
tggtactatt tcaacgaaga aggtgccatg aagacaggct gggcaagta caaggacact 1800
tggtactact tagacgctaa agaaggcgcc atgcaatata tcaaggctaa ctctaagttc 1860
attggtatca ctgaaggcgt catggtatca aatgccttta tccagtcagc ggacggaaca 1920
ggctggtact acctcaaacc agacggaaca ctggcagaca ggccagaatg a 1971

```

<210> 50

<211> 656

<212> PRT

<213> Sequência Artificial

<220>

<223> MUC-1 humano fundido a C-LyTA de St. pneum. epitopo P2
auxiliar C-Lyta

<400> 50

```

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

```

Arg	Pro	Pro	Pro	Gly	Ser	Thr	Ala	Pro	Pro	Ala	His	Gly	Val	Thr	Ser
145					150					155					160
Ala	Pro	Asp	Thr	Arg	Pro	Pro	Pro	Gly	Ser	Thr	Ala	Pro	Ala	Ala	His
				165					170						175
Gly	Val	Thr	Ser	Ala	Pro	Asp	Thr	Arg	Pro	Ala	Pro	Gly	Ser	Thr	Ala
			180					185					190		
Pro	Pro	Ala	His	Gly	Val	Thr	Ser	Ala	Pro	Asp	Asn	Arg	Pro	Ala	Leu
		195					200					205			
Ala	Ser	Thr	Ala	Pro	Pro	Val	His	Asn	Val	Thr	Ser	Ala	Ser	Gly	Ser
	210					215					220				
Ala	Ser	Gly	Ser	Ala	Ser	Thr	Leu	Val	His	Asn	Gly	Thr	Ser	Ala	Arg
225					230					235					240
Ala	Thr	Thr	Thr	Pro	Ala	Ser	Lys	Ser	Thr	Pro	Phe	Ser	Ile	Pro	Ser
				245					250						255
His	His	Ser	Asp	Thr	Pro	Thr	Thr	Leu	Ala	Ser	His	Ser	Thr	Lys	Thr
			260					265					270		
Asp	Ala	Ser	Ser	Thr	His	His	Ser	Thr	Val	Pro	Pro	Leu	Thr	Ser	Ser
		275					280					285			
Asn	His	Ser	Thr	Ser	Pro	Gln	Leu	Ser	Thr	Gly	Val	Ser	Phe	Phe	Phe
	290					295					300				
Leu	Ser	Phe	His	Ile	Ser	Asn	Leu	Gln	Phe	Asn	Ser	Ser	Leu	Glu	Asp
305					310					315					320
Pro	Ser	Thr	Asp	Tyr	Tyr	Gln	Glu	Leu	Gln	Arg	Asp	Ile	Ser	Glu	Met
			325						330					335	
Phe	Leu	Gln	Ile	Tyr	Lys	Gln	Gly	Gly	Phe	Leu	Gly	Leu	Ser	Asn	Ile
			340					345					350		
Lys	Phe	Arg	Pro	Gly	Ser	Val	Val	Val	Gln	Leu	Thr	Leu	Ala	Phe	Arg
		355					360					365			
Glu	Gly	Thr	Ile	Asn	Val	His	Asp	Val	Glu	Thr	Gln	Phe	Asn	Gln	Tyr
	370					375					380				
Lys	Thr	Glu	Ala	Ala	Ser	Arg	Tyr	Asn	Leu	Thr	Ile	Ser	Asp	Val	Ser
385					390					395					400
Val	Ser	Asp	Val	Pro	Phe	Pro	Phe	Ser	Ala	Gln	Ser	Gly	Ala	Gly	Val
				405					410					415	
Pro	Gly	Trp	Gly	Ile	Ala	Leu	Leu	Val	Leu	Val	Cys	Val	Leu	Val	Ala
			420					425					430		
Leu	Ala	Ile	Val	Tyr	Leu	Ile	Ala	Leu	Ala	Val	Cys	Gln	Cys	Arg	Arg
		435					440					445			
Lys	Asn	Tyr	Gly	Gln	Leu	Asp	Ile	Phe	Pro	Ala	Arg	Asp	Thr	Tyr	His
	450					455					460				
Pro	Met	Ser	Glu	Tyr	Pro	Thr	Tyr	His	Thr	His	Gly	Arg	Tyr	Val	Pro
465					470					475					480
Pro	Ser	Ser	Thr	Asp	Arg	Ser	Pro	Tyr	Glu	Lys	Val	Ser	Ala	Gly	Asn
				485					490					495	
Gly	Gly	Ser	Ser	Leu	Ser	Tyr	Thr	Asn	Pro	Ala	Val	Ala	Ala	Thr	Ser
			500					505					510		
Ala	Asn	Leu	Met	Ala	Ala	Ala	Tyr	Val	His	Ser	Asp	Gly	Ser	Tyr	Pro
		515					520					525			
Lys	Asp	Lys	Phe	Glu	Lys	Ile	Asn	Gly	Thr	Trp	Tyr	Tyr	Phe	Asp	Ser
	530					535					540				
Ser	Gly	Tyr	Met	Leu	Ala	Asp	Arg	Trp	Arg	Lys	His	Thr	Asp	Gly	Asn
545					550					555					560

Trp	Tyr	Trp	Phe	Asp	Asn	Ser	Gly	Glu	Met	Ala	Thr	Gly	Trp	Lys	Lys
				565					570					575	
Ile	Ala	Asp	Lys	Trp	Tyr	Tyr	Phe	Asn	Glu	Glu	Gly	Ala	Met	Lys	Thr
			580					585					590		
Gly	Trp	Val	Lys	Tyr	Lys	Asp	Thr	Trp	Tyr	Tyr	Leu	Asp	Ala	Lys	Glu
		595					600					605			
Gly	Ala	Met	Gln	Tyr	Ile	Lys	Ala	Asn	Ser	Lys	Phe	Ile	Gly	Ile	Thr
	610					615					620				
Glu	Gly	Val	Met	Val	Ser	Asn	Ala	Phe	Ile	Gln	Ser	Ala	Asp	Gly	Thr
625					630					635				640	
Gly	Trp	Tyr	Tyr	Leu	Lys	Pro	Asp	Gly	Thr	Leu	Ala	Asp	Arg	Pro	Glu
				645					650					655	

<210> 51

<211> 2037

<212> ADN

<213> Sequência Artificial

<220>

<223> ADN codificando C-LytA de St. pneum. epitopo P2 auxiliar
C-LytA fundido a MUC-1 humano

<400> 51

```

atgggatgga gctgtatcat cctcttcttg gtagcaacag ctacaggtgt ccactcccag 60
gtccaaatgg cggccgctta cgtacattcc gacggctctt atccaaaaga caagtttgag 120
aaaatcaatg gcacttggtg ctactttgac agttcaggct atatgcttgc agaccgctgg 180
aggaagcaca cagacggcaa ctggtactgg ttcgacaact caggcgaaat ggctacaggc 240
tggaagaaaa tcgtgataa gtggtactat ttcaacgaag aaggtgccat gaagacaggc 300
tgggtcaagt acaaggacac ttggtactac ttagacgcta aagaaggcgc catgcaatac 360
atcaaggcta actctaagtt cattggtatc actgaaggcg tcatggtatc aaatgccttt 420
atccagtcag cggacggaac aggctggtac tacctcaaac cagacggaac actggcagac 480
aggccagaaa tgacaccggg caccagtcct cctttcttcc tgctgctgct cctcacagtg 540
cttacagttg ttacagggtc tggcatgca agctctaccc cagggtggaga aaaggagact 600
tcggctaccc agagaagttc agtgcccagc tctactgaga agaatgctgt gagtatgacc 660
agcagcgtac tctccagcca cagccccggt tcaggctcct ccaccactca gggacaggat 720
gtcactctgg ccccggccac ggaaccagct tcaggttcag ctgccacctg gggacaggat 780
gtcacctcgg tcccagtcac caggccagcc ctgggctcca ccaccccgcc agcccacgat 840
gtcacctcag ccccggacaa caagccagcc ccgggctcca ccgccccccc agcccacggt 900
gtcacctcgg ccccggacac caggccgccc ccgggctcca ccgccccccc agcccacggt 960
gtcacctcgg ccccggacac caggccgccc ccgggctcca ccgccccccc agcccacggt 1020
gtcacctcgg ccccggacaa caggccgccc ctgggctcca ccgccccccc agcccacggt 1080
gtcacctcgg cctcaggctc tgcacaggc tcagcttcta ctctggtgca caacggcacc 1140
tctgccaggg ctaccacaac cccagccagc aagagcactc cattctcaat tcccagccac 1260
cactctgata ctctaccac ccttgccagc catagcacca agactgatgc cagtagcact 1320

```

```

caccatagca cggtagctcc tctcacctcc tccaatcaca gcacttctcc ccagttgtct 1380
actgggggtct ctttcttttt cctgtctttt cacatttcaa acctccagtt taattcctct 1440
ctggaagatc ccagcaccca ctactacca gagctgcaga gagacatttc tgaaatgttt 1500
ttgcagattt ataaacaagg gggttttctg ggctctcca atattaagtt caggccagga 1560
tctgtgggtg tacaattgac tctggccttc cgagaaggta ccatcaatgt ccacgacgtg 1620
gagacacagt tcaatcagta taaaacggaa gcagcctctc gatataacct gacgatctca 1680
gacgtcagcg tgagtgatgt gccatttctt ttctctgccc agtctggggc tggggtgcca 1740
ggctggggca tcgcgtgct ggtgctggtc tgtgttctgg ttgcgctggc cattgtctat 1800
ctcattgcct tggctgtctg tcagtgccgc cgaaagaact acgggcagct ggacatcttt 1860
ccagcccggg atacctacca tcctatgagc gagtacccca cctaccacac ccatgggcgc 1920
tatgtgcccc ctagcagtac cgatcgtagc ccctatgaga aggtttctgc aggtaatggt 1980
ggcagcagcc tctcttacac aaaccagca gtggcagcca cttctgcca cttgtag 2037

```

<210> 52

<211> 678

<212> PRT

<213> Sequência Artificial

<220>

<223> C-LytA de St. pneum. epitopo P2 auxiliar C-LytA fundido a MUC-1 humano

<400> 52

