



República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial.

(21) **PI0708393-9 A2**

(22) Data de Depósito: 06/03/2007
(43) Data da Publicação: 31/05/2011
(RPI 2108)



★ B R P I 0 7 0 8 3 9 3 A 2 ★

(51) *Int.Cl.:*
A61K 39/29 2006.01
C07K 14/18 2006.01
C12N 7/01 2006.01
A61K 48/00 2006.01
C12N 5/10 2006.01

(54) Título: **PROTEÍNA DE FUSÃO NÃO ESTRUTURAL DO VÍRUS DA HEPATITE C**

(30) Prioridade Unionista: 09/03/2006 EP 06360006.8

(73) Titular(es): Transgene S.A.

(72) Inventor(es): Anne Fournillier, François Penin, Geneviève Inchauspe, Laurence Chatel

(74) Procurador(es): Tavares Propriedade Intelectual Ltda

(86) Pedido Internacional: PCT EP2007001922 de 06/03/2007

(87) Publicação Internacional: WO 2007/101657 de 13/09/2007

(57) Resumo: PROTEÍNA DE FUSÃO NÃO ESTRUTURAL DO VÍRUS DA HEPATITE C. A presente invenção revela uma proteína de fusão isolada compreendendo pelo menos três polipeptídeos NS originados de um vírus da hepatite C (VHC) configurados na referida proteína de fusão em uma ordem diferente da ordem na qual eles aparecem na configuração nativa. A presente invenção também se refere a uma molécula de ácido nucléico codificando tal proteína de fusão e a um vetor compreendendo tal molécula de ácido nucléico. A presente invenção também prevê partículas virais infecciosas e células hospedeiras compreendendo tal molécula de ácido nucléico ou tal vetor. A presente invenção também diz respeito a um método para produzir tal proteína de fusão de forma recombinante. Por fim, a presente invenção também prevê uma composição farmacêutica compreendendo tal proteína de fusão, molécula de ácido nucléico, vetor, partículas virais infecciosas e célula hospedeira, bem como o uso terapêutico desta composição para o tratamento ou prevenção de infecções pelo VHC, doenças associadas ao VHC e condições patológicas, assim como um método para induzir ou estimular uma resposta imunológica contra o VHC em um organismo hospedeiro.



Relatório Descritivo da Patente de Invenção
para **“PROTEÍNA DE FUSÃO NÃO ESTRUTURAL DO
VÍRUS DA HEPATITE C”**.

O VHC pertence à família Flaviviridae e constitui a principal causa de hepatite aguda e doenças crônicas do fígado. A infecção viral pelo VHC está associada a uma grande incidência (54-86%) de cronicidade, que, em 5 a 24% dos casos, pode evoluir para uma cirrose e, posteriormente, para um carcinoma hepatocelular num período de 20 a 30 anos (McHutchinson, 2004, *The American Journal of Managed Care* 10, S21-29). Atualmente, na Europa e nos Estados Unidos, 20 a 30% dos transplantes de fígado e 15 a 33% dos cânceres de fígado podem ser atribuídos à infecção pelo VHC. Em 1999, a Organização Mundial da Saúde estimou que cerca de 170 milhões de pessoas no mundo inteiro apresentam contágio crônico pelo VHC, e que a cada ano, 3 a 4 milhões de pessoas são contaminadas.

O VHC é um vírus encapsulado com um genoma de ácido ribonucléico (RNA) positivo, de fita única, de aproximadamente 9.600 nucleotídeos, cuja organização é comum a todas as linhagens e isolados do VHC. A tradução do genoma do VHC na fase 0 gera uma grande poliproteína de 3010 a 3030 aminoácidos, de acordo com o genótipo que é proteoliticamente clivado por proteases virais e celulares para produzir 10 proteínas virais. A terça parte amino-terminal da poliproteína codifica as proteínas estruturais do vírion: a proteína Núcleo (C) e as

glicoproteínas do envoltório E1 e E2. Após a região estrutural, vem uma proteína de membrana integral pequena, p7, que parece funcionar como um canal iônico. O restante do genoma codifica as proteínas não estruturais (NS) NS2, NS3, NS4A, NS4, NS4B, NS5A e NS5B, que mediam os processos intracelulares do ciclo de vida do vírus (Penin, 2004, Hepatology 39, 5-19). O VHC também codifica proteínas pequenas, chamadas de F (deslocamento do quadro de leitura, ou *frame shift*) ou ARFP (proteína do quadro de leitura alternativo), que podem ser produzidas pelo deslocamento do quadro ribossômico ou pela iniciação interna em um quadro de leitura +1 alternativo dentro do gene núcleo (vide Branch e col., 2005, Semin. Liver Dis. 25:105-117; and WO2004/069864). As seqüências de codificação de poliproteínas são flanqueadas em ambas as extremidades por duas regiões não traduzidas altamente conservadas (UTR), 5'UTR e 3'UTR, respectivamente. Um sítio de entrada ribossomal interno, localizado na 5'UTR, permite a fixação ribossomal para iniciação da tradução; quanto à 3'UTR, acredita-se que ela desempenhe um papel importante na iniciação da replicação viral.

A terapia padrão atual para pacientes com infecção crônica pelo VHC é uma combinação entre o interferon alfa peguilado (PEG-IFN- α) e a ribavirina (Fried e col., 2002, N. Engl. J. Med. 347, 975-982). No entanto, essa terapia apresenta alto custo por causa dos efeitos colaterais expressivos que levam ao término prematuro do tratamento em 10% dos casos, sendo inadequada para um grande número de pacientes (por exemplo, os

que apresentam cirrose hepática descompensada, doenças auto-imunes, histórico de depressão e gravidez). Outro dado muito importante: apenas 50% dos pacientes tratados apresentaram resposta (Falck-Ytter, e col., 2002, *Ann. Intern. Med.* 136, 288-292) e o índice de resposta é ainda menor nos pacientes infectados pelo genótipo 1 (27% a 35%).

Em geral, uma série de evidências experimentais vem se acumulando no sentido de enfatizar o papel crucial exercido pela imunidade das células-T no controle da infecção pelo VHC, principalmente as respostas mediadas pelas células-T CD4⁺ e CD8⁺ direcionadas às proteínas não estruturais (NS) (por exemplo, Francavilla e col., 2004, *Eur. J. Immunol.* 34, 427-37; Grakoui e col., 2003, *Science* 302, 659-662; Shoukry e col., 2003, *J. Exp. Med.* 197, 1645-1655; Thimme e col., 2001, *J. Exp. Med.* 194, 1395-1406; Lechner et al., 2000, *J. Exp. Med.* 191, 1499-1512; Gerlach, 1999, *Gastroenterology* 117, 933-941; Folgori e col., 2006, *Nat Med.* 12, 190-197). De fato, respostas vigorosas pelas células-T específicas ao VHC geralmente são observadas em pacientes que se recuperam espontaneamente da hepatite C aguda autolimitada; já as respostas das células-T são fracas e pouco direcionadas nos pacientes com contágio crônico. Por outro lado, o papel exercido pelos anticorpos neutralizantes (imunidade mediada pelas células-B) é muito menos claro (Logvinoff, 2004, *Proc. Natl. Acad. Sci. USA* 101, 10149-54).

Desde a descoberta do vírus da Hepatite C há cerca de 15 anos, ainda não foi possível criar uma vacina para o

VHC visto a dificuldade de se obter o crescimento eficiente do vírus em cultura celular e também pela falta de modelos de laboratório da infecção viral para pronta utilização. Entretanto, recentemente surgiram diversos candidatos à vacina (Barth e Baumert, 2004, Novel vaccine strategies, p. 214-227, In State of the Art of Hepatology: Molecular and Cell Biology eds Kluwer Academic Publishers; Inchauspe and Feinstone, 2003, Clin. Liver Dis. 7, 243-259), baseados, por exemplo, em peptídeos derivados do VHC (Cerny e col., 1995, J. Clin. Invest. 95, 521- 30), em antígenos do VHC produzidos de forma recombinante, em vacinas baseadas no DNA e no vírus (Brinster e col., 2001, Hepatology 34, 1206-1217; Forns e col., 2000, Hepatology 32, 618-625) e em partículas semelhantes ao vírus da Hepatite C (Baumert e col., 1999, Gastroenterology 117, 1397-1407; Jeong e col., 2004, J. Virol. 78, 6995-7003). Vários candidatos vêm sendo avaliados em modelos pré-clínicos e a primeira leva está entrando na etapa de ensaios clínicos. As vacinas mais avançadas atualmente na fase II utilizam proteínas do envoltório E1 com adjuvante (Innogenetics) e epítomos das células-T CD4+ e CD8+ (Intercell).

Uma abordagem energeticamente investigada durante a última década consiste no uso de poliproteínas associando vários domínios antigênicos como imunógenos. A grande maioria das poliproteínas da técnica anterior resulta da fusão entre polipeptídeos antigênicos, fragmentos ou epítomos originados de várias proteínas do VHC NS de modo a formar um

único polipeptídeo (com sorte) imunogênico. Por exemplo, a WO01/30812 e a WO03/031588 revelam a fusão dos polipeptídeos do VHC a partir de NS3 a NS5B para formar um único polipeptídeo englobando as principais proteínas NS. A
5 WO03/097677 descreve a fusão dos fragmentos antigênicos de 30 a 70 aminoácidos originados dos polipeptídeos NS3, NS4B e NS5B. A WO2004/046176 revela uma poliproteína compreendendo o Núcleo, NS3, NS4B e NS5B. A WO2004/111083 descreve o adenovírus e vetores MVA para co-
10 expressar NS3, NS4 e NS5B, em que NS3 e NS4 são expressos como uma proteína de fusão e NS5B é expresso de forma independente.

Deve-se observar que a configuração das poliproteínas da técnica anterior é como no contexto nativo, com
15 os vários componentes aparecendo na ordem em que eles ocorrem naturalmente no precursor nativo do VHC. Em particular, a NS3 é sucedida pela NS4, que é sucedida pela NS5. No entanto, a configuração “nativa” não é ideal em termos de expressão e imunogenicidade das poliproteínas resultantes.

20 A previsão é de que o VHC continue sendo um sério risco à saúde global por muitos anos por causa da natureza crônica e persistente da infecção, além de sua alta prevalência e pela morbidade significativa das doenças associadas. Sendo assim, há a real necessidade de se desenvolver mais polipeptídeos
25 imunogênicos, vetores de expressão, métodos de preparação destes e usos destes, com o objetivo de melhorar a prevenção e o

tratamento das infecções pelo VHC ou das doenças ou síndromes relacionadas ao VHC.

A presente invenção refere-se a proteínas de fusão envolvendo polipeptídeos não estruturais (NS) dispostos em uma configuração não nativa. Fragmentos específicos dos polipeptídeos NS do VHC (por exemplo, NS3, NS4A e NS5B e, como opção, NS4B) foram selecionados e configurados especificamente com o objetivo de otimizar a imunogenicidade e/ou o processo de produção recombinante das proteínas de fusão resultantes. Se comparado aos polipeptídeos NS nativos, as novas proteínas de fusão do VHC proporcionadas pela presente invenção ou suas seqüências de nucleotídeos codificadoras podem permitir o aprimoramento de uma resposta imune anti-VHC e/ou o aprimoramento de uma resposta citotóxica geral quando da introdução em um organismo hospedeiro e/ou o aprimoramento da produção de lotes clínicos. Sendo assim, a invenção representa um avanço significativo no atual tratamento e prevenção das infecções pelo VHC ou das doenças ou distúrbios associados ao VHC. Particularmente, ela pode ser utilizada para reforçar as terapias já existentes ou oferecer um tratamento alternativo aos pacientes com contágio crônico, principalmente àqueles que não reagem à terapia convencional.

Esse problema técnico é solucionado por meio das concretizações, conforme definidas nas reivindicações.

Esses e outros aspectos, características e vantagens da presente invenção ficarão evidentes na descrição das

concretizações presentemente preferidas da invenção, que será apresentada a seguir. Essas concretizações são apresentadas para fins de revelação.

Logo, em um primeiro aspecto, a presente
5 invenção oferece uma proteína de fusão isolada compreendendo pelo menos três polipeptídeos NS que se originam de um vírus da hepatite C (VHC), em que os referidos polipeptídeos NS são configurados na referida proteína de fusão em uma ordem diferente da ordem na qual eles aparecem na configuração nativa.

10 Como empregados neste documento por todo o pedido, os termos “um” e “uma” são utilizados no sentido de “pelo menos um(a)”, “pelo menos um(a) primeiro(a)”, “um(a) ou mais” ou “uma multiplicidade” dos compostos ou etapas referenciados, salvo em outros contextos. Por exemplo, o termo
15 “uma célula” inclui uma multiplicidade de células, inclusive uma mistura delas.

O termo “e/ou”, sempre que empregado neste documento, inclui o significado de “e”, “ou” e “toda e qualquer combinação diferente dos elementos conectados pelo aludido
20 termo”.

O termo “cerca de” ou “aproximadamente”, conforme utilizado neste documento, significa que algo está dentro de 20%, de preferência dentro de 10% e mais preferivelmente dentro de 5% de um dado valor ou faixa de
25 valores. Os termos “aminoácidos” e “resíduos” são sinônimos e englobam aminoácidos naturais, assim como análogos de

aminoácidos (por exemplo, aminoácidos não naturais, sintéticos e modificados, inclusive isômeros ópticos D ou L).

Os termos “polipeptídeo”, “peptídeo” e “proteína” são utilizados aqui de forma alternada para referir-se a
5 polímeros de resíduos de aminoácidos que compreendem dez ou mais aminoácidos ligados via ligações peptídicas. O polímero pode ser linear, ramificado ou cíclico e pode compreender análogos de aminoácidos e/ou de ocorrência natural e pode ser interrompido por não-aminoácidos. Como indicação geral, se o
10 polímero do aminoácido for longo (por exemplo, mais de 50 resíduos de aminoácidos), ele é preferivelmente chamado de polipeptídeo ou proteína.

Dentro do contexto da presente invenção, os termos “ácido nucléico”, “molécula de ácido nucléico”,
15 “polinucleotídeo” e “seqüência de nucleotídeos” são usados alternadamente e definem um polímero de qualquer tamanho, seja de moléculas de polideoxiribonucleotídeos (DNA) (por exemplo, cDNA, DNA genômico, plasmídeos, vetores, genomas virais, DNA isolado, sondas, oligonucleotídeos iniciadores e qualquer
20 mistura destes) ou de poliribonucleotídeos (RNA) (por exemplo, mRNA, RNA anti-sentido) ou poliribo-polideoxiribonucleotídeos misturados. Eles abrangem polinucleotídeos de fita dupla ou única, lineares ou circulares, naturais ou sintéticos. Além disso, um polinucleotídeo pode compreender nucleotídeos de ocorrência
25 não natural, como nucleotídeos metilados e análogos de nucleotídeos (vide a US 5,525,711, US 4,711,955 ou a EPA 302

175 como exemplos de modificações) e podem ser interrompidos por componentes que não sejam nucleotídeos. Caso estejam presentes, modificações no nucleotídeo podem ser realizadas tanto antes como após a polimerização.

5 Conforme usado neste documento, quando utilizado para definir produtos, composições e métodos, o termo “compreendendo” quer dizer que os produtos, composições e métodos incluem os componentes ou etapas referenciados, sem excluir outros. “Consistindo essencialmente de” significa
10 excluindo outros componentes ou etapas de qualquer significância essencial. Sendo assim, uma composição consistindo essencialmente dos componentes citados não excluiria contaminantes vestigiais e veículos aceitáveis do ponto de vista farmacêutico. “Consistindo de” significa excluindo mais que
15 microelementos de outros componentes ou etapas. Por exemplo, um polipeptídeo “consiste de” uma seqüência de aminoácidos quando o polipeptídeo não contém nenhum aminoácido fora a seqüência de aminoácidos citada. Um polipeptídeo “consiste essencialmente de” uma seqüência de aminoácidos quando tal
20 seqüência de aminoácidos está presente junto com apenas alguns resíduos de aminoácidos adicionais, normalmente de cerca de 1 a cerca de 50 resíduos adicionais. Um polipeptídeo “compreende” uma seqüência de aminoácidos quando a seqüência de aminoácidos faz parte pelo menos da seqüência de aminoácidos
25 final do polipeptídeo. Tal polipeptídeo pode ter de poucos até centenas de resíduos de aminoácidos adicionais. Tais resíduos de

aminoácidos adicionais podem desempenhar um papel no trânsito de polipeptídeos, facilitar a produção ou purificação dos polipeptídeos; prolongar a meia-vida, entre outras coisas. O mesmo pode ser aplicado a seqüências de nucleotídeos.

5 Como usado neste documento, o termo “isolado” refere-se a uma proteína, polipeptídeo, peptídeo ou um ácido nucléico que é purificado ou removido de seu ambiente natural. O termo “purificado” refere-se a uma proteína, polipeptídeo, peptídeo ou ácido nucléico que é separado pelo
10 menos dos outros componentes aos quais ele está normalmente associado.

 “VHC” significa “vírus da hepatite C”. Extensas análises filogenéticas levaram à classificação dos isolados do VHC em 7 genótipos principais (1 a 6) contendo diferentes
15 subtipos (a, b, c, etc.) (Simmons e col., 2005. Hepatology 42, 962-973). Exemplos de isolados do VHC do genótipo 1a incluem, sem a isto se limitar, o VHC-I (Choo e col., 1991, Proc. Natl. Acad. Sci. USA 88, 2451-2455), -J1 (Okamoto e col., 1992, Nucleic Acids Res. 20, 6410-6410) e -H (Inchauspe e col., 1991,
20 Proc. Natl. Acad. Sci. 88, 10292-10296). Exemplos de isolados do VHC do genótipo 1b incluem, sem a isto se limitar, VHC-JA (Kato e col., 1990, Proc. Natl. Acad., Sci. 87, 9524-9528) e BK (Takamizawa e col., 1991, J. Virol. 65, 1105-1113). Exemplos de isolados do VHC do genótipo 1c incluem, sem a isto se limitar, o
25 VHC-G9 (Okamoto e col., 1994, J. Gen. Virol. 45, 629-635). Exemplos de isolados do VHC do genótipo 2a incluem, sem a isto

se limitar, o VHC-G9 (Okamoto e col., 1991, J. Gen. Virol. 72, 2697-2704). Exemplos de isolados do VHC do genótipo 2b incluem, sem a isto se limitar, o VHC-J8 (Okamoto e col., 1992, Virology 188, 331-341). Exemplos de isolados do VHC do genótipo 2c incluem, sem a isto se limitar, o VHC-BEBE1 (Nako e col., 1996, J. Gen. Virol. 141, 701-704). Exemplos de isolados do VHC do genótipo 3a incluem, sem a isto se limitar, o VHC-NZL1 (Sakamoto e col., 1994, J. Gen. Virol. 75, 1761- 1768). Exemplos de isolados do VHC do genótipo 3b incluem, sem a isto se limitar, o VHC-Tr (Chayama e col., 1994, J. Gen. Virol. 75, 3623-3628). Exemplos de isolados do VHC do genótipo 4a incluem, sem a isto se limitar, o VHC-ED43 (Chamberlain e col., 1997, J. Gen. Virol. 78, 1341-1347). Exemplos de isolados do VHC do genótipo 5a incluem, sem a isto se limitar, o VHC-EUH480 (Chamberlain e col., 1997, Biochem. Biophys. Res. Commun. 236, 44-49). Exemplos de isolados do VHC do genótipo 6a incluem, sem a isto se limitar, o VHC-EUH480 (Chamberlain e col., 1997, Biochem. Biophys. Res. Commun. 234, 393-396).