```

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

```

Leu	Leu	Thr	Val	Leu	Thr	Val	Val	Thr	Gly	Ser	Gly	His	Ala	Ser	Ser	180	185	190
Thr	Pro	Gly	Gly	Glu	Lys	Glu	Thr	Ser	Ala	Thr	Gln	Arg	Ser	Ser	Val	195	200	205
Pro	Ser	Ser	Thr	Glu	Lys	Asn	Ala	Val	Ser	Met	Thr	Ser	Ser	Val	Leu	210	215	220
Ser	Ser	His	Ser	Pro	Gly	Ser	Gly	Ser	Ser	Thr	Thr	Gln	Gly	Gln	Asp	225	230	235
Val	Thr	Leu	Ala	Pro	Ala	Thr	Glu	Pro	Ala	Ser	Gly	Ser	Ala	Ala	Thr	245	250	255
Trp	Gly	Gln	Asp	Val	Thr	Ser	Val	Pro	Val	Thr	Arg	Pro	Ala	Leu	Gly	260	265	270
Ser	Thr	Thr	Pro	Pro	Ala	His	Asp	Val	Thr	Ser	Ala	Pro	Asp	Asn	Lys	275	280	285
Pro	Ala	Pro	Gly	Ser	Thr	Ala	Pro	Pro	Ala	His	Gly	Val	Thr	Ser	Ala	290	295	300
Pro	Asp	Thr	Arg	Pro	Pro	Pro	Gly	Ser	Thr	Ala	Pro	Pro	Ala	His	Gly	305	310	315
Val	Thr	Ser	Ala	Pro	Asp	Thr	Arg	Pro	Pro	Pro	Gly	Ser	Thr	Ala	Pro	325	330	335
Ala	Ala	His	Gly	Val	Thr	Ser	Ala	Pro	Asp	Thr	Arg	Pro	Ala	Pro	Gly	340	345	350
Ser	Thr	Ala	Pro	Pro	Ala	His	Gly	Val	Thr	Ser	Ala	Pro	Asp	Asn	Arg	355	360	365
Pro	Ala	Leu	Ala	Ser	Thr	Ala	Pro	Pro	Val	His	Asn	Val	Thr	Ser	Ala	370	375	380
Ser	Gly	Ser	Ala	Ser	Gly	Ser	Ala	Ser	Thr	Leu	Val	His	Asn	Gly	Thr	385	390	395
Ser	Ala	Arg	Ala	Thr	Thr	Thr	Pro	Ala	Ser	Lys	Ser	Thr	Pro	Phe	Ser	405	410	415
Ile	Pro	Ser	His	His	Ser	Asp	Thr	Pro	Thr	Thr	Leu	Ala	Ser	His	Ser	420	425	430
Thr	Lys	Thr	Asp	Ala	Ser	Ser	Thr	His	His	Ser	Thr	Val	Pro	Pro	Leu	435	440	445
Thr	Ser	Ser	Asn	His	Ser	Thr	Ser	Pro	Gln	Leu	Ser	Thr	Gly	Val	Ser	450	455	460
Phe	Phe	Phe	Leu	Ser	Phe	His	Ile	Ser	Asn	Leu	Gln	Phe	Asn	Ser	Ser	465	470	475
Leu	Glu	Asp	Pro	Ser	Thr	Asp	Tyr	Tyr	Gln	Glu	Leu	Gln	Arg	Asp	Ile	485	490	495
Ser	Glu	Met	Phe	Leu	Gln	Ile	Tyr	Lys	Gln	Gly	Gly	Phe	Leu	Gly	Leu	500	505	510
Ser	Asn	Ile	Lys	Phe	Arg	Pro	Gly	Ser	Val	Val	Val	Gln	Leu	Thr	Leu	515	520	525
Ala	Phe	Arg	Glu	Gly	Thr	Ile	Asn	Val	His	Asp	Val	Glu	Thr	Gln	Phe	530	535	540
Asn	Gln	Tyr	Lys	Thr	Glu	Ala	Ala	Ser	Arg	Tyr	Asn	Leu	Thr	Ile	Ser	545	550	555
Asp	Val	Ser	Val	Ser	Asp	Val	Pro	Phe	Pro	Phe	Ser	Ala	Gln	Ser	Gly	565	570	575
Ala	Gly	Val	Pro	Gly	Trp	Gly	Ile	Ala	Leu	Leu	Val	Leu	Val	Cys	Val	580	585	590
Leu	Val	Ala	Leu	Ala	Ile	Val	Tyr	Leu	Ile	Ala	Leu	Ala	Val	Cys	Gln	595	600	605

Cys	Arg	Arg	Lys	Asn	Tyr	Gly	Gln	Leu	Asp	Ile	Phe	Pro	Ala	Arg	Asp
610						615					620				
Thr	Tyr	His	Pro	Met	Ser	Glu	Tyr	Pro	Thr	Tyr	His	Thr	His	Gly	Arg
625					630					635					640
Tyr	Val	Pro	Pro	Ser	Ser	Thr	Asp	Arg	Ser	Pro	Tyr	Glu	Lys	Val	Ser
				645					650					655	
Ala	Gly	Asn	Gly	Gly	Ser	Ser	Leu	Ser	Tyr	Thr	Asn	Pro	Ala	Val	Ala
			660					665					670		
Ala	Thr	Ser	Ala	Asn	Leu										
		675													

Lisboa, 4 de Julho de 2007

REIVINDICAÇÕES

1. Proteína parceira de fusão compreendendo um domínio de ligação à colina e um epítipo heterólogo auxiliar T promíscuo.
2. Proteína parceira de fusão de acordo com a reivindicação 1, em que o domínio de ligação à colina é o C-terminal de LytA ou um seu derivado, em que o derivado do C-terminal de LytA retém a capacidade de actuar como um parceiro imunológico e como um intensificador de expressão.
3. Proteína parceira de fusão de acordo com a reivindicação 2, em que o C-LytA ou seu derivado compreende, pelo menos, quatro repetições de qualquer das SEQ ID N°:1 a 6.
4. Proteína parceira de fusão de acordo com qualquer das reivindicações 1 a 3, em que o domínio de ligação à colina é seleccionado do grupo compreendendo:
 - a) o domínio do C-terminal de LytA como apresentado em SEQ ID N°:7; ou
 - b) a sequência de SEQ ID N°:8; ou
 - c) uma sequência peptídica compreendendo uma sequência de aminoácidos tendo, pelo menos, 85% de identidade, de um modo preferido, pelo menos, 90% de identidade, de um modo mais preferido, pelo menos, 95% de identidade, de um modo muito preferido, pelo menos, 97-99% de identidade, com qualquer de SEQ ID N°:1 a 6; ou
 - d) uma sequência peptídica compreendendo uma sequência de aminoácidos que tem, pelo menos, 15, 20, 30, 40, 50

ou 100 aminoácidos contíguos da sequência de aminoácidos de SEQ ID a N°:7 ou SEQ ID N°:8.

5. Proteína de fusão compreendendo uma proteína parceira de fusão como reivindicado em qualquer das reivindicações 1 a 4 e uma proteína heteróloga.
6. Proteína de fusão como reivindicado na reivindicação 5, em que a proteína heteróloga é quimicamente conjugada ao parceiro de fusão.
7. Proteína de fusão como reivindicado na reivindicação 5 ou 6, em que a proteína heteróloga é derivado de um organismo seleccionado do seguinte grupo: Vírus da imunodeficiência humana HIV-1, vírus herpes simplex humano, citomegalovírus, Rotavírus, vírus de Epstein Barr, Vírus da Varicella Zoster, de um vírus da hepatite, tal como o vírus da hepatite B, vírus da hepatite A, vírus da hepatite C e vírus da hepatite E, de vírus Sincicial Respiratório, vírus da parainfluenza, vírus do sarampo, vírus da parotidite, vírus do papiloma humano, flavivírus ou vírus de Influenza, de *Neisseria spp*, *Moraxella spp*, *Bordetella spp*; *Mycobacterium spp.*, incluindo *M. tuberculosis*; *Escherichia spp*, incluindo *E. coli enterotóxica*; *Salmonella spp*; *Listeria spp*; *Helicobacter spp*; *Staphylococcus spp.*, incluindo *S. aureus*, *S. epidermidis*; *Borrelia spp*; *Chlamydia spp.*, incluindo *C. trachomatis*, *C. pneumoniae*; *Plasmodium spp.*, incluindo *P. falciparum*; *Toxoplasma spp.*, *Candida spp.*
8. Proteína de fusão como reivindicado na reivindicação 5 ou 6, em que a proteína heteróloga é uma proteína associada a tumor ou proteína específica de tecido ou seu fragmento imunogénico.

9. Proteína de fusão como reivindicado na reivindicação 8, em que a proteína heteróloga ou seu fragmento é seleccionada de MAGE 1, MAGE 3, MAGE 4, PRAME, BAGE, LAGE 1, LAGE 2, SAGE, HAGE, XAGE, PSA, PAP, PSCA, prostateína, P501S, HASH2, Cripto, B726, NY-BR1.1, P510, MUC-1, Prostase, STEAP, tirosinase, telomerase, survivina, CASB616, P53 ou her 2 neu.
10. Proteína de fusão como reivindicado em qualquer das reivindicações 6 a 9 compreendendo ainda um marcador de afinidade ou, pelo menos, 4 resíduos de histidina.
11. Sequência de ácido nucleico codificando uma proteína da reivindicação 1 a 10.
12. Vector de expressão compreendendo uma sequência de ácido nucleico da reivindicação 11.
13. Célula hospedeira transformada com uma sequência de ácido nucleico da reivindicação 11 ou com um vector de expressão da reivindicação 12.
14. Composição imunogénica compreendendo uma proteína como reivindicado em qualquer das reivindicações 1 a 10 ou uma sequência de ADN como reivindicado na reivindicação 11 e um excipiente farmacêuticamente aceitável.
15. Composição imunogénica como reivindicado na reivindicação 14 que compreende ainda um adjuvante indutor de TH-1.
16. Composição imunogénica como reivindicado na reivindicação 15, em que o adjuvante indutor de TH-1 é seleccionado do grupo de adjuvantes compreendendo: 3D-MPL, QS21, uma mistura

de QS21 e colesterol, um oligonucleótido CpG ou uma mistura de dois ou mais dos referidos adjuvantes.

17. Processo para a preparação de uma composição imunogénica como reivindicado em qualquer das reivindicações 14 a 16, compreendendo a mistura da proteína de fusão de qualquer das reivindicações 6 a 10 ou do polinucleótido codificante da reivindicação 11 com um adjuvante, diluente ou outro veículo farmacêuticamente aceitável adequado.
18. Processo para a produção de uma proteína de fusão de qualquer das reivindicações 1 a 10, compreendendo a cultura de uma célula hospedeira da reivindicação 13 sob condições suficientes para a produção da referida proteína de fusão e a recuperação da proteína de fusão do meio de cultura.
19. Proteína de qualquer das reivindicações 1 a 10 ou uma sequência de ADN da reivindicação 11 para utilização em medicina.
20. Utilização de uma proteína como reivindicado em qualquer das reivindicações 1 a 10 ou uma sequência de ADN da reivindicação 11 na produção de uma composição imunogénica para induzir uma resposta imunológica num doente.
21. Utilização de acordo com a reivindicação 20, em que referida resposta imunológica é para ser induzida por administração sequencial i) da referida proteína seguida por pela referida sequência de ADN; ou ii) da referida sequência de ADN seguida pela referida proteína.
22. Utilização de acordo com a reivindicação 21, em que a referida sequência de ADN é revestida sobre esferas

biodegradáveis ou distribuída através de uma abordagem de bombardeamento de partículas.

23. Utilização de acordo com a reivindicação 21 ou reivindicação 22, em que a referida proteína é adjuvantada.
24. Utilização de uma proteína como reivindicado em qualquer das reivindicações 1 a 10 ou uma sequência de ADN da reivindicação 11 na produção de uma composição imunogénica para tratar imunoterapeuticamente um doente que sofre ou é susceptível ao cancro.
25. Utilização de acordo com a reivindicação 24, em que o referido cancro é o cancro da próstata, cancro do cólon, cancro do pulmão, cancro da mama ou melanoma.

Lisboa, 4 de Julho de 2007

Fig. 1 - Informação de Sequência para C-LytA.

SEQ ID N°1 - sequência de aminoácidos da repetição 1 de C-LytA
 GWQKNDTGYWYVHSD 15

SEQ ID N°2 - sequência de aminoácidos da repetição 2 de C-LytA
 GSYPKDKFEKINGTWYFDSS 21

SEQ ID N°3 - sequência de aminoácidos da repetição 3 de C-LytA
 GYMLADRWRKHTDGNWYFDNS 22

SEQ ID N°4 - sequência de aminoácidos da repetição 4 de C-LytA
 GEMATGWKKIADKWYFNEE 20

SEQ ID N°5 - sequência de aminoácidos da repetição 5 de C-LytA
 GAMKTGWVKYKDTWYLDKE 21

SEQ ID N°6 - sequência de aminoácidos da repetição 6 de C-LytA
 GAMVSNAFIQSADGTGWYLLKPD 23

SEQ ID N°7 - sequência de aminoácidos do domínio de ligação à colina de C-LytA

GWQKNDTGYW YVHSDGSYPK DKFEKINGTW YYFDSSGYML ADRWRKHTDG NWYWFDNSGE 60
 MATGWKKIAD KWYFNEEGA MKTGWVKYKD TWYLDKEG AMVSNAFIQS ADGTGWYLLK 120
 PDGTLADRPE FTVEPDGLIT VK 142

SEQ ID N°8 - sequência de aminoácidos do domínio de C-LytA da repetição 1 truncada até à repetição 6 (como parte das construções da invenção mostradas na figura 2)

YVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKT
 GWVKYKDTWYLDKEGAMVSNAFIQSADGTGWYLLKPD

SEQ ID N°9 - sequência de ADN codificando a sequência de aminoácidos de SEQ ID N°:1

ggctggcaga agaattgacac tggctactgg tacgtacatt cagac

SEQ ID N°10 - sequência de ADN codificando a sequência de aminoácidos de SEQ ID N°:2

ggctcttatc caaaagacaa gtttgagaaa atcaatggca cttgggtacta ctttgacagt tca

SEQ ID N°11 - sequência de ADN codificando a sequência de aminoácidos de SEQ ID N°:3

ggctatatgc ttgcagaccg ctggaggaag cacacagacg gcaactggta ctggttcgac aactca

SEQ ID N°12 - sequência de ADN codificando a sequência de aminoácidos de SEQ ID N°:4

ggcgaaatgg ctacaggctg gaagaaaatc gctgataagt ggtactatct caacgaagaa

SEQ ID N°13 - sequência de ADN codificando a sequência de aminoácidos de SEQ ID N°:5

Ggtgccatga agacaggctg ggtcaagtac aaggacactt ggtactactt agacgctaaa gaa

SEQ ID N°14 - sequência de ADN codificando a sequência de aminoácidos de SEQ ID N°:6

Ggcgccatgg tatcaaatgc ctttatccag tcagcggacg gaacaggctg gtactacctc
aaaccagac

SEQ ID N°15 - sequência de ADN codificando a sequência de aminoácidos de SEQ ID N°:7

ggctggcaga agaatgacac tggctactgg tacgtacatt cagacggctc ttatccaaaa 60
gacaagtttg agaaaatcaa tggcacttgg tactactttg acagttcagg ctatatgctt 120
gcagaccgct ggaggaagca cacagacggc aactgggtact ggttcgacaa ctcaggcgaa 180
atggctacag gctggaagaa aatcgctgat aagtgggtact atttcaacga agaaggtgcc 240
atgaagacag gctgggtcaa gtacaaggac acttgggtact acttagacgc taaagaaggc 300
gccatggtat caaatgcctt tatccagtca gcggacggaa caggctggta ctacctcaa 360
ccagacggaa cactggcaga caggccagaa ttcacagtag agccagatgg cttgattaca 420
gtaaaataa 429

SEQ ID N°16 - sequência de ADN codificando a sequência de aminoácidos de SEQ ID N°:8

TACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACA
GTTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGG
CGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACA
GGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGGTATCAAATGCCTTTA
TCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGAC

Fig. 2 - CPC e Construções nativas

Construção 1 - sequência codificante de CPC-P501₅₁₋₅₅₃ (ver plasmídeo da figura 7 - Y1796)

Sequência de proteína (SEQ ID N°:27)

MAAA^{R1}YVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWF^{R2}DNSGEMATG^{R3}
^{R5}WKKIADKWYYFNEEGAMKTGWVVKYKDTWYYLDAKEGA^{P2}MOYIKANSKFIGITEGV^{R6}MVSNAFIQS
ADGTGWYYLKP^{R5}DLADRPEKFMVMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFTWALS
 L GILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSV
 YAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEELFGLLTLIFLTCVAATLLVAEEAALGPTEPAEG
 LSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPTLRRLFVAELCSWMALMTFTLFYTDVFGE
 GLYQGVPRAEPTARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAPFV
 AAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDLSMTSF
 LPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTARVVPGRGICLDLAILDSAF
 LLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIFYATQVVFDDKSDLAKYSAGGHHHHHH

R1 (simples): aa5-9 (fragmento)

R4 (a cheio): aa53-72

P2 (sublinhado): 97-100

R2 (a cheio): aa10-30

R5 (simples): aa73-93

R3 (simples): aa31-52

R6a (a cheio): aa94-95

R6b (a cheio): 113-133

Sequência de nucleótidos (SEQ ID N°:28)

ATGgcgcgcgcctTACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGT
 ACTACTTTGACAGTTTACGGCTATATGCTTGACAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTT
 CGACAACTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGT
 GCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatg^{gcaat}
^{acatcaaggctaactctaagttcattggtatcactgaaggcgctc}ATGGTATCAAATGCCTTTATCCAGTCAGC
 GGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAaagttcatgtaCatg
 GTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTG
 GACGCTATGGCCGCCGCCGCCCTTCATCTGGGCACGTGCTTGGGCATCCTGCTGAGCCTCTTTCTCATCCC
 AAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCCAGGCCCCCTGGAGCTGGCACTGCTCATCCTGGGC
 GTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCAAGTGGAGGCCCTGCTCTTGACCTCTTCCGGG
 ACCCGGACCACTGTGCGCCAGGCCTACTCTGTCTATGCCTTCATGATCAGTCTTGGGGCTGCCTGGGTACCT
 CCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCCCTACCTGGGCACCCAGGAGGAGTGCTCTTTGGC
 CTGCTCACCTCATCTTCTCACCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGCAGCGCTGGGCCCCA
 CCGAGCCAGCAGAAGGGCTGTGCGCCCCCTCCTTGTGCCCCACTGCTGTCCATGCCGGGCCCGCTTGGCTTT

CCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCCGCATGCCCCGCACCCTGCGCCGGCTC
 TTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACCTTCACGCTGTTTTACACGGATTTCGTGGGCGAGG
 GGCTGTACCAGGGCGTGCCAGAGCTGAGCCGGGCACCGAGGCCCGGAGACACTATGATGAAGGCGTTCGGAT
 GGGCAGCCTGGGGCTGTTCTTGCAGTGCGCCATCTCCCTGGTCTTCTCTCTGGTCATGGACCGGCTGGTGCAG
 CGATTCCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGCCGGTGCCACATGCCTGT
 CCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCC
 CTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTTCCCTGCCCAAATACCGAGGGGACACTGGAGGT
 GCTAGCAGTGAGGACAGCCTGATGACCAGCTTCTTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGAC
 ACGTGGGTGCTGGAGGAGTGGCCTGCTCCACCTCCACCCGCGCTCTGCGGGGCTCTGCCTGTGATGTCTC
 CGTACGTGTGGTGGTGGGTGAGCCACCGAGGCCAGGGTGGTTCGGGGCCGGGGCATCTGCCTGGACCTCGCC
 ATCCTGGATAGTGCCTTCTGCTGTCCAGGTGGCCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCC
 AGTCTGTCACTGCCTATATGGTGTCTGCCGAGGCCCTGGGTCTGGTCGCCATTTACTTTGCTACACAGGTAGT
 ATTTGACAAGAGCGACTTGGCCAAATACTCAGCGgtggacaccatcaccatcaccattaa

Construção 2 - Sequência codificante de P501⁵⁵⁻⁵⁵³ HIS (controle)
(estirpe SC333 de levedura)

Sequência de proteína (SEQ ID N°:29)

MVLGIGPVLG LVCVPLLGSA SDHWRGRYGR RRPFIWALS L GILLSLFLIP RAGWLAGLLC 60
 PDPRPLELAL LILGVLLDF CGQVCFTPLE ALLSDLFRDP DHCRQAYS VY AFMISLGGCL 120
 GYLLPAIDWD TSALAPYLGT QEECLFGLLT LIFLTCVAAT LLVAEEAALG PTEPAEGLSA 180
 PSLSPHCCPC RARLAFRNLG ALLPRLHQLC CRMPRTLRL FVAELCSWMA LMTFTLFYTD 240
 FVGEGLYQGV PRAEPGTEAR RHYDEGVRMG SLGLFLQCAI SLVFSLVMDR LVQRFGTRAV 300
 YLASVAAPFV AAGATCLSHS VAVVTASAAL TGFTFSALQI LPYTLASLYH REKQVFLPKY 360
 RGDGTGASSE DSLMTSFLPG PKPGAPFPNG HVGAGSGLL PPPALCGAS ACDVSVRVVV 420
 GEPTARVVP GRGICLDLAI LDSAFLLSQV APSLFMGSIV QLSQSVTAYM VSAAGLGLVA 480
 IYFATQVVF D KSDLAKYSAG GHHHHHH 507

Sequência de nucleótidos (SEQ ID N°:30)

atgGTGCTGG GCATTGGTCC AGTGCTGGGC CTGGTCTGTG TCCCGCTCCT AGGCTCAGCC 60
 AGTGACCACT GGCGTGGACG CTATGGCCGC CGCCGGCCCT TCATCTGGGC ACTGTCCTTG 120
 GGCATCCTGC TGAGCCTCTT TCTCATCCCA AGGGCCGGCT GGCTAGCAGG GCTGCTGTGC 180
 CCGGATCCCA GGCCCCCTGGA GCTGGCACTG CTCATCCTGG GCGTGGGGCT GCTGGACTTC 240
 TGTGGCCAGG TGTGCTTCAC TCCACTGGAG GCCCTGCTCT CTGACCTCTT CCGGGACCCG 300
 GACCACTGTC GCCAGGCCTA CTCTGTCTAT GCCTTCATGA TCAGTCTTGG GGGCTGCCTG 360
 GGCTACCTCC TGCTGCCAT TGA CTGGGAC ACCAGTGCCC TGGCCCCCTA CCTGGGCACC 420
 CAGGAGGAGT GCCTCTTTGG CCTGCTCACC CTCATCTTCC TCACCTGCGT AGCAGCCACA 480
 CTGCTGGTGG CTGAGGAGGC AGCGTGGGC CCCACCGAGC CAGCAGAAGG GCTGTCGGCC 540
 CCCTCCTTGT CGCCCCACTG CTGTCCATGC CGGGCCCGCT TGGCTTTCCG GAACCTGGGC 600

GCCCTGCTTC CCCGGCTGCA CCAGCTGTGC TGCCGCATGC CCCGCACCCT GCGCCGGCTC 660
 TTCGTGGCTG AGCTGTGCAG CTGGATGGCA CTCATGACCT TCACGCTGTT TTACACGGAT 720
 TTCGTGGGCG AGGGGCTGTA CCAGGGCGTG CCCAGAGCTG AGCCGGGCAC CGAGGCCCGG 780
 AGACACTATG ATGAAGGCGT TCGGATGGGC AGCCTGGGGC TGTTCCTGCA GTGCGCCATC 840
 TCCCTGGTCT TCTCTCTGGT CATGGACCGG CTGGTGCAGC GATTCGGCAC TCGAGCAGTC 900
 TATTTGGCCA GTGTGGCAGC TTTCCCTGTG GCTGCCGGTG CCACATGCCT GTCCCACAGT 960
 GTGGCCGTGG TGACAGCTTC AGCCGCCCTC ACCGGGTTCA CCTTCTCAGC CCTGCAGATC 1020
 CTGCCCTACA CACTGGCCTC CCTCTACCAC CGGGAGAAGC AGGTGTTCTT GCCCAAATAC 1080
 CGAGGGGACA CTGGAGGTGC TAGCAGTGAG GACAGCCTGA TGACCAGCTT CCTGCCAGGC 1140
 CCTAAGCCTG GAGCTCCCTT CCCTAATGGA CACGTGGGTG CTGGAGGCAG TGGCCTGCTC 1200
 CCACCTCCAC CCGCGCTCTG CGGGGCCTCT GCCTGTGAtG TCTCCGTACG TGTGGTGGTG 1260
 GGTGAGCCCA CCGAGGCCAG GGTGGTTCCG GGCCGGGGCA TCTGCCTGGA CCTCGCCATC 1320
 CTGGATAGTG CCTTCCTGCT GTCCCAGGTG GCCCATCCC TGTTTATGGG CTCCATTGTC 1380
 CAGCTCAGCC AGTCTGTCAC TGCCTATATG GTGTCTGCCG CAGGCCTGGG TCTGGTCGCC 1440
 ATTTACTTTG CTACACAGGT AGTATTTGAC AAGAGCGACT TGGCCAAATA CTCAGCGggt 1500
 ggacaccatc accatcacca ttaa 1524

Construção 3 - Sequência codificante de natssP501₁₋₃₄ P501₅₅₋₅₅₃ HIS
(estirpe Y1800 de levedura)
Sequência de proteína (SEQ ID N°:31)

MAAVQRLWVSRLLRHRKAQLLLNNLLTFGLEVCLAAAYVHSDGSYPKDKFEKINGTW
 YYFDSSGYMLADRWKRKHTDGNWYWFNDSGEMATGWKKIADKWYYFNEEGAMKTGWVK
 YKDTWYYLDAKEGAMQYIKANSKFITEGV MVSNAFIQSADGTGWYYLKPDGTLADRPEKFMY
 MVLGIGPVLGLVCVPLLGSASDHWGRGRYGRRRPFIWALSGLLSLFLIPRAGWLAGLLCPDPRPLEL
 ALLILGVLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAP
 YLGTQEELFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPR
 LHQLCCRMPTLRRLFVAELCSWMALMTFTLFYTDVFGELYQGVPRAPGTEARRHYDEGVRMG
 SLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAPVAAGATCLSHSVAVVTASAALTGFTFSA
 LQILPYTLASLYHREKQVFLPKYRGDTGGASSEDLSMTSFLPGPKPGAPFPNGHVGAGGSGLPPPPA
 LCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVS
 AAGLGLVAIFYFATQVVFDDKSDLAKYSAGGHHHHHH

R1 (simples): aa38-42 R4 (a cheio): aa77-106 P2 (sublinhado): 130-143
 (fragmento)
 R2 (a cheio): aa43-64 R5 (simples): aa107-126
 R3 (simples): aa65-76 R6a (a cheio): aa127-128 R6b (a cheio): 146-166

natss significa sequência de sinal nativa

Sequência de nucleótidos (SEQ ID N°:32)

ATGgcGGCCGTGCAGAGGCTATGGGTATCGAGACTGCTAAGACACCGCAAAGCTCAGTTGTTGTTGGTTAACT
 TGTTGACCTTCGGGCTGGAAGTCTGTTTGGCGgcccgtTACGTACATTCCGACGGCTCTTATCCAAAAGACAA
 GTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTTCAAGGCTATATGCTTGCAGACCGCTGGAGGAAG
 CACACAGACGGCAACTGGTACTGGTTCGACAACCTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATA
 AGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTT
 AGACGCTAAAGAAGGCGCCatgcaatacatcaaggctaactctaagttcattggtatcactgaaggcgctcATG
 GTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAG
 ACAGGCCAGAAaagttcatgtaCatgGTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGTCCCGCTCCT
 AGGCTCAGCCAGTGACCACTGGCGTGACGCTATGGCCGCCGCCGCCCTTCATCTGGGCACTGTCTTGGGC
 ATCCTGTCTGAGCCTCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGTGTGCCCGGATCCAGGCCCC
 TGGAGCTGGCACTGCTCATCTGGGCGTGGGGCTGTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGA
 GGCCCTGCTCTCTGACCTCTTCCGGGACCCCGGACCACTGTGCGCCAGGCCTACTCTGTCTATGCCTTCATGATC
 AGTCTTGGGGGCTGCTGGGCTACCTCCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCCCTACCTGG
 GCACCCAGGAGGAGTGCCTCTTTGGCCTGCTCACCTCATCTTCCTCACCTGCGTAGCAGCCCACTGCTGGT
 GGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTGCGCCCCCTCCTTGTGCGCCCCACTGC
 TGTCCATGCCGGGCCCGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCC
 GCATGCCCCGCACCTTGCGCCGGCTCTTCGTGGCTGAGCTGTGTCAGCTGGATGGCACTCATGACCTTCACGCT
 GTTTTACACGGATTTCTGGGCGAGGGCTGTACCAGGGCGTGCCAGAGCTGAGCCGGGCACCGAGGCCCGG
 AGACACTATGATGAAGCGTTTCGGATGGGCAGCCTGGGGCTGTTCTCTGCAGTGCGCCATCTCCCTGGTCTTCT
 CTCTGGTCATGGACCGGCTGGTGCAGCGATTCCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCC
 TGTGGCTGCCGGTGCCACATGCCTGTCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTTC
 ACCTTCTCAGCCCTGCAGATCTGCCCTACACACTGGCCTCCCTCTACCACCGGAGAAGCAGGTGTTCTTGC
 CCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAA
 GCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCACTGGCCTGCTCCACCTCCACCGCGCTC
 TGCGGGGCTCTGCCTGTGAtGTCTCCGTACGTGTGGTGGTGGGTGAGCCACCGAGGCCAGGGTGGTTCGG
 GCCGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCTTCTCTGCTGTCCAGGTGGCCCCATCCCTGTT
 TATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGAGGCTGGGTCTGGTC
 GCCATTTACTTTGTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGggtggacaccatc
 accatcaccattaa

Construção 4 - Sequência codificante de alfapreCPC-P501₅₁₋₅₅₃ HIS
(estirpe Y1802 de levedura)

Sequência de proteína (SEQ ID N°:33)

sinal	Alfa-pre	R1	R2	R3
MAARFPSIFTAVLFAASSALAAAY	VHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWF			
R4	R5	P2		
NSGEMATGWKRIADKWYFNEEGAMKTGWVKYKDTWYYLDAKEGAM	MOYIKANSKFIGITEGV	MVSNAPFI		

R6

QSADGTGWYYLKPDGTLADRPEKFMVMVLGIGPVLGLVCVPLLGSASDHWGRYGRRRPFIWALSLGILLSLF
 LIPRAGWLAGLLCPDRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCL
 GYLLPAIDWDTALAPYLGTQECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRAR
 LAFRNLGALLPRLHQLCCRMPTLRRLFVAELCSWMALMTFTLFYTDVFGELYQGVPRAPGTEARRHYDEG
 VRMGSLGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVAAPFVAAGATCLSHSVAVVTASAALTGFTFSALQ
 ILPYTLASLYHREKQVFLPKYRGDTGGASSEDLSMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASAC
 DVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFAT
 QVVFDKSDLAKYSAGGHHHHHH

sinal alfa-pre (a cheio): AA4-22

R1 (simples): aa24-28 R4 (a cheio): aa72-91 P2 (sublinhado): 116-129
 (fragmento)

R2 (a cheio): aa29-49 R5 (simples): aa92-112

R3 (simples): aa50-71 R6a (a cheio): aa113-114 R6b (a cheio): 132-152

Alfapre significa sequência de pré-sinal alfa

Sequência de nucleótidos (SEQ ID N°:34)

TACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACA
 GTTCAGGCTATATGCTTGACAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGG
 CGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACA
 GGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatgcaatacatcaaggcta
actctaagttcatttggtatcactgaaggcgctcATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGG
 CTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAA

ATGgcGGCCAGATTTCTTCAATTTTTACTGCAGTTTTATTTCGCAGCATCTCCGCATTAGCggccgctTACG
 TACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTC
 AGGCTATATGCTTGACAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGGCGAA
 ATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACAGGCT
 GGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatgcaatacatcaaggctaactc
taagttcatttggtatcactgaaggcgctcATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGG
 TACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAgctggtattacttacgttccaccattgttgt
 tggagttggtgttgaagaaaagttcatgtaCatgGTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGT
 CCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCCGCCGCCGCCCTTCATCTGGGCACTG
 TCCTTGGGCATCCTGCTGAGCCTCTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATC
 CCAGGCCCTGGAGCTGGCACTGCTCATCCTGGCGCTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCAC
 TCCACTGGAGGCCCTGCTCTCTGACCTCTCCGGGACCCGGACCACTGTCCGACGGCCTACTCTGTCTATGCT
 TCATGATCAGTCTTGGGGCTGCCTGGGCTACCTCTGCCTGCCATTGACTGGGACACCAGTGGCCTGGCCCC
 CTACCTGGGCACCCAGGAGGAGTGCCTCTTTGGCCTGCTCACCTCATCTTCCTCACCTGCGTAGCAGCCACA

CTGCTGGTGGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTCCGCCCCCTCCTTGTGCG
 CCCACTGCTGTCCATGCCGGGCGCGCTTGGCTTTCGGAACTGGGCGCCCTGCTTCCCCGGCTGCACCAGCT
 GTGCTGCCGCATGCCCCGACCCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACC
 TTCACGCTGTTTTACACGGATTTTCGTGGGCGAGGGGCTGTACCAGGGCGTGCCAGAGCTGAGCCGGGCACCG
 AGGCCCCGAGACACTATGATGAAGGCGTTCGGATGGGCAGCCTGGGGCTGTTCTGCAGTGCGCCATCTCCCT
 GGTCTTCTCTCTGGTCATGGACCGGCTGGTGACGCGATTTCGGCACTCGAGCAGTCTATTTGGCCAGTGTTGGCA
 GCTTTCCCTGTGGCTGCCGGTGCCACATGCCTGTCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCA
 CCGGGTTACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGAGAGAAGCAGGT
 GTTCTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCTCTGCCA
 GGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCACCTCCAC
 CCGCGCTCTGCGGGGCTCTGCCTGTGATGTCCTCGTACGTGTGGTGGTGGGTGAGCCCACCGAGGCCAGGGT
 GGTTCGGGGCCGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCTGCTGTCCAGGTGGCCCCA
 TCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCTGG
 GTCTGGTCGCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGggtgg
 acaccatcaccatcaccattaa

Construção 5 - Sequência codificante de alfaprepro-P501₅₁₋₅₅₃ HIS
 (no plasmídeo pRIT 15068 e estirpe Y1790 de levedura)
 Sequência de proteína (SEQ ID N°:35)

MSFLNFTAVL FAASSALAAP VNTTTEDETA QIPAEAVIGY SDLEGDFDVA VLPFSNSTNN 60
 GLLFINTTIA SIAAKEEGVS LEKREAEAMV LGIGPVLGLV CVPLLGSASD HWRGRYGRRR 120
 PFIWALS LGLI LLSLFLIPRA GWLAGLLCPD PRPLELALLI LGVGLLDPCG QVCFTPLEAL 180
 LSDLFRDPDH CRQAYS VYAF MISLGGCLGY LLP AIDWDT S ALAPYLGTQE ECLFGLLTLI 240
 FLTCVAATLL VAEEAALGPT EPAEGLSAPS LSPHCCPCRA RLAFRNLGAL LPRLHQLCCR 300
 MPRTLRLRFV AELCSWMALM TFTLFYTDV FV GEGLYQGVP R AEPGTEARRH YDEGVRMGSL 360
 GLFLQCAISL VFSLVMDRLV QRF GTRAVYL ASVAAFPVAA GATCLSHSVA VVTASAALTG 420
 FTFSALQILP YTLASLYHRE KQVFLPKYRG DTGGASSEDS LMTSFLPGPK PGAPFPNGHV 480
 GAGGSGLLPP PPALCGASAC DVSVRVVVGE PTEARVVPGR GICLDLAILD SAFLLSQVAP 540
 SLFMGSIVQL SQSVTAYMVS AAGLGLVAIY FATQVVFDKS DLAKYSAGGH HHHHH 595

Sequência de nucleótidos (SEQ ID N°:36)

ATGAGTTTCC TCAATTTTAC TGCAGTTTAA TTCGCAGCAT CCTCCGCATT AGCTGCTCCA 60
 GTCAACACTA CAACAGAAGA TGAAACGGCA CAAATTCCGG CTGAAGCTGT CATCGGTTAC 120
 TCAGATTTAG AAGGGGATTT CGATGTTGCT GTTTTGCCAT TTTCCAACAG CACAAATAAC 180
 GGGTTATTGT TTATAAATAC TACTATTGCC AGCATTGCTG CTAAAGAAGA AGGGGTATCT 240
 CTCGAGAAAA GAGAGGCTGA AGCCatgGTG CTGGGCATTG GTCCAGTGCT GGGCCTGGTC 300
 TGTGTCCCGC TCCTAGGCTC AGCCAGTGAC CACTGGCGTG GACGCTATGG CCGCCGCCGG 360

CCCTTCATCT GGGCACTGTC CTTGGGCATC CTGCTGAGCC TCTTTCTCAT CCCAAGGGCC 420
 GGCTGGCTAG CAGGGCTGCT GTGCCCCGAT CCCAGGCCCC TGGAGCTGGC ACTGCTCATC 480
 CTGGGCGTGG GGCTGCTGGA CTTCTGTGGC CAGGTGTGCT TCACTCCACT GGAGGCCCTG 540
 CTCTCTGACC TCTTCCGGGA CCCGGACCAC TGTCCGCCAGG CCTACTCTGT CTATGCCTTC 600
 ATGATCAGTC TTGGGGGCTG CCTGGGCTAC CTCTGCCTG CCATTGACTG GGACACCACT 660
 GCCCTGGCCC CCTACCTGGG CACCCAGGAG GAGTGCCTCT TTGGCCTGCT CACCCTCATC 720
 TTCCTCACCT GCGTAGCAGC CACACTGCTG GTGGCTGAGG AGGCAGCGCT GGGCCCCACC 780
 GAGCCAGCAG AAGGGGTGTC GGCCCCCTCC TTGTCGCCCC ACTGCTGTCC ATGCCGGGCC 840
 CGCTTGGCTT TCCGGAACCT GGGCGCCCTG CTTCCCCGGC TGCACCAGCT GTGCTGCCGC 900
 ATGCCCCGCA CCCTGCGCCG GCTCTTCGTG GCTGAGCTGT GCAGCTGGAT GGCACTCATG 960
 ACCTTCACGC TGTTTACAC GGATTTCGTG GGCAGAGGGC TGTACCAGGG CGTGCCCAGA 1020
 GCTGAGCCGG GCACCGAGGC CCGGAGACAC TATGATGAAG GCGTTCGGAT GGGCAGCCTG 1080
 GGGCTGTTCC TGCAGTGCGC CATCTCCCTG GTCTTCTCTC TGGTCATGGA CCGGCTGGTG 1140
 CAGCGATTCTG GCACTCGAGC AGTCTATTTG GCCAGTGTGG CAGCTTTCCC TGTGGCTGCC 1200
 GGTGCCACAT GCCTGTCCCA CAGTGTGGCC GTGGTGACAG CTTCAGCCGC CCTCACCAGG 1260
 TTCACCTTCT CAGCCCTGCA GATCCTGCCC TACACACTGG CCTCCCTCTA CCACCGGGAG 1320
 AAGCAGGTGT TCCTGCCCAA ATACCGAGGG GACACTGGAG GTGCTAGCAG TGAGGACAGC 1380
 CTGATGACCA GCTTCCTGCC AGGCCCTAAG CCTGGAGCTC CCTTCCCTAA TGGACACGTG 1440
 GGTGCTGGAG GCAGTGGCCT GCTCCACCT CCACCCGCGC TCTGCGGGGC CTCTGCCTGT 1500
 GAtGTCTCCG TACGTGTGGT GGTGGGTGAG CCCACCGAGG CCAGGGTGGT TCCGGGCCGG 1560
 GGCATCTGCC TGGACCTCGC CATCCTGGAT AGTGCCTTCC TGCTGTCCCA GGTGGCCCCA 1620
 TCCCTGTTTA TGGGTCTCAT TGTCCAGCTC AGCCAGTCTG TCACTGCCTA TATGGTGTCT 1680
 GCCGAGGCC TGGGTCTGGT CGCCATTTAC TTTGTACAC AGGTAGTATT TGACAAGAGC 1740
 GACTTGGCCA AATACTCAGC Gggtggacac catcaccatc accattaa 1788

Fig. 3 - Estrutura da proteína de fusão CPC-p501 His expressa em *S. cerevisiae*

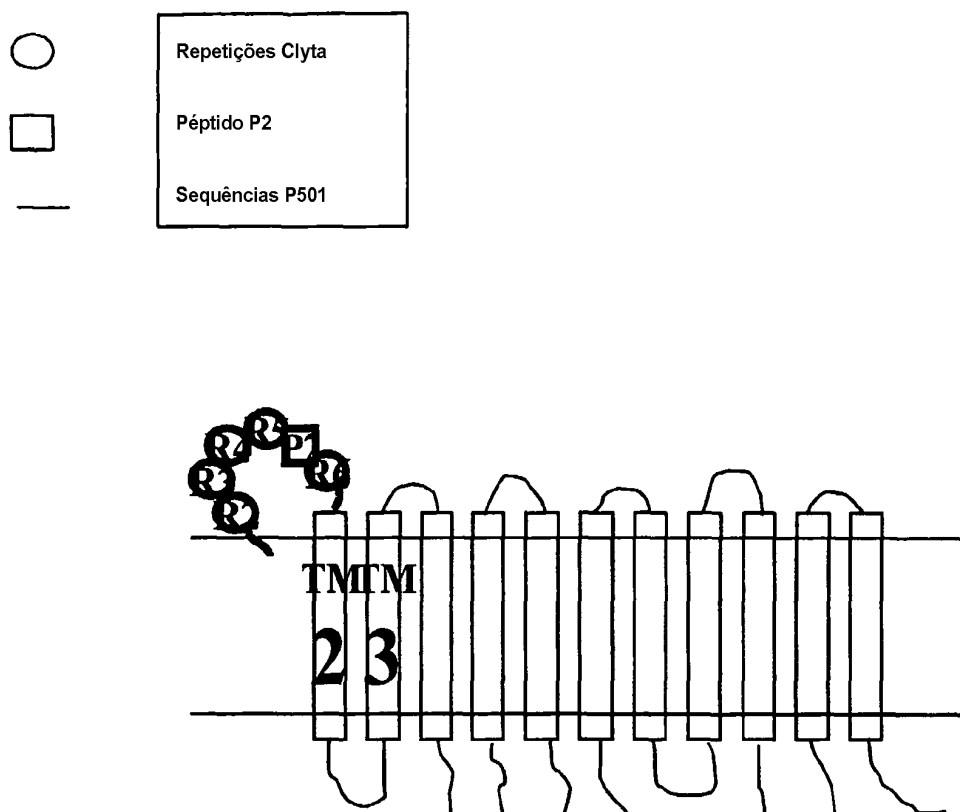


**Fig. 4 - Estrutura primária da proteína de fusão CPC-P501 His
(SEQ ID N° 41)**