O termo “fusão” ou “proteína de fusão”, conforme utilizado neste documento, refere-se à combinação com outro dentre pelo menos três polipeptídeos NS (ou fragmentos ou variantes destes) em uma cadeia de polipeptídeos. De preferência, a fusão entre os vários polipeptídeos NS é realizada por meios genéticos, isto é, fusionando-se em quadro as seqüências de nucleotídeos codificando cada um dos polipeptídeos NS. Entende-

se por “fusionado em quadro” que a expressão das seqüências de codificação fusionadas resulta em uma única proteína sem nenhuma terminador traducional entre cada um dos polipeptídeos NS. A fusão entre cada um dos polipeptídeos NS pode ser direta
5 ou por meio de um peptídeo de ligação. Como utilizado neste documento, o termo “direta” refere-se a uma fusão entre dois polipeptídeos NS sem nenhum resíduo de aminoácido adicional interveniente (por exemplo, os códons codificando um polipeptídeo NS são contíguos aos códons codificando o
10 polipeptídeo NS seguinte). Como alternativa, pode-se utilizar um peptídeo de ligação na junção de pelo menos dois polipeptídeos NS. A presença de um peptídeo de ligação pode facilitar a formação, enovelamento e/ou funcionamento correto da proteína de fusão. A presente invenção não se limita à forma, tamanho ou
15 número de seqüências de peptídeos de ligação empregado, sendo possível inserir várias cópias de uma seqüência de peptídeos de ligação na junção entre dois polipeptídeos NS. Peptídeos de ligação adequados de acordo com a invenção contêm de 3 a 30 aminoácidos de comprimento e são compostos de repetições de
20 resíduos de aminoácidos como a glicina, serina, treonina, asparagina, alanina e/ou prolina (vide, por exemplo, Wiederrecht e col., 1988, Cell 54, 841; Aumailly e col., 1990 FEBS Lett. 262, 82; e Dekker e col., 1993, Nature 362, 852). Exemplos representativos de peptídeos de ligação adequados incluem um
25 peptídeo de ligação Ser-Gli-Ser para conectar o polipeptídeo NS N-terminal ao segundo polipeptídeo NS e um peptídeo de ligação

Gli-Ser-Gli-Ser-Gli para conectar o segundo polipeptídeo NS ao terceiro polipeptídeo NS. Como alternativa, também é possível usar seqüências derivadas do VHC para conectar dois polipeptídeos NS um ao outro (por exemplo, a parte C-terminal de um polipeptídeo NS4A nativo estendendo aproximadamente da posição 1691 até aproximadamente a posição 1711 numerada em relação à poliproteína do VHC-I de comprimento total). “Pelo menos três” significa um número igual ou maior que três, com preferência especial por três ou quatro.

Conforme usado neste documento, o termo “polipeptídeo NS” refere-se a uma proteína não estrutural reconhecida na técnica, escolhida preferivelmente dentre o grupo que consiste dos polipeptídeos NS2, NS3, NS4A, NS4B e NS5B. Contudo, é preferível que a proteína de fusão não contenha nenhum polipeptídeo NS5A. No contexto da invenção, os polipeptídeos NS incluídos na proteína de fusão da invenção podem se originar independentemente de qualquer linhagem ou isolado do VHC identificado nos dias atuais, como os descritos acima em conexão com o termo “VHC”. O termo “originado” significa ser isolado, clonado, derivado ou relacionado. Sendo assim, de acordo com a presente invenção, cada um dos polipeptídeos NS incluídos na proteína de fusão pode se originar de um polipeptídeo NS nativo ou de um polipeptídeo NS modificado, como definido em detalhes abaixo.

Um polipeptídeo NS “nativo” refere-se a uma proteína, polipeptídeo ou peptídeo NS que pode ser encontrado ou

isolado de uma fonte na natureza, em vez de ser modificado artificialmente ou alterado no laboratório pelo homem. Sendo assim, o termo “polipeptídeo NS nativo” incluiria polipeptídeos NS de ocorrência natural e seus fragmentos. Tais fontes na natureza incluem amostras biológicas (por exemplo, sangue, plasma ou soro de pacientes contaminados pelo VHC ou que já foram contaminados por um VHC), células cultivadas, bem como materiais recombinantes (por exemplo, vírus ou genoma do VHC, bibliotecas genômicas ou de cDNA, plasmídeos contendo fragmentos do genoma do VHC, precursor pré-processado recombinante ou polipeptídeo NS processado, entre outros). As seqüências de nucleotídeos e aminoácidos de vários polipeptídeos/genes NS nativos foram descritas em publicações específicas e encontram-se disponíveis em bancos de dados especializados. Exemplos representativos de polipeptídeos NS nativos são apresentados na SEQ ID NO: 1-4 (A SEQ ID NO: 1-4 fornece as seqüências de aminoácidos dos polipeptídeos NS3, NS4A, NS4B e NS5B nativos da linhagem JA do VHC genótipo 1b). Entretanto, os polipeptídeos NS nativos não se limitam a essas seqüências exemplificativas. De fato, as seqüências de aminoácidos podem variar entre diferentes genótipos, subtipos e isolados do VHC, e esse escopo natural de variação genética está incluído no âmbito da invenção. Um ou mais dos polipeptídeos NS incluídos na proteína de fusão da invenção podem, de forma independente um do outro, incluir uma ou mais modificações de aminoácidos a partir das seqüências exemplificativas ou de outros

polipeptídeos NS nativos. Modificações podem ser geradas por meio de mutação e/ou pela adição de componentes químicos (por exemplo, alquilação, acetilação, amidação, fosforilação, entre outros) ou componentes de marcação. As mutações incluem
5 deleção, substituição ou adição de um ou mais resíduos de aminoácidos ou combinações dessas possibilidades. Quando várias modificações são contempladas, elas podem envolver resíduos consecutivos e/ou resíduos não consecutivos. Modificações podem ser feitas de diversas maneiras conhecidas
10 pelos versados na técnica. Por exemplo, a seqüência de nucleotídeos codificando o polipeptídeo NS pode ser modificada usando técnicas recombinantes de rotina, tal como corte enzimático seguido de modificação e ligação do fragmento definido, mutagênese direcionada ao sítio (por exemplo, usando o
15 sistema de mutagênese *in vitro* SulptorTM da Amersham, Les Ullis, France), mutagênese do DNA ou embaralhamento do DNA

Um polipeptídeo NS modificado preferido retém um alto nível de identidade de seqüência de aminoácidos com o polipeptídeo NS nativo correspondente, por exemplo, pelo
20 menos 75% de resíduos de aminoácidos idênticos ao longo da seqüência de aminoácidos de comprimento total ou um fragmento mais curto desta (por exemplo, de pelo menos 10, 15, 20, 25, 30, 40, 100 aminoácidos de comprimento). Mais especificamente, no contexto da invenção, o polipeptídeo NS modificado em uso na
25 invenção compartilha um grau de identidade com o polipeptídeo NS nativo correspondente maior do que 75%, de preferência

maior que 80%, ainda mais preferivelmente maior que 85%, preferivelmente maior que 90%, mais preferivelmente maior que 95%, ainda mais preferivelmente maior que 97%. A identidade de porcentagem entre os dois polipeptídeos é em função do número de posições idênticas compartilhadas pelas seqüências, levando em conta o número de intervalos que precisam ser introduzidos para o alinhamento ideal e o comprimento de cada intervalo. Vários programas de computador e algoritmos matemáticos encontram-se disponíveis na técnica para determinar as identidades de porcentagem entre as seqüências de aminoácidos, como por exemplo, o programa Blast (por exemplo, Altschul e col., 1997, *Nucleic Acids Res.* 25, 3389-3402; Altschul e col., 2005, *FEBS J.* 272, 5101-5109) disponível na NCBI. Por exemplo, qualquer um ou todos os polipeptídeos NS incluídos na proteína de fusão da invenção podem ser modificados de modo a representarem um genótipo ou subtipo específico, e, dessa forma, compreenderem uma seqüência de aminoácidos correspondendo a uma seqüência próxima ao consenso, que é geralmente determinada após o alinhamento de seqüência de vários NS de um genótipo ou subtipo específico. Qualquer um ou todos os polipeptídeos NS também podem ser modificados para obter um polipeptídeo NS com propriedades funcionais modificadas, tal como função enzimática reduzida e/ou capacidade reduzida de serem ancorados em uma membrana celular e/ou capacidade reduzida de serem processados após a tradução, como descrito a seguir. Os aminoácidos que são essenciais para tais propriedades

funcionais podem ser identificados por métodos de rotina, tal como por análise estrutural e funcional. Os versados na técnica podem determinar imediatamente o tipo de modificação(ões) capaz de reduzir ou interromper uma determinada atividade enzimática. Por exemplo, pode-se proceder com a mutagênese direcionada ao sítio ou técnicas PCR de modo a deletar ou substituir um ou mais aminoácidos dentro de uma região ativa da atividade enzimática (por exemplo, no sítio catalítico), de modo que a atividade enzimática nativa seja reduzida ou abolida significativamente. A redução ou falta de uma atividade biológica pode ser facilmente determinada em ensaios apropriados de acordo com a atividade enzimática a ser testada usando métodos conhecidos pelos versados na técnica. Os domínios de ancoragem de membrana são geralmente prognósticos com base em sua natureza hidrofóbica.

De acordo com a presente invenção, os pelo menos três polipeptídeos NS incluídos na proteína de fusão da invenção são configurados em uma ordem diferente da ordem na qual eles aparecem na configuração nativa. A configuração nativa é conhecida na técnica (vide o resumo geral fornecido na seção de introdução do presente pedido) e a ordem na qual os polipeptídeos NS aparecem é como encontrado em um precursor de poliproteína do VHC, isto é, a partir do N até o C-terminal NS2-NS3-NS4A-NS4B-NS5A-NS5B. De acordo com a configuração nativa:

o Um polipeptídeo NS2 sempre precede um polipeptídeo NS3 ou um polipeptídeo NS4A ou um polipeptídeo NS4B ou um polipeptídeo NS5A ou um polipeptídeo NS5B; e

5 o Um polipeptídeo NS3 sempre precede um polipeptídeo NS4A ou um polipeptídeo NS4B ou um polipeptídeo NS5A ou um polipeptídeo NS5B; e

o Um polipeptídeo NS4A sempre precede um polipeptídeo NS4B ou um polipeptídeo NS5A ou um polipeptídeo NS5B; e

10 o Um polipeptídeo NS4B sempre precede um polipeptídeo NS5A ou um polipeptídeo NS5B; e

o Um polipeptídeo NS5A sempre precede um polipeptídeo NS5B.