```

MAAAYVHSDG SYPKDKFEKI NGTWYYFDSS GYMLADRWRK HTDGNWYWFD NSGEMATGWK 60
KIADKWYYFN EEGAMKTGWV KYKDTWYYLD AKEGAMQYIK ANSKFIGITE GVMVSNAFIQ 120
SADGTGWYYL KPDGTLADRP EKFMVMVLGI GPVLGLVCVP LLGSASDHWR GRYGRRRPFI 180
WALSLGILLS LFLIPRAGWL AGLLCPDPRP LELALLILGV GLLDFCGQVC FTPLEALLSD 240
LFRDPDHCRQ AYSVYAFMIS LGGCLGYLLP AIDWDTSALA PYLGTQEECL FGLLTILFLT 300
CVAATLLVAE EAALGPTEPA EGLSAPSLSP HCCPCRARLA FRNLGALLPR LHQLCCRMPR 360
TLRRLFVAEL CSWMALMTFT LFYTDFVGEG LYQGVPRAP GTEARRHYDE GVRMGSLGLF 420
LQCAISLVFS LVMDRLVQRF GTRAVYLASV AAFPVAAGAT CLSHSVAVVT ASAALTGFTF 480
SALQILPYTL ASLYHREKQV FLPKYRGDTG GASSEDSLMT SFLPGPKPGA PFPNGHVGAG 540
GSGLLPPPPA LCGASACDVS VRVVVGEPT ARVVPGRGIC LDLAILDSAF LLSQVAPSLF 600
MGSIVQLSQS VTAYMVSAA GGLVAIYFAT QVVFDKSDLA KYSAGGHHHH HH 652

```

Fig. 5 - Sequência de nucleótidos de CPC-P501 His (pRIT15201)
(SEQ ID N° 42)

```

ATGGCGGCCG CTTACGTACA TTCCGACGGC TCTTATCCAA AAGACAAGTT TGAGAAAATC 60
AATGGCACTT GGTACTACTT TGACAGTTCA GGCTATATGC TTGCAGACCG CTGGAGGAAG 120
CACACAGACG GCAACTGGTA CTGGTTCGAC AACTCAGGCG AAATGGCTAC AGGCTGGAAG 180
AAAATCGCTG ATAAGTGGTA CTATTCAAC GAAGAAGGTG CCATGAAGAC AGGCTGGGTC 240
AAGTACAAGG ACACTTGGTA CTACTTAGAC GCTAAAGAAG GCGCCATGCA ATACATCAAG 300
GCTAACTCTA AGTTTCATTGG TACTACTGAA GGCGTCATGG TATCAAATGC CTTTATCCAG 360
TCAGCGGACG GAACAGGCTG GTACTACCTC AAACCAGACG GAACACTGGC AGACAGGCCA 420
GAAAAGTTCA TGTACATGGT GCTGGGCATT GGTCCAGTGC TGGGCCTGGT CTGTGTCCCG 480
CTCCTAGGCT CAGCCAGTGA CCACTGGCGT GGACGCTATG GCCGCCGCCG GCCCTTCATC 540
TGGGCACTGT CCTTGGGCAT CCTGCTGAGC CTCTTTCTCA TCCAAGGGC CGGCTGGCTA 600
GCAGGGCTGC TGTGCCCCGA TCCCAGGCCC CTGGAGCTGG CACTGCTCAT CCTGGGCGTG 660
GGGCTGCTGG ACTTCTGTGG CCAGGTGTGC TCACTCCAC TGGAGGCCCT GCTCTCTGAC 720
CTCTTCCGGG ACCCGGACCA CTGTCGCCAG GCCTACTCTG TCTATGCCTT CATGATCAGT 780
CTTGGGGGCT GCCTGGGCTA CCTCCTGCCT GCCATTGACT GGGACACCAG TGCCCTGGCC 840
CCCTACCTGG GCACCCAGGA GGAGTGCCTC TTTGGCCTGC TCACCCTCAT CTTCTCACC 900
TGCGTAGCAG CCACACTGCT GGTGGCTGAG GAGGCAGCGC TGGGCCCCAC CGAGCCAGCA 960
GAAGGGCTGT CGGCCCCCTC CTTGTCGCCC CACTGCTGTC CATGCCGGGC CCGCTTGGCT 1020
TTCCGGAACC TGGGCGCCCT GCTTCCCCGG CTGCACCAGC TGTGCTGCCG CATGCCCGC 1080
ACCCTGCGCC GGCTCTTCGT GGCTGAGCTG TGCAGCTGGA TGGCACTCAT GACCTTCACG 1140
CTGTTTTTACA CGGATTTCTG GGGCGAGGGG CTGTACCAGG GCGTGCCAG AGCTGAGCCG 1200
GGCACCAGAG CCCGAGACA CTATGATGAA GCGTTCGGA TGGGCAGCCT GGGGCTGTTC 1260
CTGCAGTGCG CCATCTCCCT GGTCTTCTCT CTGGTCATGG ACCGGCTGGT GCAGCGATTTC 1320
GGCACTCGAG CAGTCTATTT GGCCAGTGTG GCAGCTTTCC CTGTGGCTGC CGGTGCCACA 1380
TGCCTGTCCC ACAGTGTGGC CGTGGTGACA GCTTCAGCCG CCCTCACCAG GTTCACCTTC 1440
TCAGCCCTGC AGATCCTGCC CTACACACTG GCCTCCCTCT ACCACCGGA GAAGCAGGTG 1500
TTCCTGCCCA AATACCGAGG GGACACTGGA GGTGCTAGCA GTGAGGACAG CCTGATGACC 1560
AGCTTCCTGC CAGGCCCTAA GCCTGGAGCT CCCTTCCCTA ATGGACACGT GGGTGTCTGA 1620
GGCAGTGGCC TGCTCCACAC TCCACCCGCG CTCTGCGGGG CCTCTGCCTG TGATGTCTCC 1680
GTACGTGTGG TGGTGGGTGA GCCACCCGAG GCCAGGGTGG TTCCGGGCCG GGGCATCTGC 1740
CTGGACCTCG CCATCCTGGA TAGTGCTTTC CTGCTGTCCC AGGTGGCCCC ATCCCTGTTT 1800
ATGGGCTCCA TTGTCCAGCT CAGCCAGTCT GTCACTGCCT ATATGGTGTC TGCCGCAGGC 1860
CTGGGTCTGG TCGCCATTTA CTTTGCTACA CAGGTAGTAT TTGACAAGAG CGACTTGGCC 1920
AAATACTCAG CGGGTGGACA CCATCACCAT CACCATTAA 1959

```

Fig. 6 - Estratégia de clonagem para produção do plasmídeo pRIT 15201

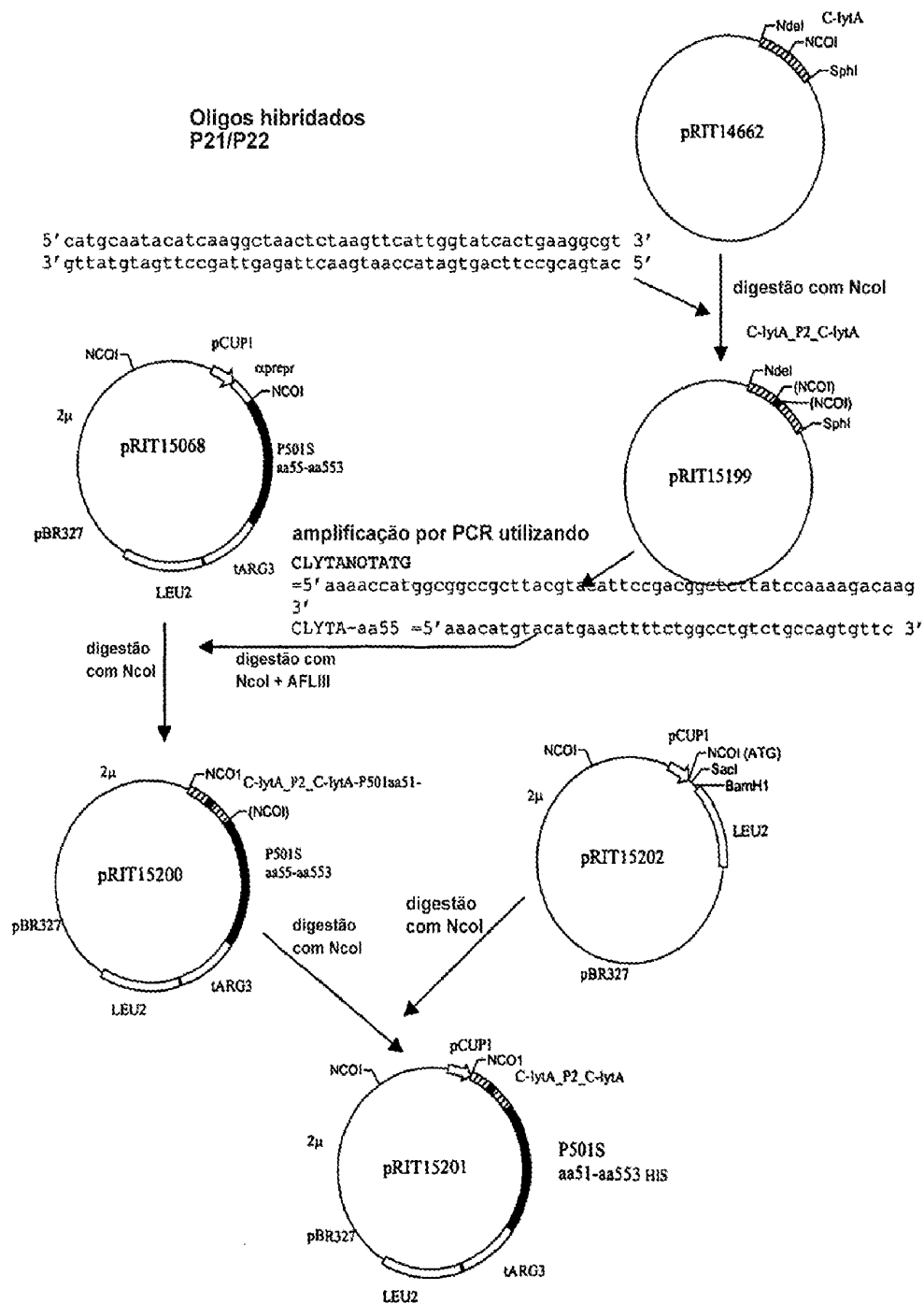


Fig. 7 - Mapa do plasmídeo pRIT 15201

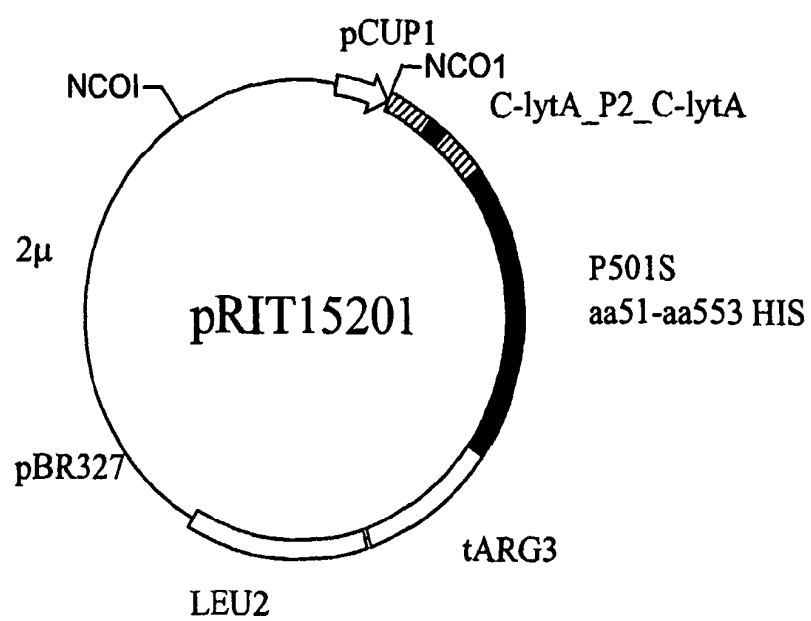
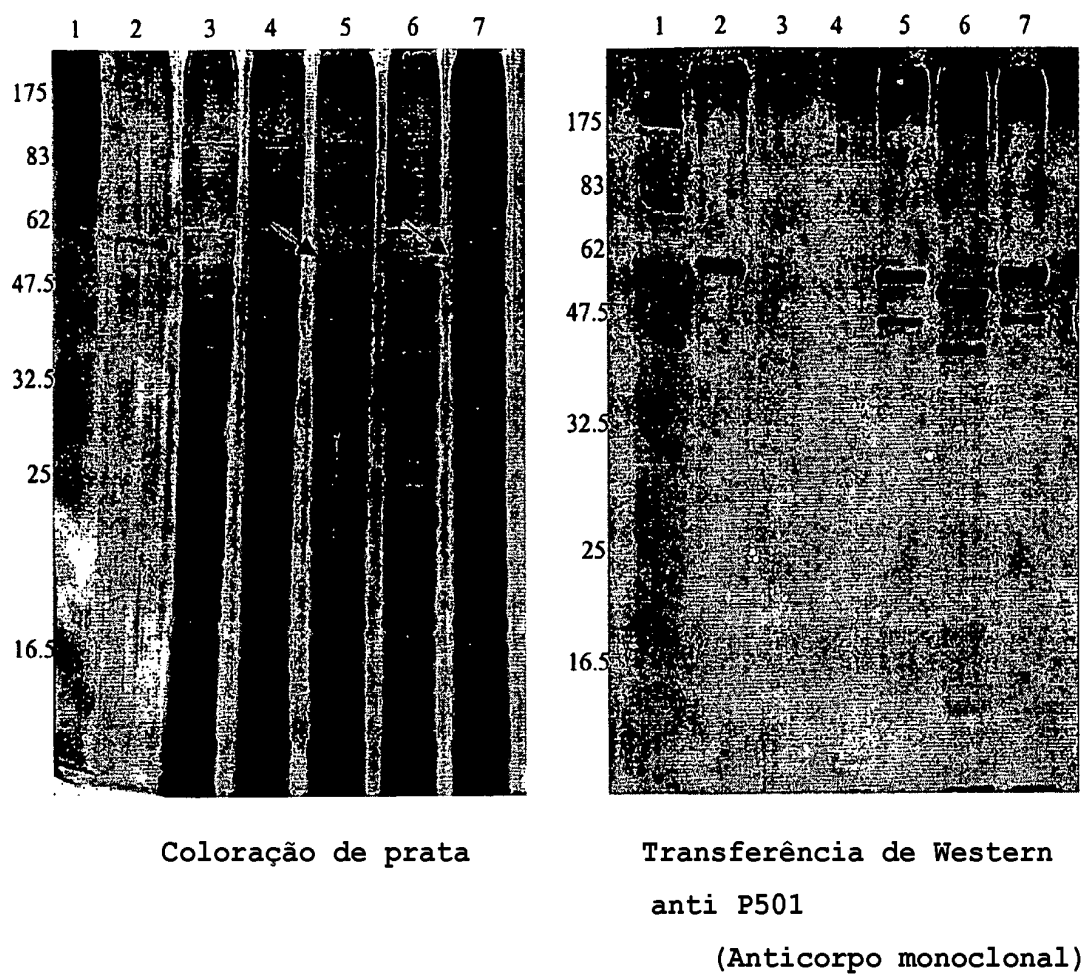


Fig. 8 - Expressão comparativa de CPC P501 e P501 na estirpe DC% de *S. Cerivisiae* (gel de Laemmli a 10%)



- 1 Marcadores Biolabs (175/83/62/47,5/32,5/16,5 Kda)
- 2 Y1796 purificada
- 3 Y1795 Extracto Bruto (controlo negativo)
- 4 SC333 Extracto Bruto
- 5 Y1796 Extracto Bruto
- 6 Y1790 Extracto Bruto
- 7 Y1802 Extracto Bruto

Fig. 9A.

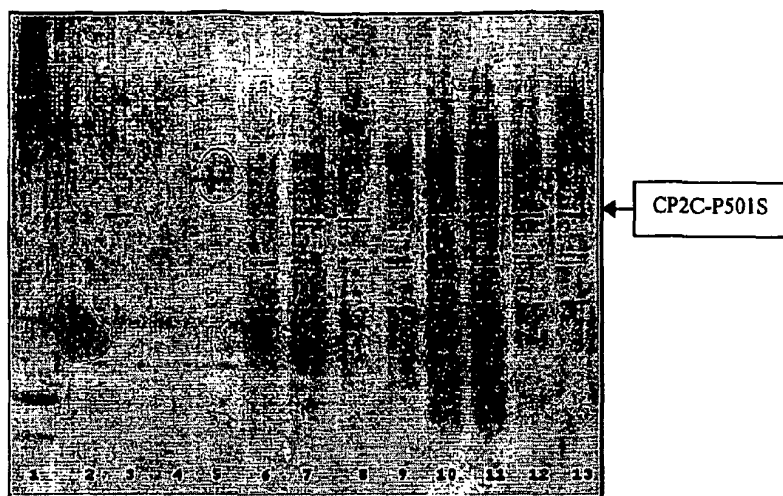
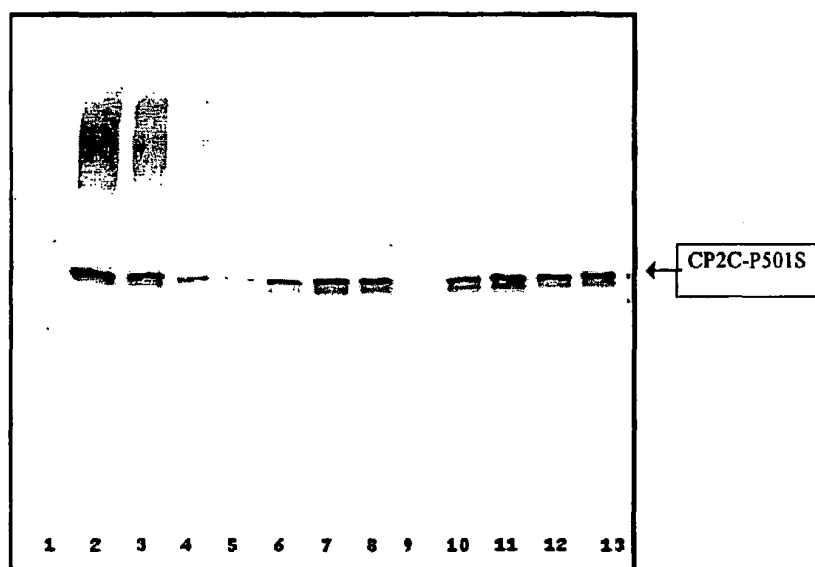


Fig. 9B.



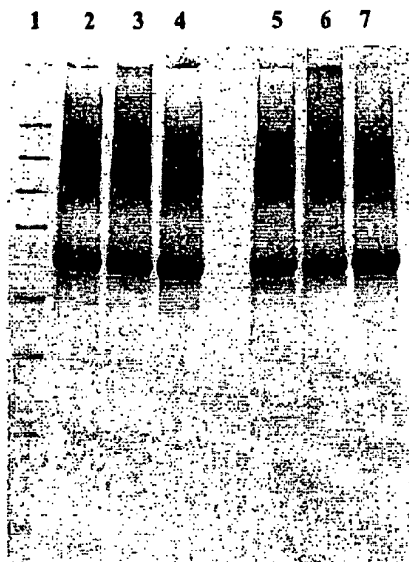
- 1 - Marcadores de Peso Molecular (Biolabs - Grow Range) 175; 83; 62; 47,5; 32,5; 25; 16,5; 6,5 Kd - 10
- 2 - Referência purificada CP2CP501S/12 135 ng
- 3 - Referência purificada CP2CP501S/12 67,8 ng
- 4 - Referência purificada CP2CP501S/12 33,9 ng
- 5 - Referência purificada CP2CP501S/12 16,9 ng
- 6 - Fermentação PRO119-21h30
- 7 - Fermentação PRO124-21h30
- 8 - Fermentação PRO124-22h30
- 9 - Fermentação PRO127-0 h
- 10 - Fermentação PRO127-4 h
- 11 - Fermentação PRO127-6 h
- 12 - Fermentação PRO127-22h20
- 13 - Fermentação PRO127-22h45

Fig. 10. Esquema de purificação de CPC-P501-His produzido por Y1796.

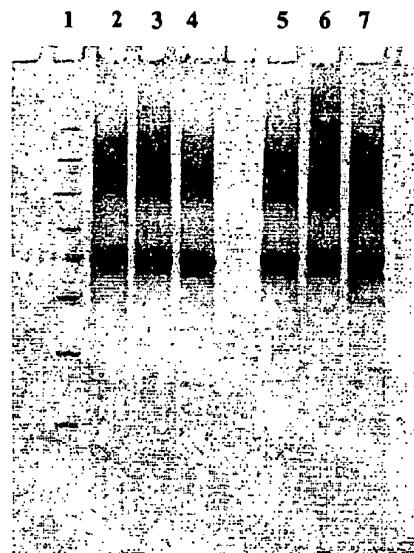
Células de <i>S. cerevisiae</i>	
↓	
Ruptura com Dyno-mill	OD 120/2 passagens/20 mM Tris pH 8,5 - 5 mM EDTA
↓	
Centrifugação	12.000 g /t.a./90 min (sobrenadante rejeitado)
↓	
Passo 1 de lavagem de sedimento	20 mM Tris pH 8,5 - 0,15 M NaCl - 2,0 M Guanidina.HCl - 0,1% de Empigen (30 min/t.a.)
↓	
Centrifugação	12.000 g /t.a./60 min (sobrenadante rejeitado)
↓	
Passo 2 de lavagem de sedimento	20 mM Tris pH 8,5 - 0,15 M NaCl - 4,0 M Ureia
↓	
Centrifugação	12.000g /t.a./30 min (sobrenadante rejeitado)
↓	
Solubilização/redução	20 mM Tris pH 8,5 - 0,15 M NaCl - 8,0 M Ureia - 1% SDS - 0,2 M Glutathione (60 min/t.a.)
↓	
Centrifugação	12.000 g /t.a./30 min (sobrenadante rejeitado)
↓	
Carbamidometilação	0,3 M Iodoacetamida (30 min/t.a./no escuro)/pH ajustado a 8,5 (com solução de NaOH a 5 M) antes da incubação
↓	
Sobrenadante R/C	
↓	
Diluição de 10 vezes e Ajuste de pH (8,5)	<u>Tampão de diluição:</u> 20 mM Tris pH 8,5 - 1 M NaCl - 8,0 M Ureia
↓	
Cromatografia de afinidade de ião metálico imobilizado em Ni⁺⁺- Sepharose FF quelante (Amersham) (coluna de 10x25 cm - 2000 mL)	<u>Tampão de equilíbrio:</u> 20 mM Tris pH 8,5 - 0,9 M NaCl - 8,0 M Ureia - 0,1% SDS <u>Tampões de lavagem:</u> 1) Tampão de equilíbrio 2) 20 mM Tris pH 8,5 - 0,15 M NaCl - 8,0 M Ureia - 0,1% SDS 3) 20 mM Tris pH 8,5 - 8,0 M Ureia - 0,1% Tween 80

	<u>Tampão de eluição:</u> 20 mM Tris pH 8,5 - 8,0 M Ureia - 0,1% Tween 80 - 0,5 M Imidazole
↓	
Diluição de 2 vezes e Ajuste de pH (10,0)	20 mM Piperazina pH 10,0 - 8,0 M Ureia - 0,1% Tween 80
↓	
Cromatografia de permuta aniônica em Q Sepharose FF (Amersham) (coluna de 2,6x6,5 cm - 35 mL)	<u>Tampão de equilíbrio:</u> 20 mM Piperazina pH 10,0 - 8,0 M Ureia - 0,1% Tween 80 <u>Tampões de lavagem:</u> 1) Tampão de equilíbrio 2) 20 mM Tris pH 8,5 - 8,0 M Ureia - 0,1% Tween 80 <u>Tampão de eluição:</u> 20 mM Tris pH 7,5 - 8,0 M Ureia - 0,1% Tween 80 - 0,5 M NaCl
↓	
Concentração/Diafiltração o (Pall-Omega 10 kDa - 200 cm²)	+/- concentração de 3 vezes <u>Tampão de diafiltração:</u> 20 mM Tris pH 7,5
↓	
Filtração estéril (Millipore - Millex GV 0,22 µm)	
↓	
Quantidade purificada	<u>Tampão final:</u> 20 mM Tris pH 7,5 - +/- 0,3% Tween 80
↓	
Armazenamento a -20 °C	

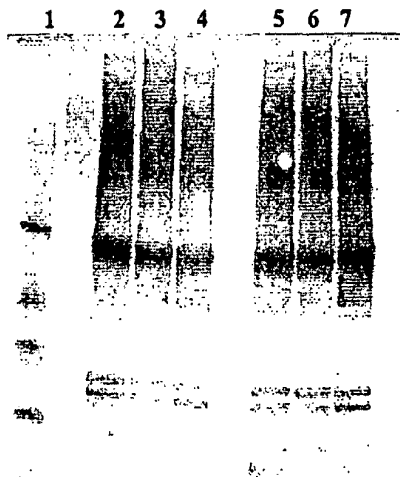
Fig. 11. Perfil da proteína purificada de CPC P501 His (géis pré-fabricados de 4-12% de poliacrilamida Novex Nu-Page).



Azul de Coomassie R250



Coloração de prata de Daiichi



Transferência de Western

Anti P501

(Anticorpo monoclonal)

- 1: Marcadores (250/150/75/50/37/25/15/10 KDa)
- 2: Quantidade A purificada (condições redutoras)
- 3: Quantidade B purificada (condições redutoras)
- 4: Quantidade C purificada (condições redutoras)
- 5: Quantidade A purificada (condições não redutoras)
- 6: Quantidade B purificada (condições não redutoras)
- 7: Quantidade C purificada (condições não redutoras)

Fig. 12. Sequência de P501S de tamanho total nativa (SEQ ID N°:17 e 43)

Sequência de nucleótidos: SEQ ID N° 17

Sequência polipeptídica: SEQ ID N° 43