Na proteína de fusão da invenção, a
15 configuração não é nativa no sentido de que pelo menos um dos polipeptídeos NS aparece numa ordem diferente da ordem da configuração nativa. Sendo assim, se a proteína de fusão compreender um polipeptídeo NS3, um polipeptídeo NS4A e um polipeptídeo NS5B, a configuração nativa seria NS3-NS4A-
20 NS5B com NS3 no N-terminal e NS5B no C-terminal. Em contrapartida, uma configuração não nativa pode ser NS5B-NS3-NS4A, NS5B-NS4A-NS3, NS4A-NS3-NS5B, NS4A-NS5B-NS3 ou NS3-NS5B-NS4A.

Em particular, a proteína de fusão da invenção
25 compreende pelo menos um dentre os seguintes:

o Um polipeptídeo NS4A fusionado diretamente, ou através de um peptídeo de ligação, ao N-terminal de um polipeptídeo NS3;

o Um polipeptídeo NS3 fusionado diretamente,
5 ou através de um peptídeo de ligação, ao N-terminal de um polipeptídeo NS5B;

o Um polipeptídeo NS4B fusionado diretamente, ou através de um peptídeo de ligação, ao N-terminal de um polipeptídeo NS5B;

10 o Um polipeptídeo NS4A fusionado diretamente, ou através de um peptídeo de ligação, ao N-terminal de um polipeptídeo NS3 que é fusionado diretamente, ou através de um peptídeo de ligação, ao N-terminal de um polipeptídeo NS4B; e/ou

15 o Um polipeptídeo NS3 fusionado diretamente, ou através de um peptídeo de ligação, ao N-terminal de um polipeptídeo NS4B que é fusionado diretamente, ou através de um peptídeo de ligação, ao N-terminal de um polipeptídeo NS5B.

20 Em tais partes específicas da proteína de fusão da invenção, cada um dos polipeptídeos NS pode ser independentemente nativo ou modificado. Por exemplo, o polipeptídeo NS4A incluído na parte NS4A-NS3 pode ser nativo, ao passo que o polipeptídeo NS3 compreende pelo menos uma
25 das modificações descritas abaixo.

Em uma concretização, todos os polipeptídeos NS incluídos na proteína de fusão da invenção se originam da mesma linhagem ou isolado do VHC. Como alternativa, pelo menos dois dos polipeptídeos NS se originam de diferentes

5 linhagens ou isolados do VHC de modo a obter proteção contra uma variedade maior de genótipos do VHC. Também seria interessante adaptar a proteína de fusão à região geográfica específica onde ela será utilizada pela inclusão de pelo menos um polipeptídeo NS a partir de um VHC endêmico nesta região (por

10 exemplo, os genótipos Ia, Ib, 2 e 3 são os mais prevalentes na América do Norte, na Europa e na Ásia; o genótipo 4 é predominante na África do Norte e Central; o genótipo 5 foi, até agora, identificado principalmente na África do Sul e isolados do genótipo 6 foram encontrados principalmente no Vietnã e em

15 Hong Kong). Por exemplo, se a proteína de fusão compreender um polipeptídeo NS3, um polipeptídeo NS4A e um polipeptídeo NS5B, o polipeptídeo NS3 pode se originar de uma primeira linhagem do HCV, e os polipeptídeos NS4A e NS5B de uma segunda linhagem do VHC. Como alternativa, o polipeptídeo

20 NS4A pode se originar de uma primeira linhagem do VHC, e os polipeptídeos NS3 e NS5B de uma segunda linhagem do VHC. Como alternativa, o polipeptídeo NS5A pode se originar de uma primeira linhagem do VHC, e os polipeptídeos NS3 e NS4A de uma segunda linhagem do VHC. Como alternativa, cada um dos

25 polipeptídeos NS3, NS4A e NS5B se origina de diferentes linhagens do VHC. Quando a proteína de fusão da invenção

também compreende um polipeptídeo NS4B, ele pode se originar da mesma linhagem do VHC ou de uma linhagem do VHC diferente da dos outros polipeptídeos.

As concretizações preferidas da invenção estão
5 voltadas para uma fusão que compreende, ou consiste essencialmente de, ou consiste de um polipeptídeo NS4A, um polipeptídeo NS3, e um polipeptídeo NS5B com NS4A no N-terminal e NS5B no C-terminal (fusão NS4A-3-5B).

Como opção, a fusão também pode incluir um
10 polipeptídeo NS4B (ou fragmentos ou variantes deste) para melhorar ainda mais a atividade imunogênica da proteína de fusão resultante. O polipeptídeo NS4B é, de preferência, incluído entre os polipeptídeos NS3 e NS5B. Portanto, a presente invenção também se refere a uma proteína de fusão que compreende, ou
15 consiste essencialmente de, ou consiste de um polipeptídeo NS4A, um polipeptídeo NS3, um polipeptídeo NS4B e um polipeptídeo NS5B com NS4A no N-terminal e NS5B no C-terminal (fusão NS4A-3-4B-5B). Com exceção dos polipeptídeos NS3, NS4A, NS5B e do opcional NS4B, prefere-se, contudo, que
20 a proteína de fusão da invenção não contenha outros polipeptídeos do VHC, e em especial nenhum núcleo, e como discutido acima em conexão com os polipeptídeos NS, nenhum polipeptídeo NS5A.

Mais especificamente, o termo “polipeptídeo
25 NS3” refere-se a uma proteína, polipeptídeo ou peptídeo NS3 nativo obtido de qualquer linhagem HCV ou a uma proteína,

polipeptídeo ou peptídeo NS3 modificado conforme definido acima. O polipeptídeo NS3 em uso na invenção tem um comprimento de pelo menos 200 aminoácidos, para melhor proveito, pelo menos 300 aminoácidos, de preferência, pelo menos 400 aminoácidos, mais preferivelmente, pelo menos 500 aminoácidos, e ainda mais preferivelmente, pelo menos 600 aminoácidos. Para fins de ilustração, a proteína NS3 do VHC-1 nativa tem 631 aminoácidos de comprimento e está localizada aproximadamente nas posições 1027 a 1657 no precursor da poliproteína. Um polipeptídeo NS3 nativo compreende dois domínios ativos distintos, um domínio serina protease no N-terminal e um domínio helicase no C-terminal. A protease NS3 é responsável por processar o precursor da poliproteína do VHC nas junções NS3/NS4A, NS4A/NS4B, NS4B/NS5A e NS5A/NS5B, e exige o NS4A como cofator para a atividade proteolítica (Tomei e col., 1993, J. Virol. 67, 4017-4026) e para o enovelamento correto (Yao e col., 1999, Structure (London) 7 pp 1353). Um motivo protéico, consistindo de um ácido aspártico ou glutâmico na posição 6 e uma cisteína ou treonina na posição 1 a montante do sítio de clivagem e uma serina ou alanina na posição 1 a jusante do sítio de clivagem, é especificamente reconhecido pela protease NS3 do VHC. Acredita-se que a helicase NS3 seja importante para o enovelamento do intermediário do RNA de fita dupla durante a replicação viral (Tai e col., 1996, J. Virol. 70, 8477-8484).

Entende-se por “polipeptídeo NS4A” uma proteína, polipeptídeo ou peptídeo NS4A nativo de qualquer linhagem do VHC ou uma proteína, polipeptídeo ou peptídeo NS4A modificado conforme definido acima. O polipeptídeo NS4A em uso na invenção tem um comprimento de pelo menos 10 aminoácidos, para melhor proveito, pelo menos 11 aminoácidos, de preferência, pelo menos 12 aminoácidos, mais preferivelmente, pelo menos 13 aminoácidos. Entende-se por “polipeptídeo NS4B” uma proteína, polipeptídeo ou peptídeo NS4B nativo de qualquer linhagem do VHC ou uma proteína, polipeptídeo ou peptídeo NS4B modificado conforme definido acima. O polipeptídeo NS4B em uso na invenção tem um comprimento de pelo menos 20 aminoácidos, para melhor proveito, pelo menos 25 aminoácidos, de preferência, pelo menos 30 aminoácidos, mais preferivelmente, pelo menos 31 aminoácidos, e ainda mais preferivelmente, pelo menos 32 aminoácidos. Para fins de ilustração, a proteína NS4 do VHC-1 nativa (A-B) tem 315 aminoácidos de comprimento e está localizada aproximadamente nas posições 1658 a 1972 no precursor da poliproteína. O polipeptídeo NS4A (posições 1658 a 1711) é um co-fator na protease NS3, mas também é necessário para a estabilidade do NS3 e a localização na membrana do retículo endoplasmático (ER) (Failla e col., 1994, J. Virol. 68, 3753-3760). O polipeptídeo NS4B (posições 1712 a 1972) é uma proteína de membrana integral, cuja função ainda é desconhecida. Os algoritmos previstos sugerem a presença de quatro a seis

segmentos transmembrana. Sua expressão, no entanto, induz à formação de uma rede membranosa derivada do ER (Egger e col., 2002, J.Virol. 76, 5974-5984).