```

#####
GCCACCATGGTCCAGAGGCTGTGGGTGAGCCGCTGCTGCGGCACCGG
      M V Q R L W V S R L L R H R   14

AAAGCCCAGCTCTTGCTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTTGGCCGCA
K A Q L L L V N L L T F G L E V C L A A   34

GGCATCACCTATGTGCCGCTCTGCTGCTGGAAGTGGGGGTAGAGGAGAAGTTCATGACC
G I T Y V P P L L L E V G V E E K F M T   54

ATGGTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGTCCCGCTCCTAGGCTCAGCC
M V L G I G P V L G L V C V P L L G S A   74

AGTGACCACTGGCGTGGACGCTATGGCCGCCGCCGCCCTTCATCTGGGCACTGTCTTGTG
S D H W R G R Y G R R R P F I W A L S L   94

GGCATCCTGCTGAGCCTCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGC
G I L L S L F L I P R A G W L A G L L C   114

CCGGATCCCAGGCCCTGGAGCTGGCACTGCTCATCCTGGGCGTGGGGCTGCTGGACTTC
P D P R P L E L A L L I L G V G L L D F   134

TGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCTGCTCTCTGACCTCTTCCGGGACCCG
C G Q V C F T P L E A L L S D L F R D P   154

GACCACTGTGCCAGGCCTACTCTGTCTATGCCTTCATGATCAGTCTTGGGGGCTGCCTG
D H C R Q A Y S V Y A F M I S L G G C L   174

GGCTACCTCCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCCCTACCTGGGCACC
G Y L L P A I D W D T S A L A P Y L G T   194

CAGGAGGAGTGCTCTTTGGCCTGCTCACCTCATCTTCTCACCTGCGTAGCAGCCACA
Q E E C L F G L L T L I F L T C V A A T   214

CTGCTGGTGGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTGGGCC
L L V A E E A A L G P T E P A E G L S A   234

CCCTCCTTGTGCCCCACTGCTGTCCATGCCGGGCCCGCTTGGCTTTCCGGAACCTGGGC
P S L S P H C C P C R A R L A F R N L G   254

GCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCCGATGCCCGCACCTGCGCCGGGCTC
A L L P R L H Q L C C R M P R T L R R L   274

TTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACCTTCACGCTGTTTACACGGAT
F V A E L C S W M A L M T F T L F Y T D   294

TTCGTGGGCGAGGGGCTGTACCAGGGCGTGCCAGAGCTGAGCCGGGCACCGAGGCCCGG
F V G E G L Y Q G V P R A E P G T E A R   314

AGACACTATGATGAAGGCGTTCGGATGGGCAGCCTGGGGCTGTTCTGCACTGCGCCATC
R H Y D E G V R M G S L G L F L Q C A I   334

TCCCTGGTCTTCTCTCTGGTCATGGACCGCTGGTGCAGCGATTGGCACTCGAGCAGTC
S L V F S L V M D R L V Q R F G T R A V   354

```

TATTTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCCGGTGCCACATGCCTGTCCCACAGT
Y L A S V A A F P V A A G A T C L S H S 374

GTGGCCGTGGTGACAGCTTCAGCCGCCCTCACCGGTTACCTTCTCAGCCCTGCAGATC
V A V V T A S A A L T G F T F S A L Q I 394

CTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCTGCCCAAATAC
L P Y T L A S L Y H R E K Q V F L P K Y 414

CGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGC
R G D T G G A S S E D S L M T S F L P G 434

CCTAAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTC
P K P G A P F P N G H V G A G G S G L L 454

CCACCTCCACCCGCGCTCTGCGGGGCCTCTGCCTGTGATGTCTCCGTACGTGTGGTGGTG
P P P P A L C G A S A C D V S V R V V V 474

GGTGAGCCCACCGAGGCCAGGGTGGTTCCGGGCCGGGCATCTGCCTGGACCTCGCCATC
G E P T E A R V V P G R G I C L D L A I 494

CTGGATAGTGCTTCTCTGCTGTCCCAGGTGGCCCCATCCCTGTTTATGGGCTCCATTGTC
L D S A F L L S Q V A P S L F M G S I V 514

CAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCCTGGGTCTGGTCGCC
Q L S Q S V T A Y M V S A A G L G L V A 534

ATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAG
I Y F A T Q V V F D K S D L A K Y S A * 554

GTCTGAG

Fig. 13. Sequência da cassete de expressão de CPC-P501S de JNW735 (SEQ ID N°:18 e 44)

Sequência de nucleótidos: SEQ ID N° 18

Sequência polipeptídica: SEQ ID N° 44

```

#####
GCCACCATGGCGGCCGCTTACGTACATTCCGACGGCTCTTATCCAAAA
      M A A A Y V H S D G S Y P K 14

GACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTCAGGCTATATGCTT
D K F E K I N G T W Y Y F D S S G Y M L 34

GCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACCTCAGGCGAA
A D R W R K H T D G N W Y W F D N S G E 54

ATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCC
M A T G W K K I A D K W Y Y F N E E G A 74

ATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGC
M K T G W V K Y K D T W Y Y L D A K E G 94

GCCATGCAATACATCAAGGCTAACTCTAAGTTCATTGGTATCACTGAAGGCGTCATGGTA
A M Q Y I K A N S K F I G I T E G V M V 114

TCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGACGGA
S N A F I Q S A D G T G W Y Y L K P D G 134

ACACTGGCAGACAGGCCAGAAAAGTTCATGTACATGGTGTGGGCATTGGTCCAGTGCTG
T L A D R P E K F M Y M V L G I G P V L 154

GGCCTGTGTGTGTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGC
G L V C V P L L G S A S D H W R G R Y G 174

CGCCGCCGGCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCTCTTCTCATC
R R R P F I W A L S L G I L L S L F L I 194

CCAAGGCGCGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCAGGCCCTGGAGCTGGCA
P R A G W L A G L L C P D P R P L E L A 214

CTGCTCATCCTGGGCGTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTG
L L I L G V G L L D F C G Q V C F T P L 234

GAGGCCCTGCTCTCTGACCTCTTCCGGGACCCGGACCACTGTGCCAGGCCCTACTCTGTC
E A L L S D L F R D P D H C R Q A Y S V 254

TATGCCTTCATGATCAGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGG
Y A F M I S L G G C L G Y L L P A I D W 274

GACACCAGTGCCCTGGCCCCCTACCTGGGCACCCAGGAGGAGTGCCTCTTTGGCCTGCTC
D T S A L A P Y L G T Q E E C L F G L L 294

ACCTCATCTTCTCCTCCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGCAGCGCTG
T L I F L T C V A A T L L V A E E A A L 314

GGCCCCACCGAGCCAGCAGAAGGGCTGTGGCCCCCTCCTTGTGCCCCACTGCTGTCCA
G P T E P A E G L S A P S L S P H C C P 334

```

TGCCGGGCCCCGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTG
 C R A R L A F R N L G A L L P R L H Q L 354

TGCTGCCGCATGCCCCGCACCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATG
 C C R M P R T L R R L F V A E L C S W M 374

GCACTCATGACCTTCACGCTGTTTACACGGATTTCGTGGGCGAGGGGCTGTACCAGGGC
 A L M T F T L F Y T D F V G E G L Y Q G 394

GTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGGAGACACTATGATGAAGGCGTTCGGATG
 V P R A E P G T E A R R H Y D E G V R M 414

GGCAGCCTGGGGCTGTTCCCTGCAGTGCGCCATCTCCCTGGTCTTCTCTCTGGTCATGGAC
 G S L G L F L Q C A I S L V F S L V M D 434

CGGCTGGTGCAGCGATTCCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTCCCT
 R L V Q R F G T R A V Y L A S V A A F P 454

GTGGCTGCCGGTGCCACATGCCTGTCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCC
 V A A G A T C L S H S V A V V T A S A A 474

CTCACCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTAC
 L T G F T F S A L Q I L P Y T L A S L Y 494

CACCGGGAGAAGCAGGTGTTCCGCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGT
 H R E K Q V F L P K Y R G D T G G A S S 514

GAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAAT
 E D S L M T S F L P G P K P G A P F P N 534

GGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCACCTCCACCCGCGCTCTGCGGGGCC
 G H V G A G G S G L L P P P P A L C G A 554

TCTGCCTGTGATGTCTCCGTACGTGTGGTGGTGGGTGAGCCCACCGAGGCCAGGGTGGTT
 S A C D V S V R V V V G E P T E A R V V 574

CCGGGCCGGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCTGCTGTCCAG
 P G R G I C L D L A I L D S A F L L S Q 594

GTGGCCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTAT
 V A P S L F M G S I V Q L S Q S V T A Y 614

ATGGTGTCTGCCGCAGGCCTGGGTCTGGTCGCCATTTACTTTGCTACACAGGTAGTATTT
 M V S A A G L G L V A I Y F A T Q V V F 634

GACAAGAGCGACTTGGCCAAATACTCAGCGTAGGTTCGAG
 D K S D L A K Y S A * 645

Fig. 14. Sequências P501S de codão otimizado (SEQ ID N°:19-20)

SEQ ID N°:19

ATGGTGCAGCGGCTCTGGGTGAGCCGCTCCTGCGGCATCGCAAGGCCAGCTCCTGCTGGTGAATCTGCTCA
 CATTTCGGCCTGGAGGTGTGCCTGGCCGCCGGCATCACCTACGTGCCCCCTCCTGCTGGAGGTGGGAGTCTGA
 GGAGAAGTTCATGACCATGGTGTGGGCATTGGGCCGCTCCTGGGCCTCGTGTGCGTGCCTCTCCTCGGCAGC
 GCTTCCGACCATTTGGCGGGCCGGTATGGCCGCAGGAGACCTTCATCTGGGCTCTGAGTCTCGGCATCCTGC
 TGAGCCTGTTCCTGATCCCTCGGGCCGGCTGGCTGGCCGGGCTGCTGTGCCCCGATCCTCGGCCCTGGAGCT
 GGCCCTGTGTATCCTCGGCGTGGGCTGTGGACTTCTGCGGCCAGGTGTGCTTCACGCCCCTGGAGGCACTG
 CTGAGCGACCTGTTCCGGGACCCGACCATTTGCCGCCAGGCGTACAGCGTGTACGCTTCATGATCTCCCTGG
 GAGGCTGCCTGGGCTACCTGCTCCCCGCCATCGATTGGGACACCAGCGCACTCGCCCCCTATCTCGGAACACA
 GGAGGAATGCCTGTTCCGATTGTTGACGCTCATCTTCCTCACGTGCGTCGCGGCCACCCTGTTGGTGGCCGAG
 GAGGCCGCCCTGGGGCCACCAGCGCGGCCGAGGACTGAGCGCCCCGAGCCTGAGTCCACACTGCTGCCCTT
 GCCGGGCCCGCTGGCCTTCCGTAATCTGGGCGCCCTCCTGCCTCGGCTCCATCAGCTGTGTTGCAGAATGCC
 TAGGACGCTGCGGCGCCTGTTCTGCTCGCTGAGTTGTGCTCCTGGATGGCTCTCATGACCTTCACCTGTTTAT
 ACGGACTTCGTCGGGGAGGGCCTGTACCAGGGGGTGGCGCGCGCCGAGCCCGGGACAGAGGCGCGGCCCACT
 ACGACGAGGGAGTGCGTATGGGCTCCCTGGGCTCTTCTTGCACTGCGCCATCAGTCTGGTTTTCTCTCTGGT
 CATGGACAGGCTGGTGCAGCGCTTCGGAACCCGGGCGGTGTACCTGGCGAGCGTGGCCGCTTCCCCGTGGCT
 GCCGGCGCCACCTGCCTCTCTCACTCGGTGGCCGTGGTACCGCCAGCGCCGCCCTGACCGGGTTACCTTCT
 CTGCCCTGCAGATTCTGCCTTACACCTGGCCAGCCTGTACCATCGCGAGAAACAGGTGTTTCTCCCCAAGTA
 CAGAGGCGACACCGGGGGCGCTCCAGCGAGGACAGCCTCATGACCTCCTTCTGCTGGCCCCAAGCCCCGC
 GCCCCCTTCCCCAACGGGCACGTGGGCGCCGGCGGGAGTGGGCTCCTGCCCCCCCCTCCTGCGCTGTGCGGGG
 CCAGCGCTGCGACGTGAGCGTGCCTGCTGGTGGTGGGCGAGCCACCGAGGCCCGCTGGTGGCGGGCAGAGG
 CATTGTCTGGACCTGGCCATCCTCGACTCCGCCTTCTCCTCAGCCAGGTGGCCCCGTCCCTCTTCATGGGC
 TCTATCGTCCAGCTGTCTCAGAGCGTACCGCTTACATGGTGTCCGCTGCTGGACTGGGCTTGGTGGCTATTT
 ATTTCCGCCACCCAGGTGGTGTTCGACAAGAGCGACCTGGCCAAATACTCCGCCTGA

SEQ ID N°:20

ATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAAAGGCCAGTTGCTGCTGGTGAACCTGCTGA
 CTTTCGGACTGGAGGTGTGCCTGGCTGCCGGGATCACGTACGTGCCCCCTGCTGCTGGAGGTGGGCGTGA
 GGAGAAGTTCATGACAATGGTGTGGGCATCGGCCCGCTCCTGGGCCTCGTGTGTGTGCCCCCTCCTCGGGAGT
 GCGTCCGATCATTGGCGGGGCCGCTACGGCCGCCGAGACCGTTTCATCTGGGCCCTGAGCCTGGGGATCCTGC
 TCTCTCTCTTCTGATCCCCCGGCCGGCTGGCTGGCCGGCCTGCTGTGTCCCGACCCCCGCCCCCTCTGGAGCT
 GGCCCTCCTGATCCTGGGCGTGGGCTTGTGGACTTCTGCGGCCAGGTGTGTTTCACTCCCCTGGAGGCTCTG
 CTCTCCGACCTCTTCCGCGACCCGACCACTGTAGGCAGGCTTACAGCGTGTACGCTTCATGATCAGTCTGG

GGGGATGCCTGGGCTATCTGCTGCCCCGCTATCGACTGGGACACCAGCGCCCTGGCCCCCTACCTGGGGACTCA
GGAGGAGTGCCTGTTTCGGCCTGCTCACCTTGATCTTCCTGACGTGCGTCGCCGCCACCCTGCTGGTGGCCGAG
GAGGCGGCCCTGGGGCCCACCGAGCCCCGCCGAGGGCCTGAGCGCTCCCAGCCTGAGCCCCCATTGCTGCCCCGT
GCAGGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCTGCATCAGCTGTGCTGTGCGCATGCC
TCGCACCCTGCGCCGCCTGTTTCGTCGCTGAGCTCTGTTCTGGATGGCCCTGATGACGTTACCCCTCTTCTAC
ACCGACTTCGTGGGGGAGGGCCTGTACCAGGGCGTGCCCGAGGGCCGAGCCCGGCACCGAGGCTAGGCGCCATT
ACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCCTGCAGTGCGCCATCAGTCTGGTGTTCCTCTCTGGT
GATGGACCGGCTGGTGCAGCGCTTCGGCACCCGGGCGGTGTACCTCGCCTCTGTGGCGGGCTTTCCCGTCGCC
GCCGGCGCGACCTGCCTGTCTCATTTCTGTGCGCGTGGTGACCGCCAGCGCCGCCCTGACCGGCTTCACCTTCA
GTGCGCTCCAGATTCTGCCCTACACCCTGGCGTCTCTGTACCATCGCGAGAAGCAGGTGTTCTGCCCAAGTA
CCGCGGGGACACAGGGGGAGCTTCCTCTGAGGACAGCCTGATGACCAGCTTCTTGCCCGGCCCCAAGCCGGGG
GCCCCCTTTCCCAACGGCCATGTGCGGGCGGGCGGCAGCGGCCTGCTCCCTCCCCCCCCCGCCCTGTGCGGCG
CTAGTGCCCTGCGACGTGAGCGTGCGGGTGGTGGTGGGGAGCCCACCGAGGCTAGGGTCGTGCCTGGCCGGGG
GATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCCTGCTCTCCAGGTGGCGCCAGCCTGTTTCATGGGC
AGTATCGTGAGCTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCCGCCGGCCTGGGGTTGGTGGCCATCT
ACTTTGCCACCCAGGTCGTGTTTCGACAAGAGCGATCTCGCCAAGTATAGCGCCTGA

Fig. 15. Sequência 19 de codão otimizado re-manipulada (SEQ ID N°:21)

GACGGCTAGCGCCACCATGGTGCAGCGGCTCTGGGTGAGCCGCTCCTGCGGCATCGCAAGGCCAGCTCCTG
 CTGGTGAATCTGCTCACATTTCGGCCTGGAGGTGTGCCTGGCCGCCGGCATCACCTACGTGCCCCCCTCCTGC
 TGGAGGTGGGAGTCGAGGAGAAGTTCATGACCATGGTGCTGGGCATTGGGCCCCTCCTGGGCCCTCGTGTGCGT
 GCCTCTCCTCGGCAGCGCTTCGACCATTTGGCGCGGCCGGTATGGCCGAGGAGACCCTTCATCTGGGCTCTG
 AGTCTCGGCATCCTGCTGAGCCTGTTCTGATCCCTCGGGCCGGCTGGCTGGCCGGGCTGCTGTGCCCGATC
 CTCGGCCCCCTGGAGCTGGCCCTGCTGATCCTCGGCGTGGGCCTGCTGGACTTCTGCGGCCAGGTGTGCTTCAC
 GCCCCTGGAGGCACTGCTGAGCGACCTGTTCCGGGACCCCGACCATTTGCCGCCAGGCGTACAGCGTGTACGCC
 TTCATGATCTCCCTGGGAGGCTGCCTGGGCTACCTGCTCCCGCCATCGATTGGGACACCAGCGCACTCGCCC
 CCTATCTCGGAACACAGGAGGAATGCCTGTTCCGGA~~CTG~~~~CTG~~ACGCTCATCTTCCTCACGTGCGTCGCGGCCAC
 CTTGTTGGTGGCCGAGGAGGCCGCCCTGGGGCCACCGAGCCGGCCGAGGGACTGAGCGCCCCGAGCCTGAGT
 CCACACTGCTGCCCTTGCCGGGCCCCGCTGGCCTTCGTAATCTGGGCGCCCTCCTGCCTCGGCTCCATCAGC
 TGTGTTGCAGAATGCCTAGGACGCTGCGGCGCCTGTTCTGTCGCTGAGTTGTGCTCCTGGATGGCTCTCATGAC
 CTTACCCCTGTTTTATACGGA~~CTG~~CTTCGTCGGGAGGGCCTGTACCAGGGGGTGCCGCGCGCCGAGCCCGGACA
 GAGGCGCGCCGCCACTACGACGAGGGAGTGCCTATGGGCTCCCTGGGCCTCTTCTTGCA~~CTG~~GTGCGCCATCAGTC
 TGGTTTTCTCTCTGGTCATGGACAGGCTGGTGCAGCGCTTCGGAACCCGGGCGGTGTACCTGGCGAGCGTGGC
 CGCCTTCCCCGTGGCTGCCGGCGCCACCTGCCTCTCTCACTCGGTGGCCGTGGTCACCGCCAGCGCCGCCCTG
 ACCGGGTTACCTTCTCTGCCCTGCAGATTCTGCCTTACACCCTGGCCAGCCTGTACCATCGCGAGAAACAGG
 TGTCTCTCCCAAGTACAGAGGCGACACCGGGGGCGCCTCCAGCGAGGACAGCCTCATGACCTCCTTCCTGCC
 TGGCCCCAAGCCCGCGCCCCCTTTCCCAACGGGCACGTGGGCGCCGGCGGGAGTGGGCTCCTGCCCCCCCCCT
 CCTGCGCTGTGCGGGGCCAGCGCCTGCGACGTGAGCGTGC~~CGT~~GGTGGGCGAGCCACCGAGGCCCGCG
 TGGTGCCGGGCGAGGCATTTGTCTGGACCTGGCCATCCTCGACTCCGCCCTTCTCTCAGCCAGGTGGCCCC
 GTCCCTCTTCATGGGCTCTATCGTCCAGCTGTCTCAGAGCGTCACCGCTTACATGGTGTCCGCTGCTGGACTG
 GGCTTGGTGGCTATTTATTTGCGCACCCAGGTGGTGTTCGACAAGAGCGACCTGGCCAAATACTCCGCCTGAC
TCGAGGCAG

Fig. 16. Sequência 20 de codão otimizado re-manipulada (SEQ ID N°:22)

GACGGCTAGCGCCACCATGGTGACGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAAAGGCCCAGTTGCTG
CTGGTGAACCTGCTGACTTTCGACTGGAGGTGTGCTGGCTGCCGGGATCACGTACGTGCCCCCCTGCTGC
TGGAGGTGGGCGTGGAGGAGAAGTTCATGACAATGGTGTGGGCATCGGCCCCGTCTGGGCCTCGTGTGTGT
GCCCCCTCTCGGGAGTGCGTCCGATCATTGGCGGGGCCGCTACGGCCGCGCAGACCGTTCATCTGGGCCCTG
AGCCTGGGCTATCCTGCTCTCTCTCTTCTGATCCCCGGGCCGGCTGGCTGGCCGGCCTGCTGTGTCCCACC
CCCCCCTCTGGAGCTGGCCCTCCTGATCCTGGGCGTGGGCTGGCTGGACTTCTGCGGCCAGGTGTGTTTCAC
TCCCCGGAGGCTCTGCTCTCCGACCTCTTCCGCGACCCCGACCACTGTAGGCAGGCTTACAGCGTGACGCC
TTCATGATCAGTCTGGGGGGATGCTGGGCTATCTGCTGCCCCGTATCGACTGGGACACCAGCGCCCTGGCCC
CCTACCTGGGGACTCAGGAGGAGTGCTGTTTCGGCCTGCTCACCTTGATCTTCTGACGTGCGTCGCCGCCAC
CCTGCTGGTGGCCGAGGAGGCGGCCCTGGGGCCACCAGCCCCGAGGGCCTGAGCGCTCCAGCCTGAGC
CCCCATTGCTGCCCCGTGCAGGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCCTGCATCAGC
TGTGCTGTGCGATGCCTCGCACCTGCGCCGCTGTTCTGCTGAGCTCTGTTCTGGATGGCCCTGATGAC
GTTACCCCTCTTCTACACCGACTTCGTGGGGAGGGCCTGTACCAGGGCGTGCCAGGGCCGAGCCCGGCACC
GAGGCTAGGCGCCATTACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCTGCGAGTGCGCCATCAGTC
TGGTGTCTCTCTGGTGTATGGACCGCTGGTGCAGCGCTTCGGCACCCGGGCCGTGTACCTCGCCTCTGTGGC
GGCTTTCCCCGTGCGCGCCGGCGCGACCTGCCTGTCTCATTCTGTGCGCGTGGTGACCGCCAGCGCCGCCCTG
ACCGGCTTCACCTTCAGTGCGCTCCAGATTCTGCCCTACACCTGGCGTCTCTGTACCATCGCGAGAAGCAGG
TGTTCTGCCCCAAGTACCGCGGGGACACAGGGGGAGCTTCCTCTGAGGACAGCCTGATGACCAGCTTCTTGCC
CGGCCCCAAGCGGGGGCCCCCTTCCCCAACGGCCATGTGCGGGCGGGCGGCAGCGGCTGCTCCCTCCCCC
CCCGCCCTGTGCGGCGCTAGTGCTGCGACGTGAGCGTGCGGGTGGTGGTGGGGAGCCCCACCGAGGCTAGGG
TCGTGCTTGGCCGGGGATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCTGCTCTCCAGGTGGCGCC
CAGCCTGTTCATGGGCAGTATCGTGACGTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCGCCGGCCTG
GGGTGGTGGCCATCTACTTTGCCACCCAGGTGCTGTTTCGACAAGAGCGATCTCGCCAAGTATAGCGCCTGAC
TCGAGGCAG

Fig. 17. A sequência inicial para a otimização de CPC (SEQ ID N°:23)

Quatro aminoácidos da sequência de P501S estão numa caixa de texto

ATGGCGCCGCTTACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGT
ACTACTTTGACAGTTTCAGGCTATATGCTTGACAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGT
CGACAACCTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGT
GCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGCAAT
ACATCAAGGCTAACTCTAAGTTTCATTGGTATCACTGAAGGCGTCATGGTATCAAATGCCTTTATCCAGTCAGC
GGACGGAAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAAGTTCATGTAC

Fig. 18. Sequências de CPC de códon otimizado representativas (SEQ ID N°:24-25)

SEQ ID N°:24

ATGGCCCGCCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACAAGTTGAGAAGATCAACGGGACATGGT
ACTACTTCGACTCCTCCGGCTACATGCTCGCCGACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGT
CGATAACTCGGGAGAGATGGCCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCAACGAGGAGGGC
GCCATGAAGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCATGCAGT
ATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCAACGCCTTTATCCAGAGCGC
CGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCCCTCGCGGATCGGCCGAGAAGTTCATGTAC

SEQ ID N°:25

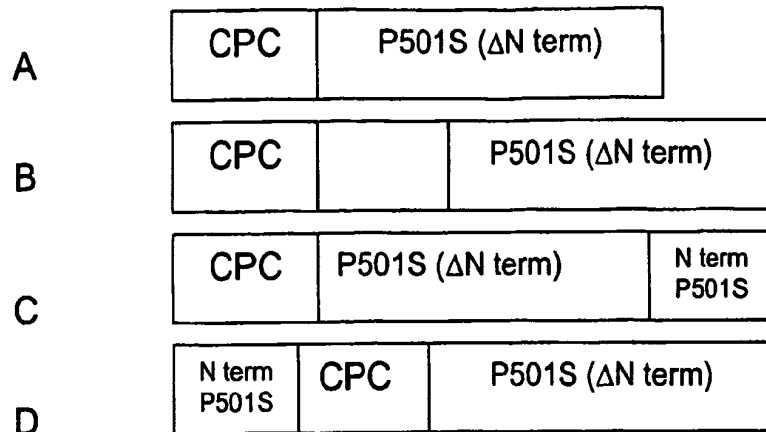
ATGGCCCGCCCTACGTGCACAGCGACGGGTCCTACCCAAAGGACAAGTTGAGAAGATCAACGGCACGTGGT
ACTATTTGACAGCAGCGGCTACATGCTCGCCGATCGCTGGCGCAAGCACACCGACGGGAAGTGGTACTGGT
CGACAACCTCTGGCGAGATGGCTACGGGGTGGAAGAAGATCGCGGACAAGTGGTACTACTTCAACGAGGAGGGC
GCCATGAAGACCGGGTGGGTGAAGTACAAGGACACCTGGTACTACCTGGACGCTAAGGAGGGCGCCATGCAGT
ACATCAAGGCCAACTCGAAGTTCATCGGGATCACCGAGGGCGTGATGGTCAGTAACGCCTTTATCCAGAGCGC
GGACGGCACAGGCTGGTATTACCTGAAGCCCGATGGCACCCCTGGCGGACAGACCTGAGAAATTCATGTAC

Fig. 19. Sequência de CPC de códon otimizado manipulada (SEQ ID N°:26)

SEQ ID N°:26

GACGGCTAGCGCCACCATGGCCCGCCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACAAGTTGAGAAG
ATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCGACCGCTGGCGGAAGCACACCGACG
GCAACTGGTACTGGTTCGATAACTCGGGAGAGATGGCCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTA
TTTCAACGAGGAGGGCGCCATGAAGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAG
GAGGGCGCCATGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCAACG
CCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCCCTCGCGGATCGGCCGA
GAAGTTCATGTAC

Fig. 20. Sequências e construções de candidato de fusão de P501S CPC



Construção A = SEQ ID N° 37 (nucleótidos) e 45 (polipéptido)

```

GCGGCCGCGCCACCATGGCCGCGCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACA
      M A A A Y V H S D G S Y P K D K

AGTTCGAGAAGATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCG
      F E K I N G T W Y Y F D S S G Y M L A D

ACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTTCGATAACTCGGGAGAGATGG
      R W R K H T D G N W Y W F D N S G E M A

CCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCACGAGGAGGGCGCCATGA
      T G W K K I A D K W Y Y F N E E G A M K

AGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCA
      T G W V K Y K D T W Y Y L D A K E G A M

TGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCA
      Q Y I K A N S K F I G I T E G V M V S N

ACGCCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCC
      A F I Q S A D G T G W Y Y L K P D G T L

TCGCGGATCGGCCCGAGAAGTTCATGTACATGGTGTCTGGGCATCGGCCCGTCCTGGGCC
      A D R P E K F M Y M V L G I G P V L G L

TCGTGTGTGTGCCCTCCTCGGGAGTGCGTCCGATCATTGGCGGGGCCGCTACGGCCGCC
      V C V P L L G S A S D H W R G R Y G R R

GCAGACCGTTTCATCTGGGCCCTGAGCCTGGGCATCCTGCTCTCTCTTCTGATCCCCC
      R P F I W A L S L G I L L S L F L I P R

GGGCCGCTGGCTGGCCGGCCTGCTGTGTCCCGACCCCCGCCCTCTGGAGCTGGCCCTCC
      A G W L A G L L C P D P R P L E L A L L

TGATCCTGGGCGTGGGCCTGCTGGACTTCTGCGGCCAGGTGTGTTTCACTCCCCTGGAGG

```

I L G V G L L D F C G Q V C F T P L E A
 CTCTGCTCTCCGACCTCTTCCGCGACCCCGACCACTGTAGGCAGGCTTACAGCGTGTACG
 L L S D L F R D P D H C R Q A Y S V Y A
 CCTTCATGATCAGTCTGGGGGGATGCCTGGGCTATCTGCTGCCCCGTATCGACTGGGACA
 F M I S L G G C L G Y L L P A I D W D T
 CCAGCGCCCTGGCCCCCTACCTGGGGACTCAGGAGGAGTGCCTGTTCCGGCCTGCTCACCT
 S A L A P Y L G T Q E E C L F G L L T L
 TGATCTTCTGACGTGCGTCCGCCACCCTGCTGGTGGCCGAGGAGGCGGCCCTGGGGC
 I F L T C V A A T L L V A E E A A L G P
 CCACCGAGCCCGCGAGGGCCTGAGCGCTCCAGCCTGAGCCCCATTGCTGCCCCGTGCA
 T E P A E G L S A P S L S P H C C P C R
 GGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCTGCATCAGCTGTGCT
 A R L A F R N L G A L L P R L H Q L C C
 GTCGCATGCCTCGCACCCCTGCGCCGCTGTTGCTGCTGAGCTCTGTTCTGGATGGCCC
 R M P R T L R R L F V A E L C S W M A L
 TGATGACGTTACCCCTCTTCTACACCGACTTCGTGGGGAGGGCCTGTACCAGGGCGTGC
 M T F T L F Y T D F V G E G L Y Q G V P
 CCAGGGCCGAGCCCGGCACCGAGGCTAGGCGCCATTACGACGAGGGCGTCAGGATGGGCT
 R A E P G T E A R R H Y D E G V R M G S
 CTCTGGGCTCTTCTGTCAGTGCGCCATCAGTCTGGTGTCTCTCTGTTGATGACCAGGC
 L G L F L Q C A I S L V F S L V M D R L
 TGGTGCAGCGCTTCGGCACCCGGGCGGTGTACCTCGCCTCTGTGGCGGCTTTCCCCGTCG
 V Q R F G T R A V Y L A S V A A F P V A
 CCGCCGGCGCGACCTGCCTGTCTCATTTCTGTGCGCGTGGTGACCGCCAGCGCCGCCCTGA
 A G A T C L S H S V A V V T A S A A L T
 CCGGCTTCACCTTCAGTGCCTCCAGATTCTGCCCTACACCCTGGCGTCTCTGTACCATC
 G F T F S A L Q I L P Y T L A S L Y H R
 GCGAGAAGCAGGTGTTCTGCCCCAAGTACCGCGGGACACAGGGGAGCTTCTCTGAGG
 E K Q V F L P K Y R G D T G G A S S E D
 ACAGCCTGATGACCAGCTTCTTGCCCCGCCCCAAGCCGGGGGCCCTTTCCCCAACGGCC
 S L M T S F L P G P K P G A P F P N G H
 ATGTCGGGGCGGGCGGCAGCGGCCTGCTCCCTCCCCCCCCCGCCTGTGCGGCGCTAGTG
 V G A G G S G L L P P P P A L C G A S A
 CCTGCGACGTGAGCGTGCGGGTGGTGGTGGGGGAGCCACCGAGGCTAGGGTCTGCTGCTG
 C D V S V R V V V G E P T E A R V V P G
 GCCGGGGGATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCTGCTCTCCAGGTGG
 R G I C L D L A I L D S A F L L S Q V A
 CGCCAGCCTGTTTCATGGGCAGTATCGTGCAGCTGAGCCAGAGCGTGACCGCCTACATGG
 P S L F M G S I V Q L S Q S V T A Y M V

TGAGCGCCGCCGGCCTGGGGTTGGTGGCCATCTACTTTGCCACCCAGGTCGTGTTGACA
S A A G L G L V A I Y F A T Q V V F D K

AGAGCGATCTCGCCAAGTATAGCGCCTGAGGATCC
S D L A K Y S A *

Construção B = SEQ ID N°38 (nucleótidos) e 46 (polipeptido)

GCGGCCGCGCCACCATGGCCGCCGCTACGTGCATAGCGACGGGAGCTACCCCAAGGACA
M A A A Y V H S D G S Y P K D K

AGTTCGAGAAGATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCG
F E K I N G T W Y Y F D S S G Y M L A D

ACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTTCGATAACTCGGGAGAGATGG
R W R K H T D G N W Y W F D N S G E M A

CCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCAACGAGGAGGGCGCCATGA
T G W K K I A D K W Y Y F N E E G A M K

AGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCA
T G W V K Y K D T W Y Y L D A K E G A M

TGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCAGGGAGTGATGGTCAGCA
Q Y I K A N S K F I G I T E G V M V S N

ACGCCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCC
A F I Q S A D G T G W Y Y L K P D G T L

TCGCGGATCGGCCCGAGATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAA
A D R P E M V Q R L W V S R L L R H R K

AGGCCCAGTTGCTGCTGGTGAACCTGCTGACTTTCGGACTGGAGGTGTGCCTGGCTGCCG
A Q L L L V N L L T F G L E V C L A A G

GGATCACGTACGTGCCCCCCTGCTGCTGGAGGTGGGCGTGGAGGAGAAGTTCATGACAA
I T Y V P P L L L E V G V E E K F M T M

TGGTGTGGGCATCGGCCCGCTCCTGGGCCTCGTGTGTGTGCCCCCTCCTCGGGAGTGCGT
V L G I G P V L G L V C V P L L G S A S

CCGATCATTGGCGGGGCCGCTACGGCCGCCGAGACCGTTCATCTGGGCCCTGAGCCTGG
D H W R G R Y G R R R P F I W A L S L G

GCATCCTGCTCTCTCTCTCTCTGATCCCCGGGCCGGCTGGCTGGCCGGCCTGCTGTGTC
I L L S L F L I P R A G W L A G L L C P

CCGACCCCCGCCCTCTGGAGCTGGCCCTCCTGATCCTGGGCGTGGGCCTGCTGGACTTCT
D P R P L E L A L L I L G V G L L D F C

GCGGCCAGGTGTGTTTCACTCCCTGGAGGCTCTGCTCTCCGACCTCTTCCGCGACCCCG
G Q V C F T P L E A L L S D L F R D P D

ACCACTGTAGCGAGGCTTACAGCGTGTACGCCTTCATGATCAGTCTGGGGGGATGCCTGG
H C R Q A Y S V Y A F M I S L G G C L G

GCTATCTGCTGCCCGCTATCGACTGGGACACCAGCGCCCTGGCCCCCTACCTGGGGACTC
Y L L P A I D W D T S A L A P Y L G T Q

AGGAGGAGTGCCCTGTTTCGGCCTGCTCACCTTGATCTTCCTGACGTGCGTCGCCGCCACCC
 E E C L F G L L T L I F L T C V A A T L
 TGCTGGTGGCCGAGGAGGCGGCCCTGGGGCCACCGAGCCCGCCGAGGGCCTGAGCGCTC
 L V A E E A A L G P T E P A E G L S A P
 CCAGCCTGAGCCCCCATGCTGCCCCGTGCAGGGCTAGGCTCGCCTTCAGGAATCTGGGCG
 S L S P H C C P C R A R L A F R N L G A
 CTTTGCTGCCCCGCTGCATCAGCTGTGCTGTGCGCATGCCTCGCACCTGCGCCGCTGT
 L L P R L H Q L C C R M P R T L R R L F
 TCGTCGCTGAGCTCTGTTCTCTGGATGGCCCTGATGACGTTACCCCTCTTCTACACCGACT
 V A E L C S W M A L M T F T L F Y T D F
 TCGTGGGGGAGGGCCTGTACCAGGGCGTGCCAGGGCCGAGCCCGGCACCGAGGCTAGGC
 V G E G L Y Q G V P R A E P G T E A R R
 GCCATTACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCCTGCAGTGCGCCATCA
 H Y D E G V R M G S L G L F L Q C A I S
 GTCTGGTGTCTCTCTGTTGATGGACCGGCTGGTGCAGCGCTTCGGCACCCGGGCGGTGT
 L V F S L V M D R L V Q R F G T R A V Y
 ACCTCGCCTCTGTGGCGGCTTTCCCCGTGCGCGCCGGCGCGACCTGCCTGTCTCATTCTG
 L A S V A A F P V A A G A T C L S H S V
 TCGCCGTGGTGACCGCCAGCGCCGCCCTGACCGGCTTCACCTTCAGTGCGCTCCAGATTC
 A V V T A S A A L T G F T F S A L Q I L
 TGCCCTACACCTGGCGTCTCTGTACCATCGCGAGAAGCAGGTGTTCTGCCCCAAGTACC
 P Y T L A S L Y H R E K Q V F L P K Y R
 GCGGGGACACAGGGGAGCTTCCTCTGAGGACAGCCTGATGACCAGCTTCTTGCCCGGCC
 G D T G G A S S E D S L M T S F L P G P
 CCAAGCCGGGGGCCCCCTTTCCCCAACGGCCATGTGCGGGCGGGCGGCAGCGGCCTGCTCC
 K P G A P F P N G H V G A G G S G L L P
 CTCCCCCCCCCGCCCTGTGCGGCGCTAGTGCCTGCGACGTGAGCGTGCGGGTGGTGGTGG
 P P P A L C G A S A C D V S V R V V V G
 GGGAGCCCACCGAGGCTAGGGTCGTGCCTGGCCGGGGGATCTGCCTGGACCTGGCCATCC
 E P T E A R V V P G R G I C L D L A I L
 TCGACTCCGCCTTCCTGCTCTCCCAGGTGGCGCCAGCCTGTTTCATGGGCAGTATCGTGC
 D S A F L L S Q V A P S L F M G S I V Q
 AGCTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCCCGCCGCTGGGGTTGGTGGCCA
 L S Q S V T A Y M V S A A G L G L V A I
 TCTACTTTGCCACCCAGGTCGTGTTTCGACAAGAGCGATCTCGCCAAAGTATAGCGCCTGAG
 Y F A T Q V V F D K S D L A K Y S A *

GATCC

Construção C = SEQ ID N°39 (nucleótidos) e 47 (polipéptido)

GCGGCCGCGCCACCATGGCCCGCGCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACA
 M A A A Y V H S D G S Y P K D K

AGTTTCGAGAAGATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCG
 F E K I N G T W Y Y F D S S G Y M L A D

ACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTTCGATAACTCGGGAGAGATGG
 R W R K H T D G N W Y W F D N S G E M A

CCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCAACGAGGAGGGCGCCATGA
 T G W K K I A D K W Y Y F N E E G A M K

AGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCA
 T G W V K Y K D T W Y Y L D A K E G A M

TGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCA
 Q Y I K A N S K F I G I T E G V M V S N

ACGCCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCC
 A F I Q S A D G T G W Y Y L K P D G T L

TCGCGGATCGGCCCCGAGAAGTTCATGTACATGGTGCTGGGCATCGGCCCCGTCCTGGGGC
 A D R P E K F M Y M V L G I G P V L G L

TCGTGTGTGTGCCCCCTCCTCGGGAGTGCCTCCGATCATTGGCGGGGCCGCTACGGCCGCC
 V C V P L L G S A S D H W R G R Y G R R

GCAGACCGTTTCATCTGGGCCCTGAGCCTGGGCATCCTGCTCTCTCTCTCCTGATCCCCC
 R P F I W A L S L G I L L S L F L I P R

GGGCCGGCTGGCTGGCCGGCCTGCTGTGTCCCGACCCCGCCCTCTGGAGCTGGCCCTCC
 A G W L A G L L C P D P R P L E L A L L

TGATCCTGGGCGTGGGCGCTGCTGGACTTCTGCGGCCAGGTGTGTTTCACTCCCTGGAGG
 I L G V G L L D F C G Q V C F T P L E A

CTCTGCTCTCCGACCTCTTCCGCGACCCCGACCACTGTAGGCAGGCTTACAGCGTGACG
 L L S D L F R D P D H C R Q A Y S V Y A

CCTTCATGATCAGTCTGGGGGGATGCCTGGGCTATCTGCTGCCCCGCTATCGACTGGGACA
 F M I S L G G C L G Y L L P A I D W D T

CCAGCGCCCTGGCCCCCTACCTGGGGACTCAGGAGGAGTGCCTGTTTCGGCCTGCTCACCT
 S A L A P Y L G T Q E E C L F G L L T L

TGATCTTCCTGACGTGCGTCGCCGCCACCCTGCTGGTGGCCGAGGAGGCGGCCCTGGGGC
 I F L T C V A A T L L V A E E A A L G P

CCACCGAGCCCGCCGAGGGCCTGAGCGCTCCCAGCCTGAGCCCCCATTTGCTGCCCCGTGCA
 T E P A E G L S A P S L S P H C C P C R

GGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCTGCATCAGCTGTGCT
 A R L A F R N L G A L L P R L H Q L C C

GTCGCATGCCTCGCACCTGCGCCCGCTGTTGCTGCTGAGCTCTGTTCTCTGGATGGCCC
 R M P R T L R R L F V A E L C S W M A L

TGATGACGTTACCCCTCTTCTACACCGACTTCGTGGGGGAGGGCCTGTACCAGGGCGTGC
 M T F T L F Y T D F V G E G L Y Q G V P

CCAGGGCCGAGCCCGGCACCGAGGCTAGGCGCCATTACGACGAGGGCGTCAGGATGGGCT
 R A E P G T E A R R H Y D E G V R M G S
 CTCTGGGCCTCTTCTGCACTGCGCCATCAGTCTGGTGTCTCTCTGGTGATGGACCGGC
 L G L F L Q C A I S L V F S L V M D R L
 TGGTGACGCGCTTCGGCACCCGGGCGGTGTACCTCGCCTCTGTGGCGGCTTTCCCCGTCG
 V Q R F G T R A V Y L A S V A A F P V A
 CCGCCGGCGCGACCTGCCTGTCTCATTTCTGTGCGCGTGGTGACCGCCAGCGCCGCCCTGA
 A G A T C L S H S V A V V T A S A A L T
 CCGGCTTCACCTTCAGTGCCTCCAGATTCTGCCCTACACCCTGGCGTCTCTGTACCATC
 G F T F S A L Q I L P Y T L A S L Y H R
 GCGAGAAGCAGGTGTTCTGCCCCAAGTACCGCGGGGACACAGGGGGAGCTTCTCTGAGG
 E K Q V F L P K Y R G D T G G A S S E D
 ACAGCCTGATGACCAGCTTCTTGCCCCGGCCCCAAGCCGGGGGCCCTTTCCCCAACGGCC
 S L M T S F L P G P K P G A P F P N G H
 ATGTGGGGGGGGCGGCAGCGGCCTGCTCCCTCCCCCCCCCGCCCTGTGCGGCGCTAGTG
 V G A G G S G L L P P P P A L C G A S A
 CCTGCGACGTGAGCGTGCGGGTGGTGGTGGGGGAGCCCCACCGAGGCTAGGGTCTGTCCTG
 C D V S V R V V V G E P T E A R V V P G
 GCCGGGGGATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCTGTCTCTCCAGGTGG
 R G I C L D L A I L D S A F L L S Q V A
 CGCCCAGCCTGTTTCATGGGCAGTATCGTGACGTGAGCCAGAGCGTGACCGCCTACATGG
 P S L F M G S I V Q L S Q S V T A Y M V
 TGAGCGCCGCGCCTGGGGTGGTGGCCATCTACTTTGCCACCCAGGTCGTGTTGACA
 S A A G L G L V A I Y F A T Q V V F D K
 AGAGCGATCTCGCCAAGTATAGCGCCATGGTGACGGCTGTGGGTGTCCCGGCTGCTGC
 S D L A K Y S A M V Q R L W V S R L L R
 GCCATAGAAAGGCCAGTTGCTGCTGGTGAACCTGCTGACTTTTCGGAAGGTGGAGGTGTGCC
 H R K A Q L L L V N L L T F G L E V C L
 TGGCTGCGGGATCAGTACGTGCCCCCCTGCTGCTGGAGGTGGGCGTGGAGGAGTGAG
 A A G I T Y V P P L L L E V G V E E *