Entende-se por “polipeptídeo NS5B” uma
5 proteína, polipeptídeo ou peptídeo NS5B nativo de qualquer
linhagem do VHC ou uma proteína, polipeptídeo ou peptídeo
NS5B modificado conforme definido acima. O polipeptídeo
NS5B em uso na invenção tem um comprimento de pelo menos
200 aminoácidos, para melhor proveito, pelo menos 300
10 aminoácidos, de preferência, pelo menos 400 aminoácidos, mais
preferivelmente, pelo menos 500 aminoácidos, e ainda mais
preferivelmente, pelo menos 550 aminoácidos. Para fins de
ilustração, a proteína NS5B do VHC-1 nativa tem 591
aminoácidos de comprimento e está localizada aproximadamente
15 nas posições 2421 a 3011 no precursor da poliproteína. A proteína
NS5B nativa age como uma RNA polimerase dependente do
RNA, que supostamente permite a síntese de um intermediário de
RNA de fita negativa e cópias de progênie de fita positiva
(Lohman e col., 1997, J. Virol. 71, 8416-8428). Ela se encontra
20 associada à membrana do retículo endoplasmático por sua
seqüência de aminoácidos C-terminal 21 (Schmidt-Mende e col.,
2001, J. Biol. Chem. 276, 44052-44063). Para fins de clareza, as
extensões de aminoácidos citadas neste documento em conexão
com os polipeptídeos NS3, NS4A, NS4B e NS5B são
25 apresentadas com respeito a suas posições no precursor de
poliproteína do VHC-1 (conforme descrito por Choo e col., 1991,

Proc. Natl. Acad. Sci. USA 88, 2451-2455 ou em GenBank sob o número de acesso M62321). Entretanto, como discutido acima, a presente invenção também abrange polipeptídeos NS de outras linhagens e isolados do VHC, bem como polipeptídeos NS modificados. Portanto, salvo indicação em contrário pelo contexto, quando se faz referência neste documento a um polipeptídeo NS ou a um peptídeo deste por referência ao precursor de poliproteína do VHC-1, isso se refere ao polipeptídeo NS ou seu peptídeo do VHC-1, a um polipeptídeo NS ou seu peptídeo de outra linhagem ou isolado do VHC, ou a um polipeptídeo NS modificado ou seu peptídeo. Uma concretização preferida da presente invenção está voltada para uma proteína de fusão compreendendo os polipeptídeos NS3, NS4A, NS5B e o polipeptídeo NS4B opcional originados da linhagem do VHC JA genótipo 1b (Kato e col., 1990, Proc. Natl. Acad., Sci. 87, 9524-9528).

Em uma concretização, o polipeptídeo NS 3 incluído na proteína de fusão da presente invenção é modificado em comparação com o polipeptídeo NS3 nativo correspondente de modo a apresentar atividade significativamente reduzida da protease, e, assim, ser incapaz de mediar a clivagem da proteína de fusão em polipeptídeos individuais. A atividade da protease foi localizada na parte N-terminal do polipeptídeo NS3, aproximadamente a partir da posição 1027 até a posição 1207, numerada em relação à poliproteína do VHC-1 de comprimento total (aproximadamente a partir da posição 1 até a posição 181 do

polipeptídeo NS3 do VHC-JA ilustrado SEQ ID NO: 1). Mais especificamente, a atividade da protease NS3 foi atribuída aos resíduos da tríade catalítica His (H) na posição 1083, Asp (D) na posição 1107 e Ser (S) na posição 1165 (respectivamente, posições 57, 81 e 139 da SEQ ID NO: 1). Exemplos representativos de polipeptídeos NS3 deficientes em protease adequados são descritos em Bartenshlager e col., 1993, J. Virol. 67, 3835-3844 e Tomei e col., 1993, J. Virol. 67, 4017-4026. De preferência, uma vez que os resíduos Asp e Ser catalíticos residem dentro de um domínio antigênico previsto, o polipeptídeo NS3 em uso na presente invenção compreende a substituição do resíduo His na posição 1083 (correspondendo à posição 57 do SEQ ID NO:1) ou do resíduo de aminoácido localizado em uma posição equivalente de um polipeptídeo NS3 nativo de outro serótipo do VHC, para qualquer resíduo de aminoácido que não seja o His, com especial preferência por uma substituição para um resíduo Ala (H1083A). A interrupção da atividade da protease pode ser determinada usando ensaios bem conhecidos na técnica (Takeshita e col., 1997, Anal. Biochem. 247, 242-246; Kakiuchi e col., 1997 J. Biochem 122, 749-755; Sali e col., 1998, Biochemistry 37, 3392-3401; Kakiuchi e col., 1999, J. Virol. Meth. 80, 77-84).

Como alternativa ou em combinação, o polipeptídeo NS3 incluído na proteína de fusão da invenção é modificado se comparado ao polipeptídeo NS3 nativo correspondente, de modo a apresentar uma atividade da helicase

significativamente reduzida. Quatro resíduos estavam envolvidos na atividade da helicase associada ao NS3, respectivamente, Thr na posição 1295, Thr na posição 1437, Arg na posição 1490 e Arg na posição 1493, numerados em relação à poliproteína do VHC-I de comprimento total (que corresponde aos resíduos Thr nas posições 269 e 411 e aos resíduos Arg nas posições 464 e 467 do polipeptídeo NS3 do VHC-JA nativo apresentado no SEQ ID NO: 1). Exemplos representativos de polipeptídeos NS3 deficientes em helicase adequados são descritos em Kim e col. (1997, J. Virol. 71, 9400) e Lin and Kim (1999, J. Virol. 73, 8798-8807). De preferência, o polipeptídeo NS3 em uso na presente invenção compreende a substituição do resíduo Arg na posição 1490 (correspondendo à posição 464 do SEQ ID NO: 1) ou do resíduo de aminoácido localizado em uma posição equivalente de um polipeptídeo NS3 nativo de outro serótipo do VHC por qualquer resíduo de aminoácido que não seja o Arg e a substituição do resíduo Thr na posição 1295 (correspondendo à posição 269 do SEQ ID NO: 1) ou do resíduo de aminoácido localizado em uma posição equivalente de um polipeptídeo NS3 nativo de outro serótipo do VHC por qualquer resíduo de aminoácido que não seja o Thr, com especial preferência por uma substituição por um resíduo Ala em ambos os casos (T1295A e R1490A). A interrupção da atividade da helicase pode ser determinada usando ensaios bem conhecidos na técnica, por exemplo, avaliando-se a atividade de desenovelamento.

Mais preferivelmente, o polipeptídeo NS3 em uso na presente invenção é mutado para reduzir ou interromper tanto as atividades de protease como de helicase do polipeptídeo NS3 nativo e compreende as modificações discutidas acima em conexão com essas funções enzimáticas, com especial preferência pelas mutações H1083A, T1295A e R1490A.

Como alternativa ou em combinação com as modificações propostas acima em conexão com o polipeptídeo NS3, o polipeptídeo NS5B incluído na proteína de fusão da invenção é modificado se comparado ao polipeptídeo NS5B nativo correspondente de modo a apresentar uma atividade de RNA polimerase dependente do RNA significativamente reduzida. Dois resíduos Asp estavam envolvidos nessa função enzimática, respectivamente, nas posições 2640 e 2738 numeradas em relação à poliproteína do VHC-1 de comprimento total (que corresponde aos resíduos Asp nas posições 220 e 318 do polipeptídeo NS5B do VHC-JA apresentados no SEQ ID NO: 4). Exemplos representativos de polipeptídeos NS5B deficientes em polimerase adequados são descritos em Lohmann e col. (1997, J. Virol. 71, 8416-8428). De preferência, o polipeptídeo NS5B em uso na presente invenção compreende pelo menos a substituição do resíduo Asg na posição 2640 (correspondendo à posição 220 do SEQ ID NO: 4) ou do resíduo de aminoácido localizado em uma posição equivalente de um polipeptídeo NS5B nativo de outro serótipo do VHC por qualquer resíduo de aminoácido que não seja o Asp e/ou a substituição do resíduo Asp na posição

2738 (correspondendo à posição 318 do SEQ ID NO: 4) ou do resíduo de aminoácido localizado em uma posição equivalente de um polipeptídeo NS5B nativo de outro serótipo do VHC por qualquer resíduo de aminoácido que não seja o Asp, com especial
5 preferência por uma substituição por um resíduo Asn em ambos os casos. A interrupção da atividade da RNA polimerase pode ser determinada usando ensaios bem conhecidos na técnica, por exemplo, ensaios convencionais de replicase baseados na incorporação do substrato de nucleotídeos radioativo em um
10 produto de RNA nascente (vide, por exemplo, Behrens e col., 1996, EMBO J. 15, 12-22; Ferrari e col., 1999, J. Virol. 73, 1649-1654).

Em outra concretização, os polipeptídeos NS incluídos na proteína de fusão da invenção também podem
15 compreender modificações adicionais se comparado aos polipeptídeos NS nativos correspondentes. Modificações adequadas são aquelas benéficas ao processamento, estabilidade e solubilidade da proteína de fusão resultante, por exemplo, as que visam a inativar os sítios de clivagem potenciais, a ancoragem da
20 membrana e/ou a glicosilação, como descrito abaixo.

De forma vantajosa, a proteína de fusão da invenção não compreende um ou mais dos sítios de clivagem reconhecidos pelo NS3 normalmente presentes no precursor de poliproteína do VHC nas junções NS3/NS4A, NS4A/NS4B,
25 NS4B/NS5A e NS5A/NS5B, de modo a evitar o processamento em polipeptídeos NS individuais. Embora o polipeptídeo NS3

preferido em uso na presente invenção apresente uma atividade da protease significativamente reduzida, a inativação dos sítios de clivagem reconhecidos pelo NS3 introduz um nível adicional de segurança. A sequência consenso do sítio de clivagem é Asp/Glu-
5 X4-Cys/Thr-Ser/Ala, em que X é qualquer resíduo de aminoácido. Foi mostrado que a clivagem pelo polipeptídeo NS3 nativo ocorre após o resíduo Cys/Thr. De preferência, o polipeptídeo NS3 em uso na invenção não compreende o resíduo Thr ou Cys normalmente presente no C-terminal de um
10 polipeptídeo NS3 nativo (posição 1657 numerada em relação à poliproteína do VHC-1 que corresponde à posição 36 do SEQ ID NO:1 ou a uma posição equivalente de um polipeptídeo NS3 nativo de outro serótipo do VHC). Como alternativa ou em combinação, o polipeptídeo NS4A não compreende o resíduo Cys
15 normalmente presente no C-terminal de um polipeptídeo NS4A nativo (posição 1711 numerada em relação à poliproteína do VHC-1 que corresponde à posição 54 do SEQ ID NO:2 ou a uma posição equivalente de um polipeptídeo NS4A nativo de outro serótipo do VHC). Como alternativa ou em combinação, o
20 polipeptídeo NS4A opcional não compreende o resíduo Cys normalmente presente no C-terminal de um polipeptídeo NS4B nativo (posição 1972 numerada em relação à poliproteína do VHC-1 que corresponde à posição 261 do SEQ ID NO:3 ou a uma posição equivalente de um polipeptídeo NS4B nativo de
25 outro serótipo do VHC).