 GATCC

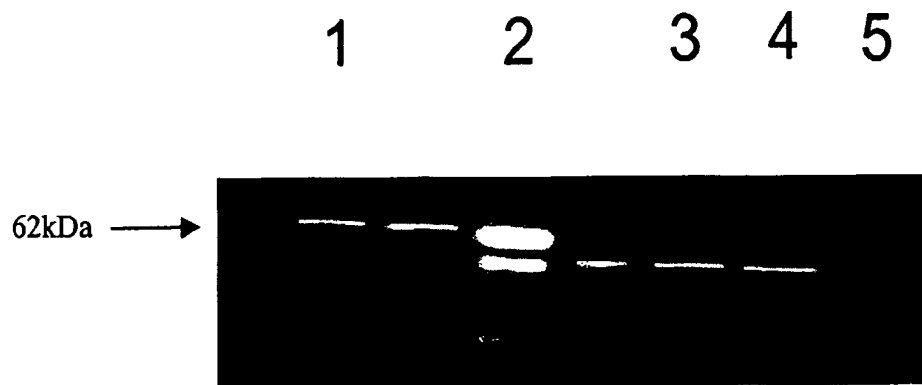
Construção D = SEQ ID N°40 (nucleótidos) e 48 (polipéptido)

GCGGCCGCGCCACCATGGTGACGGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAAAGG
 M V Q R L W V S R L L R H R K A
 CCCAGTTGCTGCTGGTGAACCTGCTGACTTTTCGGAAGGTGTGCTGGCTGCCGGGA
 Q L L L V N L L T F G L E V C L A A G I
 TCACGTACGTGCCCCCCTGCTGCTGGAGGTGGGCGTGGAGGAGATGGCCGCCGCTACG
 T Y V P P L L L E V G V E E M A A A Y V

TGCATAGCGACGGGAGCTACCCCAAGGACAAGTTCGAGAAGATCAACGGGACATGGTACT
H S D G S Y P K D K F E K I N G T W Y Y
ACTTCGACTCCTCCGGCTACATGCTCGCCGACCGCTGGCGGAAGCACACCGACGGCAACT
F D S S G Y M L A D R W R K H T D G N W
GGTACTGGTTCGATAACTCGGGAGAGATGGCCACCGGCTGGAAGAAGATCGCGGACAAGT
Y W F D N S G E M A T G W K K I A D K W
GGTACTATTTCAACGAGGAGGGCGCCATGAAGACCGGCTGGGTGAAGTATAAGGACACCT
Y Y F N E E G A M K T G W V K Y K D T W
GGTACTACCTCGACGCCAAGGAGGGCGCCATGCAGTATATCAAGGCCAACAGCAAGTTCA
Y Y L D A K E G A M Q Y I K A N S K F I
TCGGCATCACCGAGGGAGTGATGGTCAGCAACGCCTTTATCCAGAGCGCCGACGGCACCGG
G I T E G V M V S N A F I Q S A D G T G
GATGGTACTACTTGAAGCCGGACGGCACCTCGCGGATCGGCCCGAGAAGTTCATGTACA
W Y Y L K P D G T L A D R P E K F M Y M
TGGTGCTGGGCATCGGCCCCGTCTGGGCTCGTGTGTGTGCCCCCTCTCGGGAGTGCCT
V L G I G P V L G L V C V P L L G S A S
CCGATCATTTGGCGGGGCGCTACGGCCGCCGAGACCGTTCATCTGGGCCCTGAGCCTGG
D H W R G R Y G R R R P F I W A L S L G
GCATCCTGCTCTCTCTCTCTCTGATCCCCGGGCGGCTGGCTGGCCGGCCTGCTGTGTC
I L L S L F L I P R A G W L A G L L C P
CCGACCCCCGCCCTCTGGAGCTGGCCCTCCTGATCCTGGGCGTGGGCTGCTGGACTTCT
D P R P L E L A L L I L G V G L L D F C
GCGGCCAGGTGTGTTTCACTCCCCTGGAGGCTCTGCTCTCCGACCTCTTCCGCGACCCCG
G Q V C F T P L E A L L S D L F R D P D
ACCACTGTAGGCAGGCTTACAGCGTGTACGCCTTCATGATCAGTCTGGGGGGATGCCTGG
H C R Q A Y S V Y A F M I S L G G C L G
GCTATCTGCTGCCCCTATCGACTGGGACACCAGCGCCCTGGCCCCCTACCTGGGGACTC
Y L L P A I D W D T S A L A P Y L G T Q
AGGAGGAGTGCCTGTTCGGCCTGCTCACCTTGATCTTCCTGACGTGCGTCGCCGCCACCC
E E C L F G L L T L I F L T C V A A T L
TGCTGGTGGCCGAGGAGGCGGCCCTGGGGCCCACCGAGCCCGCGAGGGCCTGAGCGCTC
L V A E E A A L G P T E P A E G L S A P
CCAGCCTGAGCCCCATTGCTGCCCCTGCAGGGCTAGGCTCGCCTTCAGGAATCTGGGCG
S L S P H C C P C R A R L A F R N L G A
CTTTGCTGCCCCGCCTGCATCAGCTGTGCTGTGTCGATGCCTCGCACCTGCGCCGCTGT
L L P R L H Q L C C R M P R T L R R L F
TCGTCGCTGAGCTCTGTTCTGGATGGCCCTGATGACGTTACCCCTCTTCTACACCGACT
V A E L C S W M A L M T F T L F Y T D F
TCGTGGGGGAGGGCCTGTACCAGGGCGTGCCAGGGCCGAGCCCGGCACCGAGGCTAGGC

V G E G L Y Q G V P R A E P G T E A R R
 GCCATTACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCCTGCAGTGCGCCATCA
 H Y D E G V R M G S L G L F L Q C A I S
 GTCTGGTGTCTCTCTGGTGATGGACCGGCTGGTGCAGCGCTTCGGCACCCGGGCGTGT
 L V F S L V M D R L V Q R F G T R A V Y
 ACCTCGCCTCTGTGGCGGCTTTCCCCGTGCGCGCCGGCGGACCTGCCTGTCTCATTCTG
 L A S V A A F P V A A G A T C L S H S V
 TCGCCGTGGTGACCGCCAGCGCGCCCTGACCGGCTTCACCTTCAGTGCGCTCCAGATTC
 A V V T A S A A L T G F T F S A L Q I L
 TGCCCTACACCCTGGCGTCTCTGTACCATCGCGAGAAGCAGGTGTTCCCTGCCCAAGTACC
 P Y T L A S L Y H R E K Q V F L P K Y R
 GCGGGGACACAGGGGAGCTTCCTCTGAGGACAGCCTGATGACCAGCTTCTTGCCCGGCC
 G D T G G A S S E D S L M T S F L P G P
 CCAAGCCGGGGGCCCCCTTTCCCCAACGGCCATGTGCGGGCGGGCGGCAGCGCCTGCTCC
 K P G A P F P N G H V G A G G S G L L P
 CTCCCCCCCCCGCCTGTGCGGCGCTAGTGCCTGCGACGTGAGCGTGCGGGTGGTGGTGG
 P P P A L C G A S A C D V S V R V V V G
 GGGAGCCCACCGAGGCTAGGGTCGTGCCCTGGCCGGGGGATCTGCCTGGACCTGGCCATCC
 E P T E A R V V P G R G I C L D L A I L
 TCGACTCCGCCTTCTGTCTCTCCAGGTGGCGCCAGCCTGTTTCATGGGCAGTATCGTGC
 D S A F L L S Q V A P S L F M G S I V Q
 AGCTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCCCGGCCTGGGGTTGGTGGCCA
 L S Q S V T A Y M V S A A G L G L V A I
 TCTACTTTGCCACCCAGGTCGTGTTTCGACAAGAGCGATCTCGCCAAGTATAGCGCCTGAG
 Y F A T Q V V F D K S D L A K Y S A *
 GATCC

Fig. 21. Análise de transferência de Western de células CHO após transfecção transitória com P501 S (JNW680), CPC-P501S (JNW735) e controlo com vector vazio.



Pista	Amostra
1	CPC-P501S (JNW735)
2	Proteína CPC-P501S (62,5 ng)
3	P501S (JNW680)
4	P501S (JNW680)
5	Controlo com vector vazio

Fig. 22. Respostas de anticorpo anti-P501S após imunização no dia 0, 21 e 42 com pVAC-P501S (JNW680, murganhos B1-9) ou vector vazio (pVAC, murganhos A1-6).

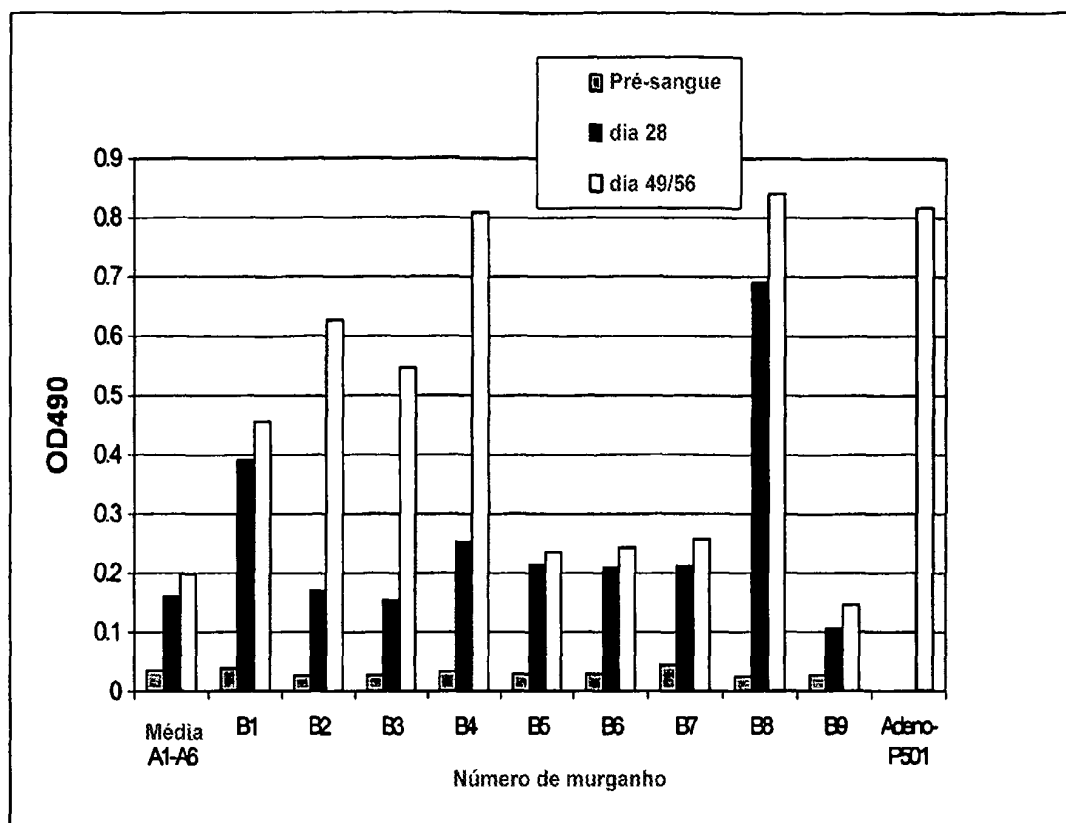


Fig. 23. Rastreo da biblioteca peptídica utilizando murganhos C57BL/6 imunizados no dia 0, 21, 42 e 70 com pVAC-P501S (JNW680).

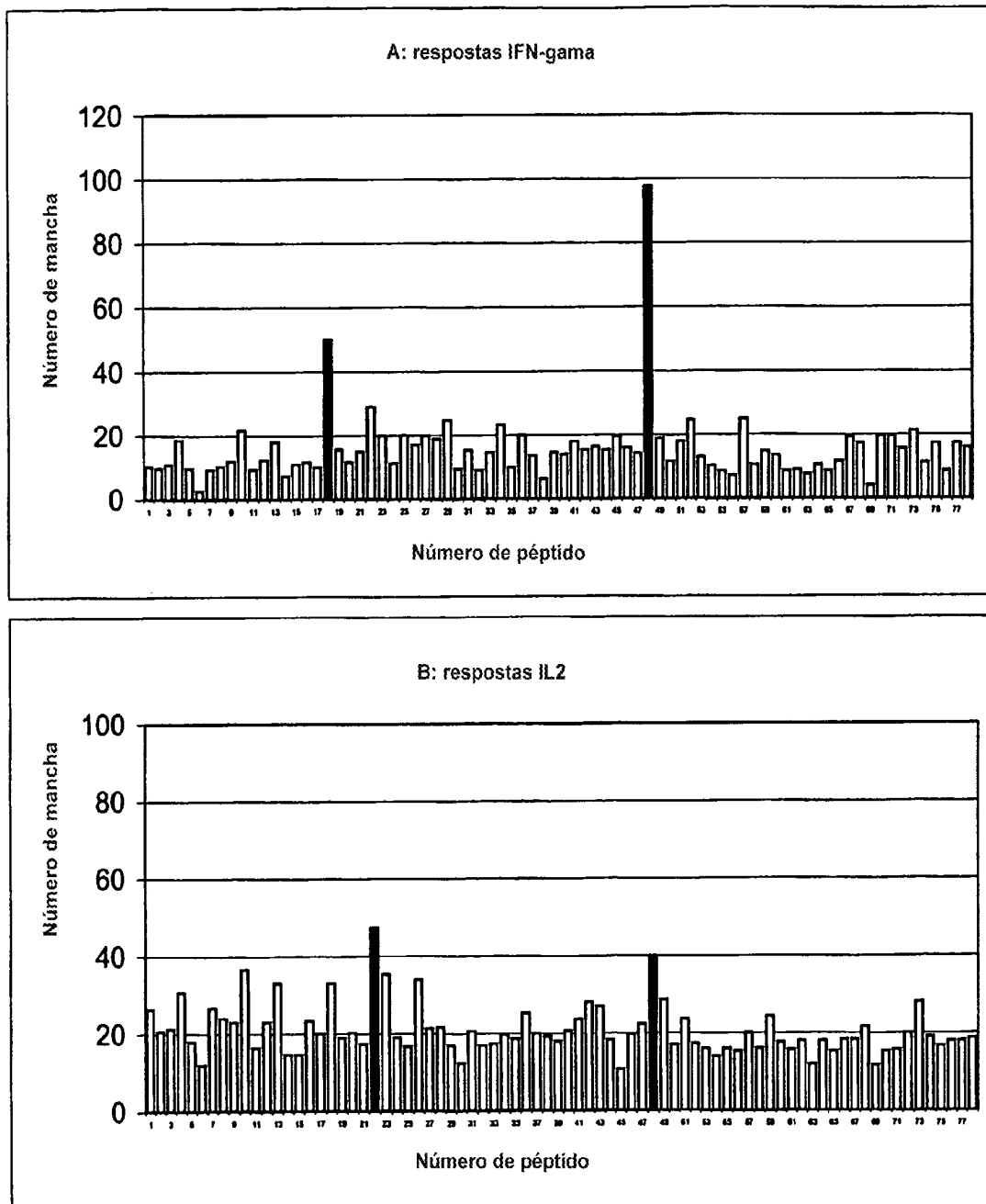


Fig. 24. Respostas celulares por ELISPOT no dia 77 após imunização PMID no dia 0, 21, 42 e 70 com pVAC-P501S (JNW680, B6-9) e pVAC vazio (A4-6).

O Gráfico A mostra as respostas IFN- γ enquanto que o Gráfico B mostra as respostas IL-2.

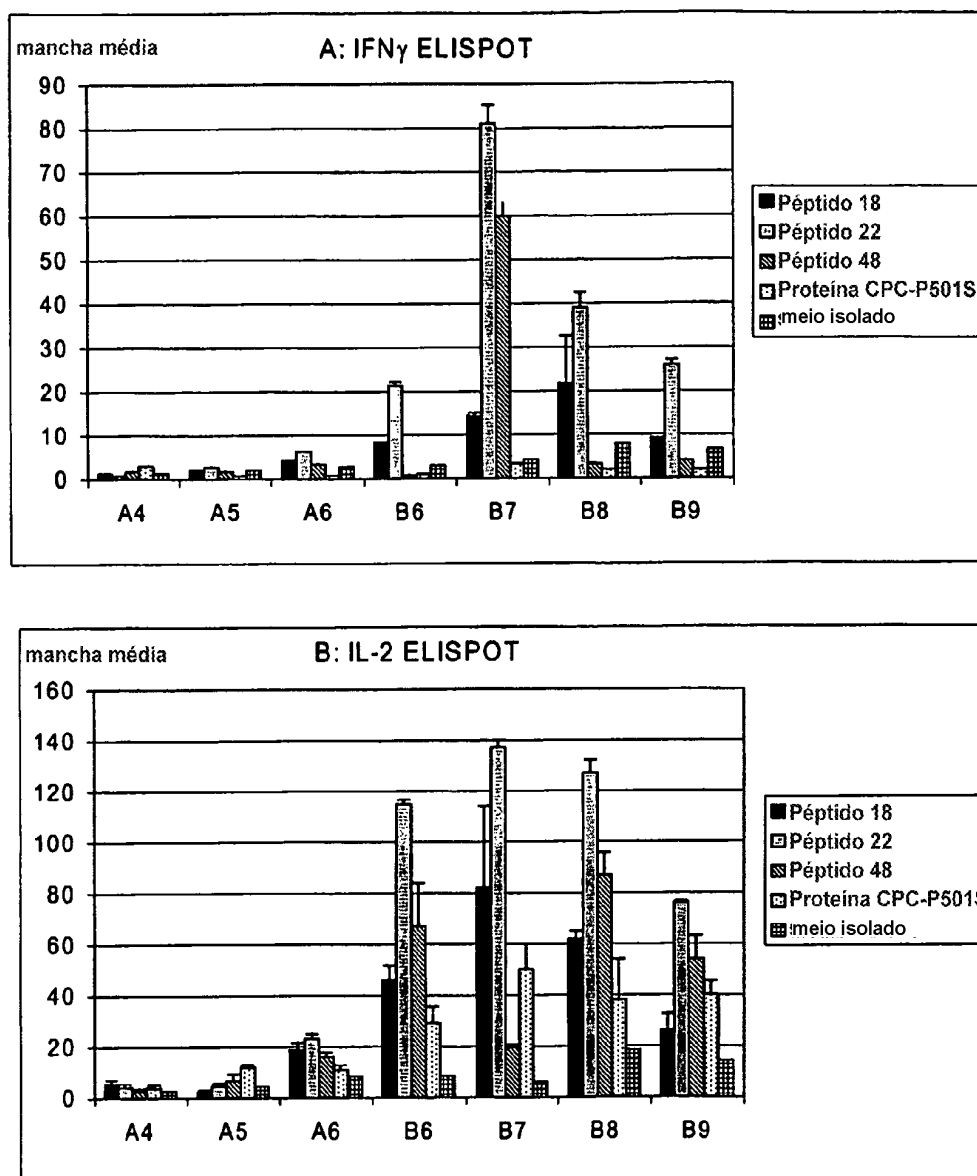


Fig. 25. Comparação de P501S e CPC-P501S.

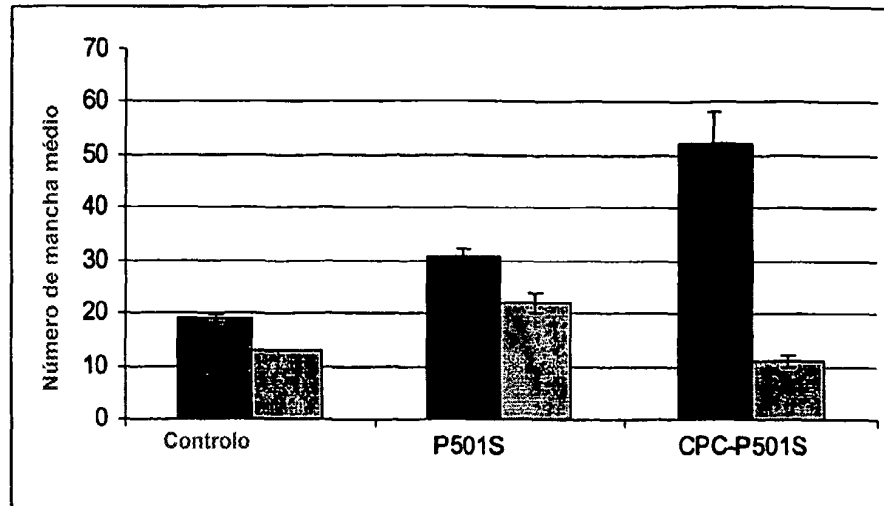


Fig. 26. Resposta imunológica (linfoproliferação em células de baço) após imunização de proteína com CPC-P501S.

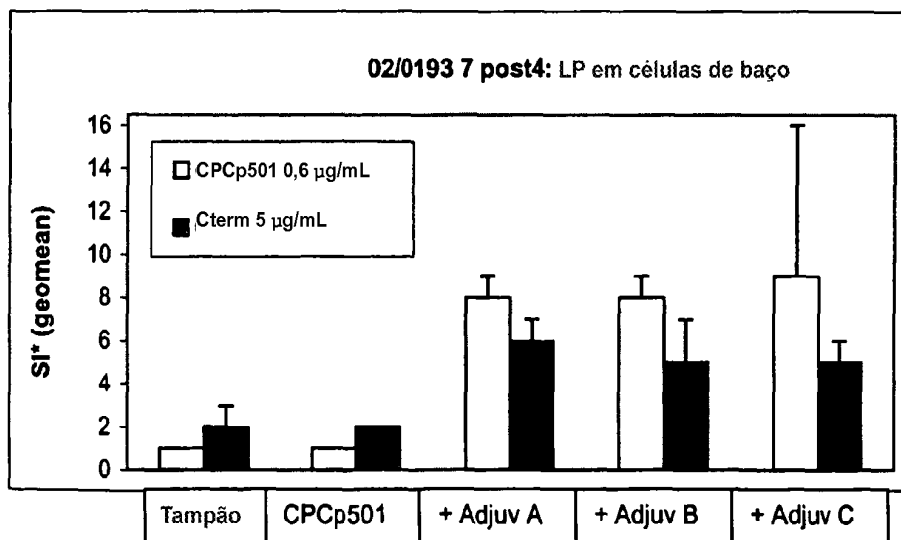


Fig. 27. Avaliação da resposta imunológica a construções CPC-P501S diferentes.

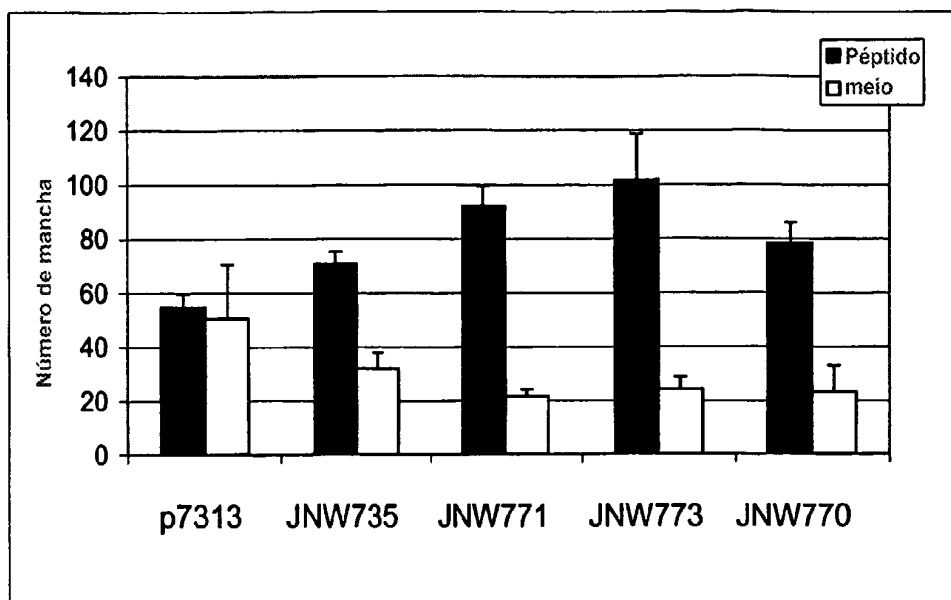


Fig. 28. Sequências de ADN e de polipéptido de MUC1-CPC

Fig. 28A. Sequência de ADN (SEQ ID N° 49)

ATGACACCGGGCACCCAGTCTCCTTTCTTCTCTGCTGCTGCTCCTCACAGTGCTTACAGTTGTTACAGGTTCTG
 GTCATGCAAGCTCTACCCAGGTGGAGAAAAGGAGACTTCGGCTACCCAGAGAAGTTCAGTGCCAGCTCTAC
 TGAGAAGAATGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCGGTTCAGGCTCCTCCACCACT
 CAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACAGCTTCAGGTTTCAGCTGCCACCTGGGGACAGGATG
 TCACCTCGGTCCCAGTCACCAGGCCAGCCCTGGGCTCCACCACCCCGCCAGCCACGATGTACCTCAGCCCC
 GGACAACAAGCCAGCCCCGGGCTCCACCGCCCCCCCAGCCACGGTGTCACTCGGCCCCGGACACCAGGCCG
 CCCCCGGGCTCCACCGCCCCCCCAGCCACGGTGTCACTCGGCCCCGGACACCAGGCCGCCCCGGGCTCCA
 CCGCGCCCGCAGCCACGGTGTCACTCGGCCCCGGACACCAGGCCGCCCCGGGCTCCACCGCCCCCCCAGC
 CCATGGTGTCACTCGGCCCCGGACAACAGGCCCGCCTTGGCGTCCACCGCCCCCTCAGTCCACAATGTCACC
 TCGGCTCAGGCTCTGCATCAGGCTCAGCTTCTACTCTGGTGCAACGGCACCTCTGCCAGGGTACCACAA
 CCCCAGCCAGCAAGAGCACTCCATTCTCAATTCCAGCCACCACTCTGATACTCCTACCACCTTGCCAGCCA
 TAGCACCAGACTGATGCCAGTAGCACTCACCATAGCACGGTACCTCCTCTCACCTCCTCAATCACAGCACT
 TCTCCCCAGTTGTCTACTGGGGTCTCTTTCTTTTCTGCTCTTTTCACATTTCAAACCTCCAGTTTAATTCCT
 CTCTGGAAGATCCAGCACCGACTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTTGCAGATTTA
 TAAACAAGGGGGTTTTCTGGGCCTCTCCAATATTAAGTTTCAGGCCAGGATCTGTGGTGGTACAATTGACTCTG
 GCCTTCCGAGAAGGTACCATCAATGTCCACGACGTGGAGACACAGTTCAATCAGTATAAAACGGAAGCAGCCT
 CTCGATATAACCTGACGATCTCAGACGTCAGCGTGAGTGATGTGCCATTTCTTTCTCTGCCCAGTCTGGGGC
 TGGGGTGCCAGGCTGGGGCATCGCGTGCTGGTCTGGTCTGTGTTCTGGTTGCGCTGGCCATTGTCTATCTC
 ATTGCCCTTGGCTGTCTGTGAGTGCCGCCGAAAGAACTACGGGCAGCTGGACATCTTTCCAGCCCCGGGATACCT
 ACCATCCTATGAGCGAGTACCCACCTACCACACCCATGGGCGCTATGTGCCCCCTAGCAGTACCGATCGTAG
 CCCCTATGAGAAGGTTTTCTGCAGGTAATGGTGGCAGCAGCCTCTCTTACACAAACCCAGCAGTGGCAGCCACT
 TCTGCCAATTGATGGCGGCCGCTTACGTACATTCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCA
 ATGGCACTTGGTACTACTTTGACAGTTCAGGCTATATGCTTGACAGCCGCTGGAGGAAGCACACAGACGGCAA
 CTGGTACTGGTTTCGACAACCTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTT
 AACGAAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAG
 GCGCCATGCAATACATCAAGGCTAACTCTAAGTTCAATTGGTATCACTGAAGGCGTCATGGTATCAAATGCCTT
 TATCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAATGA

Fig. 28B. Sequência de polipéptido de MUC1-CPC (SEQ ID N° 50)

MTPGTQSPFFLLLLLVLTVVTSGSHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHSPGSGSSTT
 QGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNKPAPGSTAPPAHGVTSPAPDTRP
 PPGSTAPPAHGVTSPAPDTRPPPGSTAPAAHGVTSPAPDTRPAPGSTAPPAHGVTSPAPDNRPALASTAPPVHNVT
 SASGSASGSASTLVHNGTSARATTPASKSTPFSIPSHSDTPTTLASHSTKTDASSTHHSTVPPLTSSNHST
 SPQLSTGVSFFFLSFHISNLQFNSSLEDPSDYYQELQRDISEMFLQIYKGGFLGLSNIKFRPGSVVVQLTL

AFREGTINVHDTVETQFNQYKTEAASRYNLTISDVSVDVFPFSAQSGAGVPGWGIALLVLCVLVALAIVYL
IALAVCQCRKNYQQLDIFPARDTYHPMSEYPTYHTHGRYVPPSSTRSPYEKVSAGNGGSSLSYTNPAVAAT
SANLMAAAYVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYYF
NEEGAMKTGWVKYKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFIQSADGTGWYYLKPDGTLADRPE

Fig. 29. Sequência e construção de ss-CPC-MUC-1

Fig. 29A. Sequência de ADN (SEQ ID N° 51)

ATGGGATGGAGCTGTATCATCTCTTCTTGGTAGCAACAGCTACAGGTGTCCACTCCCAGGTCCAAATGGCGG
CCGCTTACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTT
TGACAGTTCAGGCTATATGCTTGACAGCCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAAC
TCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGA
AGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGCAATACATCAA
GGCTAACTCTAAGTTCATTGGTATCACTGAAGGCGTCATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGA
ACAGGCTGGTACTACCTCAAACAGACGGAACACTGGCAGACAGGCCAGAAATGACACCGGGCACCCAGTCTC
CTTTCTTCTGCTGCTGCTCCTCACAGTGCTTACAGTTGTTACAGGTTCTGGTCATGCAAGCTCTACCCAGG
TGGAGAAAAGGAGACTTCGGCTACCCAGAGAAGTTCAGTGCCAGCTCTACTGAGAAGAATGCTGTGAGTATG
ACCAGCAGCGTACTCTCCAGCCACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGG
CCCCGGCCACGGAACAGCTTCAGGTTAGCTGCCACCTGGGGACAGGATGTACCTCGGTCCCAGTCACCAG
GCCAGCCCTGGGCTCCACCACCCCGCCAGCCACGATGTACCTCAGCCCCGGACAACAAGCCAGCCCCGGGC
TCCACCGCCCCCCCCAGCCACGGTGTACCTCGGCCCCGGACACCAGGCCGCCCCGGGCTCCACCGCCCCC
CAGCCCCAGGTGTACCTCGGCCCCGGACACCAGGCCGCCCCGGGCTCCACCGCGCCCGCAGCCACGGTGT
CACCTCGGCCCCGGACACCAGGCCGCCCCGGGCTCCACCGCCCCCAGCCCATGGTGTACCTCGGCCCCG
GACAACAGGCCCGCTTGGCGTCCACCGCCCCCTCCAGTCCACAATGTACCTCGGCCTCAGGCTCTGCATCAG
GCTCAGCTTCTACTCTGGTGACAACGGCACCTCTGCCAGGGTACCACAACCCAGCCAGCAAGAGCACTCC
ATTCTCAATTTCCAGCCACCACTCTGATACTCTACCACCCTTGCCAGCCATAGCACCAGACTGATGCCAGT
AGCACTCACCATAGCACGGTACCTCCTCTCACCTCCTCCAATCACAGCACTTCTCCCCAGTTGTCTACTGGG
TCTCTTTCTTTTCTGTCTTTTACATTTCAAACCTCCAGTTTAATTCCTCTCTGGAAGATCCCAGCACCGA
CTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTTGCAGATTATATAACAAGGGGGTTTTCTGGG
CTCTCCAATATTAAGTTCAGGCCAGGATCTGTGGTGGTACAATTGACTCTGGCCTTCCGAGAAGGTACCATCA
ATGTCCACGACGTGGAGACACAGTTCAATCAGTATAAACGGAAGCAGCCTCTCGATATAACCTGACGATCTC
AGACGTACGCGTGAGTGATGTGCCATTTCCTTTCTCTGCCAGTCTGGGGCTGGGGTGCCAGGCTGGGGCATC
GCGCTGCTGGTGCTGGTCTGTGTTCTGGTTGCGCTGGCCATTGTCTATCTCATTGCCTTGGCTGTCTGTCAGT
GCCGCCGAAAGAACTACGGGCAGCTGGACATCTTTCCAGCCCCGGGATACCTACCATCCTATGAGCGAGTACCC
CACCTACCACACCCATGGGCGCTATGTGCCCCCTAGCAGTACCGATCGTAGCCCCCTATGAGAAGGTTTCTGCA
GGTAATGGTGGCAGCAGCTCTCTTACACAAACCCAGCAGTGGCAGCCACTTCTGCCAACTTGTAG

Fig. 29B. Sequência polipeptídica de sequência de proteína de ss-CPC-MUC-1 (SEQ ID N° 52)

MGWSCIILFLVATATGVHSQVQMAAAYVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFDN
SGEMATGWKKIADKWYYFNEEGAMKTGWVKYKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFIQSADG
TGWYYLKPDGTLADRPEMTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSM
TSSVLSSHSPGSGSSTTQGQDVTLPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNKPAPG
STAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPAAHGVTSAPDTRPAPGSTAPPAHGVTSAP
DNRPALASTAPPVHNVTASGSASGSASTLVHNGTSARATTTASKSTPFSIPSHHSDTPTTLASHSTKTDAS
STHHSTVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNSSLEDPSDYYQELQRDISEMFLQIYKQGGFLG
LSNIKFRPGSVVVQLTLAFREGTINVHDTVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGI
ALLVLVLCVLVALAIVYLIALAVCQCRRKNYGQLDIFPARDTYHPMSEYPTYHTHGRYVPPSSTDRSPYEKVSA
GNGGSSLSYTNPAVAATSANL