Em ainda outra concretização, a proteína de fusão da invenção pode ser modificada ainda mais de modo a excluir um ou mais domínios hidrofóbicos que normalmente estão envolvidos na ancoragem da membrana dos polipeptídeos NS4A, NS4B e/ou NS5B nativos. Seis domínios hidrofóbicos de ancoragem à membrana foram identificados no polipeptídeo NS4B nativo, sendo que um está presente na parte N-terminal do polipeptídeo NS4A nativo e o outro na parte C-terminal do polipeptídeo NS5B nativo. A deleção pelo menos parcial deles pode permitir melhorar a solubilidade da proteína de fusão e, dessa forma, facilitar sua produção por meios recombinantes. Isso também pode permitir limitar a citotoxicidade geral da proteína de fusão se comparado à administração dos polipeptídeos do VHC individuais em um determinado organismo hospedeiro.

Sob esse aspecto, o polipeptídeo NS4A é vantajosamente modificado pela deleção ou substituição de um ou mais resíduos de aminoácidos hidrófobos normalmente presentes dentro da parte N-terminal do polipeptídeo NS4A nativo. Um polipeptídeo NS4A preferido é deletado em até os 20 primeiros aminoácidos no N-terminal (por exemplo, deleção preferida aproximadamente a partir da posição 1658 até aproximadamente a posição 1677, numerada em relação à poliproteína do VHC-1 de comprimento total que corresponde à deleção aproximadamente a partir da posição 1 até aproximadamente a posição 20 do polipeptídeo NS4A do VHC-JA nativo ilustrado na SEQ ID NO: 2) enquanto, contudo, retém a parte do polipeptídeo NS4A nativo

estendendo aproximadamente a partir da posição 1678 até aproximadamente a posição 1690 numerada em relação à poliproteína do VHC-1 de comprimento total (aproximadamente a partir da posição 21 até aproximadamente a posição 33 do polipeptídeo NS4A do VHC-JA nativo apresentado no SEQ ID NO: 2). Ele pode ou não reter a parte C-terminal do polipeptídeo NS4A nativo, por exemplo, a parte que se estende aproximadamente da posição 1691 para aproximadamente a posição 1711 numerada em relação à poliproteína do VHC-1 de comprimento total (que corresponde aproximadamente a partir da posição 34 até aproximadamente a posição 54 do polipeptídeo NS4A do VHC-JA nativo apresentado no SEQ ID NO: 2). Mais preferivelmente, o polipeptídeo NS4A incluído na proteína de fusão da invenção consiste da parte de um polipeptídeo NS4A nativo que se estende a partir da posição 1678 até a posição 1690 numerada em relação à poliproteína do VHC-1 de comprimento total (que corresponde à parte NS4A que se estende da posição 21 à posição 33 do polipeptídeo NS4A do VHC-JA apresentado no SEQ ID NO: 2).

Como alternativa ou em combinação, um polipeptídeo NS5B preferido é deletado em 10 a 30 resíduos de aminoácidos normalmente presentes no C-terminal de um polipeptídeo NS5B nativo, com especial preferência pela deleção dos últimos 21 resíduos de aminoácidos C-terminal (por exemplo, da posição 2991 até a posição 3011, numerada em relação à

poliproteína do VHC-1 que corresponde da posição 571 à posição 591 do SEQ ID NO: 4).

Como alternativa ou em combinação, um polipeptídeo NS4B opcional preferido é deletado em pelo menos
5 um dos seis domínios hidrófobos de ancoragem à membrana. Mais preferivelmente, o polipeptídeo NS4B opcional é truncado por um ou mais aminoácidos tanto no N quanto no C-terminal do polipeptídeo NS4B nativo correspondente de modo a essencialmente reter o domínio antigênico interno. De
10 preferência, o polipeptídeo NS4B opcional em uso na invenção compreende, ou como alternativa, consiste essencialmente do segmento de aminoácido que se estende aproximadamente a partir da posição 1789 até aproximadamente a posição 1820, numeradas em relação à poliproteína do VHC-1 de comprimento total (que
15 corresponde aproximadamente à posição 78 até aproximadamente a posição 109 do polipeptídeo NS4B do VHC-J apresentado no SEQ ID NO. 3).

Em ainda outra concretização, a proteína de fusão da invenção é adicionalmente modificada de modo a
20 suprimir a potencial N-glicosilação quando da expressão em uma determinada célula hospedeira. Sob esse aspecto, um sítio de glicosilação consenso Asn-Val-Ser-Val foi identificado no polipeptídeo NS5B nativo a partir da posição 2789 à posição 2793 numeradas em relação à poliproteína do VHC-1 de comprimento
25 total (posição 369 à posição 372 da proteína NS5B do VHC-JA nativo apresentado no SEQ ID NO: 4). Um exemplo de mutação

sob esse aspecto consiste na substituição do resíduo Sew na posição 2791 numerada em relação à poliproteína do VHC-1 de comprimento total (posição 371 do SEQ ID NO: 4) por um resíduo diferente de um resíduo Ser, tal como um resíduo Gly.

5 De preferência, a proteína de fusão da invenção é imunogênica, no sentido de que ela é capaz de induzir ou estimular uma resposta imune específica ao antígeno, seja ela humoral ou celular, ou ambas, quando da introdução em um organismo hospedeiro. A presente invenção também abrange
10 proteínas de fusão criadas por engenharia genética de modo a aprimorar a imunogenicidade (por exemplo, por oxidação da ligação de dissulfeto). De preferência, o enovelamento dos vários polipeptídeos NS incluídos na proteína a de fusão da invenção é como o enovelamento dos polipeptídeos NS nativos. No contexto
15 da invenção, prefere-se que a proteína de fusão retenha um ou mais domínios antigênicos reconhecidos por um receptor de células-T. Diversos domínios antigênicos do VHC que podem ser retidos apropriadamente na proteína de fusão são descritos em publicações específicas (vide, por exemplo, Chien e col., 1992, Proc. Natl. Acad. Sci. USA 89, 10011-10015; Chien e col., 1993,
20 J. Gastroent. Hepatol. 8, S33-39 e WO03/097677).

De preferência, o polipeptídeo NS3 incluído na proteína de fusão da invenção retém as partes do polipeptídeo NS3 nativo que se estendem aproximadamente da posição 1038
25 até aproximadamente a posição 1082, aproximadamente da posição 1096 até aproximadamente a posição 1181, e

aproximadamente da posição 1244 até aproximadamente a posição 1274 numerada em relação à poliproteína do VHC-1 de comprimento total (por exemplo, aproximadamente a partir da posição 12 até aproximadamente a posição 56, aproximadamente a partir da posição 70 até aproximadamente a posição 155 e/ou aproximadamente a partir da posição 218 até aproximadamente a posição 248 do polipeptídeo NS3 do VHC-JA nativo revelado no SEQ ID NO: 1). De preferência, o polipeptídeo NS5B incluído na proteína de fusão da invenção retém a parte do polipeptídeo NS5B nativo que se estende aproximadamente da posição 2573 até aproximadamente a posição 2601 numeradas em relação à poliproteína do VHC-1 de comprimento total (por exemplo, aproximadamente a partir da posição 155 até aproximadamente a posição 182 do polipeptídeo NS5B do VHC-J nativo revelado no SEQ ID NO: 4). Quando a proteína de fusão da invenção compreende um polipeptídeo NS4B, ela de preferência retém a parte do polipeptídeo NS4B nativo que se estende aproximadamente da posição 1789 até aproximadamente a posição 1820 numeradas em relação à poliproteína do VHC-1 de comprimento total (por exemplo, aproximadamente a partir da posição 78 até aproximadamente a posição 109 do polipeptídeo NS4B do VHC-J nativo revelado no SEQ ID NO: 3).

A atividade imunogênica da proteína de fusão da invenção pode ser avaliada por uma série de técnicas de uso comum, como as descritas mais adiante em associação com o método da invenção ou ilustradas nos Exemplos. No contexto da

invenção, prefere-se que a atividade imunogênica da proteína de fusão seja maior em extensão e/ou natureza do que a atividade imunogênica usual proporcionada por um polipeptídeo NS nativo. Por exemplo, quando da introdução da proteína de fusão em um organismo hospedeiro, a resposta imune pode ser de maior intensidade, ou de natureza mais ampla (por exemplo, envolvendo mais células imunes como as células CD4+ e CD8+) ou seu escopo pode ser diferente (por exemplo, direcionado a um epítipo críptico não dominante) do que quando da introdução de um polipeptídeo NS nativo. Isso proporcionaria melhor efeito terapêutico, com isso possibilitando a redução dos regimes de dosagem e melhorando a qualidade de vida.

Um polipeptídeo NS3 preferido se origina da proteína NS3 do VHC-JA nativa apresentada no SEQ ID NO: 1 que é modificada pelo menos de tal forma que:

o ela retenha a parte que se estende da posição 12 até a posição 56, da posição 70 até a posição 155 e da posição 218 até a posição 248;

o ela compreenda a substituição do resíduo His na posição 57 por um resíduo de aminoácido diferente de um His, tal como um resíduo Ala;

o ela compreenda a substituição do resíduo Thr na posição 269 por um resíduo de aminoácido diferente de um Thr, tal como um resíduo Ala; e

o ela compreenda a substituição do resíduo Arg na posição 464 por um resíduo de aminoácido diferente de um Arg, tal como um resíduo Ala; e

o ela não compreenda o resíduo Thr na posição 631.

5 Ainda mais preferivelmente, o polipeptídeo NS3 em uso na invenção compreende, essencialmente consiste, ou alternativamente consiste da seqüência de aminoácidos apresentada no SEQ ID NO: 5.

Um polipeptídeo NS4A preferido se origina da
10 proteína NS4A do VHC-JA nativa apresentada no SEQ ID NO: 2 que é modificada pelo menos de tal forma que:

o ela contenha a parte do SEQ ID NO: 2 da posição 21 até a posição 33;

o ela não contenha a parte do SEQ ID NO: 2 da
15 posição 1 até a posição 20;

Ainda mais preferivelmente, o polipeptídeo NS4A tem uma seqüência de aminoácidos que consiste essencialmente da parte do SEQ ID NO: 2 da posição 21 até a posição 33 precedida por um resíduo Met iniciador. Mais
20 preferivelmente, o polipeptídeo NS4A consiste essencialmente da seqüência de aminoácidos apresentada no SEQ ID NO: 6 precedida por um resíduo Met iniciador.

Um polipeptídeo NS4B preferido opcional se origina da proteína NS4B do VHC-JA nativa apresentada no SEQ
25 ID NO: 3 que é modificada pelo menos de tal forma que:

o ela compreenda a parte que se estende do resíduo Ser na posição 78 até o resíduo Leu na posição 109;

o ela não compreenda o resíduo Cys na posição 261.

5 Ainda mais preferivelmente, o polipeptídeo NS4B opcional consiste essencialmente, ou alternativamente consiste da seqüência de aminoácidos conforme apresentada no SEQ ID NO: 7.

10 Um polipeptídeo NS5B preferido se origina da proteína NS5B do VHC-JA nativa apresentada no SEQ ID NO: 4 que é modificada pelo menos de tal forma que:

o ela compreenda a parte que se estende do resíduo Arg na posição 154 até o resíduo Leu na posição 182;

15 o ela compreenda a substituição do resíduo Asp na posição 220 por um resíduo de aminoácido diferente de um Asp, tal como um resíduo Asn;

o ela compreenda a substituição do resíduo Asp na posição 318 por um resíduo de aminoácido diferente de um Asp, tal como um resíduo Asn; e

20 o ela não compreende a parte que se estende do resíduo Trp na posição 571 até o resíduo Arg na posição 591.

 Ainda mais preferivelmente, o polipeptídeo NS5B opcional compreende, consiste essencialmente, ou alternativamente consiste da seqüência de aminoácidos conforme
25 apresentada no SEQ ID NO: 8.

As proteínas de fusão mais preferidas da invenção compreendem, ou consistem essencialmente ou alternativamente de uma seqüência de aminoácidos que é homóloga ou idêntica à seqüência de aminoácidos apresentada no SEQ ID NO: 9 ou 10. A seqüência apresentada no SEQ ID NO: 9 (Figura 1) corresponde à fusão dos polipeptídeos NS4A, NS3 e NS5B preferidos descritos acima, com um Met iniciador N-terminal na posição 1, o polipeptídeo NS4A estendendo-se a partir do resíduo de aminoácido 2 até o resíduo de aminoácido 14, o peptídeo de ligação estendendo-se a partir do resíduo de aminoácido 15 até o resíduo de aminoácido 17, o polipeptídeo NS3 estendendo-se a partir do resíduo de aminoácido 18 até o resíduo de aminoácido 647, o peptídeo de ligação estendendo-se a partir do resíduo de aminoácido 648 até o resíduo de aminoácido 652 e o polipeptídeo NS5B estendendo-se a partir do resíduo de aminoácido 653 até o resíduo de aminoácido 1222. A seqüência apresentada no SEQ ID NO: 10 (Figura 2) corresponde à fusão dos polipeptídeos NS4A, NS3, NS4B e NS5B mais preferidos descritos acima, com um Met iniciador N-terminal na posição 1, o polipeptídeo NS4A estendendo-se a partir do resíduo de aminoácido 2 até o resíduo de aminoácido 14, o peptídeo de ligação estendendo-se a partir do resíduo de aminoácido 15 até o resíduo de aminoácido 17, o polipeptídeo NS3 estendendo-se a partir do resíduo de aminoácido 18 até o resíduo de aminoácido 647, o polipeptídeo NS4B estendendo-se a partir do resíduo de aminoácido 648 até o resíduo de aminoácido 679 e o polipeptídeo

NS5B estendendo-se a partir do resíduo de aminoácido 680 até o resíduo de aminoácido 1249.

Também incluídos no âmbito da presente invenção estão novos fragmentos peptídicos das proteínas de fusão da invenção, em especial os apresentados no SEQ ID NO: 9 ou no SEQ ID NO: 10. Como utilizado neste documento, um fragmento compreende pelo menos 10, 15, 20, 50 ou mais resíduos de aminoácidos contíguos das proteínas de fusão reveladas neste documento. Tais fragmentos podem ser escolhidos baseando-se em sua capacidade de realizar uma função, por exemplo, de se ligar a um substrato ou de agir como um imunógeno. Fragmentos peptídicos adequados são geralmente aqueles que compreendem um domínio ou motivo protéico da proteína de fusão contendo novas estruturas imunogênicas. Domínios imunogênicos previstos são prontamente identificáveis por programas de computador, além de serem bem conhecidos e encontrarem-se prontamente disponíveis aos versados na técnica. Os fragmentos peptídicos da invenção podem ser sintetizados usando métodos de síntese protéica conhecidos.

A proteína de fusão da presente invenção e seus fragmentos peptídicos podem ser produzidos por qualquer método adequado, por exemplo, por técnicas de sintetização peptídica direta convencionais (por exemplo, Bodanszky, 1984 em Principles of peptide synthesis, Springer-Verlag), e pela tecnologia do DNA recombinante, conforme descrito abaixo em conexão com os vetores da invenção.

Sendo assim, a presente invenção também oferece uma molécula de ácido nucléico isolada codificando a proteína de fusão da invenção, com especial preferência ou moléculas de ácido nucléico que codificam uma proteína de fusão compreendendo, ou como alternativa, consistindo essencialmente de uma seqüência de aminoácidos como apresentada no SEQ ID NO: 9 ou no SEQ ID NO: 10 ou uma seqüência de aminoácidos homóloga ao SEQ ID NO: 9 ou ao SEQ ID NO: 10.

De preferência, as moléculas de ácido nucléico da invenção são otimizadas para proporcionar expressão de alto nível em uma célula hospedeira específica, por exemplo, em células hospedeiras de mamíferos, de levedura (por exemplo, *Saccharomyces cerevisiae*, *Saccharomyces pombe* ou *Pichia pastoris*) ou procarióticas (por exemplo, *E. coli*). De fato, foi observado que, quando mais de um códon está disponível para codificar para um dado aminoácido, os padrões de uso de códon dos organismos são extremamente pouco aleatórios (vide, por exemplo Wada e col., 1992, Nucleic Acids Res. 20, 2111-2118) e a utilização dos códons pode ser apreciavelmente diferente entre os diferentes hospedeiros (vide, por exemplo, Nakamura e col., 1996, Nucleic Acids Res. 24, 214-215). Uma vez que as seqüências de nucleotídeos de codificação NS usadas na invenção são de origem viral (VHC), elas podem ter um padrão de utilização de códon inapropriado para expressão eficiente nas células hospedeiras, como as células bacterianas ou eucarióticas inferiores.

Geralmente, a otimização do códon é realizada substituindo-se um ou mais códons “nativos” (por exemplo, o VHC) correspondendo a um códon pouco utilizado nesta célula hospedeira específica por um ou mais códons codificando o mesmo aminoácido que é utilizado com mais frequência. Isso pode ser obtido pela mutagênese convencional ou por técnicas químicas sintéticas (por exemplo, resultando em uma molécula de ácido nucléico sintética). Não é necessário substituir todos os códons nativos correspondendo aos códons pouco utilizados, visto que a expressão aumentada pode ser obtida mesmo com substituição parcial. Além do mais, alguns desvios da aderência estrita para utilização otimizada dos códons podem ser realizados para acomodar a introdução do(s) sítio(s) de restrição na molécula de ácido nucléico resultante.

Além da otimização da utilização dos códons, a expressão nas células hospedeiras pode ser melhorada ainda mais por modificações adicionais da seqüência de nucleotídeos. Por exemplo, a molécula de ácido nucléico da invenção pode ser modificada de modo a impedir a agregação de códons raros, não ideais, presentes em áreas concentradas e/ou suprimir ou modificar elementos de seqüência parcialmente negativos, pois se espera que estes influenciem negativamente os níveis de expressão. Tais elementos de seqüência negativa incluem, sem limitação, as regiões que apresentam um teor de GC altíssimo (>80%) ou baixíssimo (<30%); segmentos de seqüências ricas e AT ou ricas em GC; seqüências de repetições diretas ou invertidas

instáveis; estruturas secundárias de RNA; e/ou elementos regulatórios críticos internos, como caixas TATA internas, sítios-chi, sítios de entrada ribossomal e/ou sítios doadores/receptores de *splicing* (montagem de genes).

5 A molécula de ácido nucléico otimizada da invenção é, de preferência, capaz de expressar a proteína de fusão da invenção em uma determinada célula hospedeira em um nível superior, isto é, pelo menos 110%, vantajosamente pelo menos 150% e preferivelmente pelo menos 200%, se comparada ao nível
10 exposto por uma molécula de ácido nucléico envolvendo os genes do VHC nativos correspondentes sob condições idênticas (por exemplo, mesmo tipo de célula, mesmas condições de cultura, mesmo vetor de expressão, etc.). Os níveis de expressão podem ser avaliados por técnicas convencionais, como o *Western*
15 *blotting*, usando um anticorpo específico para um dos polipeptídeos do VHC incluídos na proteína de fusão da invenção.

De acordo com uma concretização preferida, a presente invenção proporciona moléculas de ácido nucléico que compreendem, ou como alternativa, consistem essencialmente de,
20 ou alternativamente consistem de uma seqüência de nucleotídeos que é homóloga, ou ainda mais preferivelmente, idêntica a qualquer uma das seqüências de nucleotídeos apresentadas no SEQ ID NO: 11-16. A molécula de ácido nucléico do SEQ ID NO: 11 codifica uma proteína de fusão NS4A-3-5B, cuja
25 seqüência de nucleotídeos foi otimizada para expressão em células de mamíferos (por exemplo, humanos). A molécula de

ácido nucléico do SEQ ID NO: 12 codifica uma fusão NS4A-3-4B-5B, cuja seqüência de nucleotídeos foi otimizada para expressão em mamíferos (por exemplo, células humanas). A molécula de ácido nucléico do SEQ ID NO: 13 codifica uma fusão NS4A-3-5B, cuja seqüência de nucleotídeos foi otimizada para expressão em leveduras (por exemplo, *P. pastoris*). A molécula de ácido nucléico do SEQ ID NO: 14 codifica uma fusão NS4A-3-4B-5B, cuja seqüência de nucleotídeos foi otimizada para expressão em leveduras (por exemplo, *P. pastoris*). A molécula de ácido nucléico do SEQ ID NO: 15 codifica uma fusão NS4A-3-5B, cuja seqüência de nucleotídeos foi otimizada para expressão em procariotos (por exemplo, *E. coli*). A molécula de ácido nucléico do SEQ ID NO: 16 codifica uma fusão NS4A-3-4B-5B, cuja seqüência de nucleotídeos foi otimizada para expressão em procariotos (por exemplo, *E. coli*). Obviamente, as seqüências podem ser equipadas com seqüências não codificadoras 5' e 3' proporcionando, se necessário, um iniciador Met na extremidade 5', um ou mais códons de PARADA na extremidade 3' e sítios de restrição adequados em ambas as extremidades para facilitar as etapas de clonagem. Exemplos de moléculas de ácido nucléico também são proporcionados nas Figuras 3 a 8.

A presente invenção também se refere a moléculas de ácido nucléico capazes de se hibridizarem sob condições rigorosas nas moléculas de ácido nucléico como definidas acima. As condições rigorosas de hibridização são

conhecidas pelos versados na técnica. Tipicamente, a hibridização é realizada em 6 vezes citrato de sódio/cloreto de sódio (SSC) a cerca de 45°C, seguido de uma ou mais lavagens em 0,2 vezes SSC, 0,1% SDS a 50-65°C. De preferência, tais ácidos nucleicos de hibridização compartilham um grau de identidade de sequência com as moléculas de ácido nucleico definidas acima maior do que 70% de todo o comprimento da sequência de nucleotídeos ou de um fragmento mais curto dela (por exemplo, de um comprimento de pelo menos 30, 45, 50, 60, 80, 100, 150, 200 nucleotídeos). Mais preferivelmente, o grau de identidade entre elas é maior do que 70%, para melhor proveito, maior do que 80%, sendo mais recomendável maior do que 85%, de preferência maior do que 90%, mais preferivelmente maior do que 95%, ainda mais preferivelmente maior do que 97%. Exemplos representativos de tais moléculas de ácido nucleico de hibridização incluem, sem limitação, moléculas de ácido nucleico complementares a pelo menos cerca de 20, 30, 40, 50, 100 ou 150 nucleotídeos contíguos incluídos em qualquer um dos SEQ ID NOs: 11-16 (por exemplo, moléculas de ácido nucleico anti-sentido). Variações das sequências exemplificativas, como códons degenerados, ou polimorfismo entre diferentes isolados/linhagens do VHC ou variações decorrentes de técnicas como o embaralhamento de DNA também são abrangidas pela presente invenção.

Outra concretização da invenção refere-se a fragmentos da molécula de ácido nucleico da invenção, por exemplo, fragmentos gerados pela endonuclease de restrição e

pela PCR. Tais fragmentos podem ser usados como sondas, oligonucleotídeos iniciadores ou fragmentos codificando uma parte imunogênica da proteína de fusão. A molécula de ácido nucléico da presente invenção pode ser gerada usando dados de
5 sequência acessíveis na técnica e as informações de seqüências fornecidas neste documento. A seqüência de DNA codificando para cada um dos polipeptídeos do VHC incluídos na proteína de fusão da presente invenção pode ser isolada diretamente de células contendo VHC, do cDNA e bibliotecas genômicas, de
10 genomas virais ou qualquer vetor da técnica anterior conhecido por incluí-la, pela biologia molecular convencional ou técnicas de PCR, e, se necessário, pode ser adicionalmente modificada por técnicas de mutagênese de rotina (por exemplo, para otimizar a expressão em uma célula hospedeira específica, para corresponder
15 a uma seqüência específica do genótipo ou subtipo, como descrito acima. A fusão das seqüências codificando cada um dos polipeptídeos do VHC pode ser realizada, por exemplo, por ligação em quadro tanto diretamente ou através de uma seqüência codificando uma ligação peptídica. Como alternativa, a molécula
20 de ácido nucléico da invenção também pode ser gerada por síntese química no processo automatizado (por exemplo, montada a partir de oligonucleotídeos sintéticos sobrepostos, como descrito, por exemplo, em Edge, 1981, Nature 292, 756; Nambair e col., 1984, Science 223, 1299; Jay e col., 5 1984, J. Biol. Chem.
25 259, 6311).

Também é proporcionado pela presente invenção um vetor compreendendo uma ou mais cópias da(s) molécula(s) de ácido nucléico da invenção. Por exemplo, pode ser vantajoso incluir, no mesmo vetor, moléculas de ácido nucléico codificando proteínas de fusão originadas de diferentes genótipos ou subtipos.

O termo “vetor”, como utilizado neste documento, refere-se tanto a vetores de expressão como aqueles que não são de expressão, e incluem tanto vetores virais como não virais, inclusive vetores extracromossomais (por exemplo, plasmídeos multicópia) e vetores integrantes projetados para serem incorporados ao(s) cromossomo(s) hospedeiro(s). De particular importância no contexto da invenção são os vetores para uso na terapia gênica (ou seja, capazes de distribuir a molécula de ácido nucléico a um organismo hospedeiro), assim como vetores de expressão para uso em variados sistemas de expressão.

Uma variedade de sistemas de vetor-hospedeiro pode ser usada para expressar a proteína de fusão da invenção, inclusive organismos procarióticos como bactérias (por exemplo, *E. coli* ou *Bacillus subtilis*) transformados com vetores de bacteriófagos, plasmídeos ou cosmídeos contendo a molécula de ácido nucléico de codificação da fusão; leveduras (por exemplo, *Saccharomyces cerevisiae*, *Saccharomyces pombe*, *Pichia pastoris*) transformadas com vetores de expressão de levedura contendo a molécula de ácido nucléico de codificação da fusão;

sistemas de células de insetos (por exemplo, células Sf 9) infectadas com vetores de expressão viral (por exemplo, baculovírus) contendo a molécula de ácido nucléico de codificação da fusão; sistemas de células de plantas infectados com vetores de expressão viral (por exemplo, vírus do mosaico da couve-flor CaMV; vírus do mosaico do tabaco TMV) ou transformados com vetores de expressão de plasmídeos (por exemplo, plasmídeo Ti) contendo a molécula de ácido nucléico de codificação da fusão; ou sistemas de células de mamíferos (por exemplo, células cultivadas) transfectados ou infectados com vetores de expressão de plasmídeo ou vírus contendo a molécula de ácido nucléico de codificação da fusão.

Vetores adequados para uso em sistemas procarióticos incluem, sem limitação, o pBR322 (Gibco BRL), pUC (Gibco BRL), pBluescript (Stratagene), p Poly (Lathe e col., 1987, Gene 57, 193-201), pTrc (Amann e col., 1988, Gene 69, 301-315); pET 11d (Studier e col., 1990, Gene Expression Technology: Methods in Enzymology 185, 60-89); pIN (Inouye e col., 1985, Nucleic Acids Res. 13, 3101-3109; Van Heeke e col., 1989, J. Biol. Chem. 264, 5503-5509); e vetores pGEX em que a molécula de ácido nucléico da invenção pode ser expressa em fusão com a glutathione S-transferase (GST). O plasmídeo pGEX-2T (Amersham Biosciences Product code: 27-4801-01, Genbank accession No. U13850) é particularmente adequado no contexto da invenção. Vetores adequados para expressão em levedura (por exemplo, *S. cerevisiae*) incluem, mas sem a isto se limitar,

pYepSec1 (Baldari e col., 1987, EMBO J. 6, 229-234), pMFa (Kujan e col., 1982, Cell 30, 933-943), pJRY88 (Schultz e col., 1987, Gene 54, 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.) and pTEF-MF (Dualsystems Biotech Product code: P03303). Os vetores adequados para expressão em células hospedeiras de mamíferos podem ser de origem viral ou não viral (por exemplo, DNA de plasmídeo). Vetores de plasmídeo adequados incluem, sem restrição, pREP4, pCEP4 (Invitrogene), pCI (Promega), pCDM8 (Seed, 1987, Nature 329, 840) e pMT2PC (Kaufman e col., 1987, EMBO J. 6, 187-195), pVAX e pgWiz (Gene Therapy System Inc; Himoudi e col., 2002, J. Virol. 76, 12735-12746). Além disso, os sistemas de vetor de hospedeiro utilizados no contexto da presente invenção também podem compreender um ou mais elemento(s) adicionais, permitindo a manutenção, propagação ou expressão da molécula de ácido nucléico da presente invenção na célula hospedeira. Tais elementos adicionais incluem, sem limitação, genes marcadores de modo a facilitar a identificação e o isolamento das células hospedeiras recombinantes (por exemplo, por complementação de uma auxotrofia celular ou por resistência a antibióticos), elementos de estabilização (por exemplo, seqüência *cer* como descrito em Summers and Sherrat, 1984, Cell 36, 1097-1103 e sistema DAP como descrito na US 5,198,343) e elementos integrantes (por exemplo, seqüências virtuais LTR e transpósons).

Genes marcadores adequados para expressão nas células hospedeiras procarióticas incluem a tetraciclina e

genes de resistência à ampicilina. Além disso, os genes de resistência podem ser usados para expressão em células hospedeiras de mamíferos, tal como a diidrofolato redutase (dhfr) que confere resistência ao metotrexato (Wigler e col., 1980, Proc. Natl. Acad. Sci. USA 77, 3567; O'Hare e col., 1981, Proc. Natl. Acad. Sci. USA 78, 1527); gpt, que confere resistência ao ácido micofenólico (Mulligan and Berg, 1981, Proc. Natl. Acad. Sci. USA 78, 2072); neo, que confere resistência ao aminoglicosídeo G-418 (Colberre-Garapin e col., 1981, J. Mol. Biol. 150, 1); zeo, que confere resistência à zeomicina, kana, que confere resistência à canamicina e hydro, que confere resistência à hydromicina (Santerre e col., 1984, Gene 30, 147). Os genes URA3 e LEU2 podem ser usados para expressão em sistemas de levedura, o que proporciona complementação dos mutantes de levedura ura3 ou leu2. Em particular, o gene URA3 a partir do qual seu promotor foi deletado pode ser vantajosamente utilizado.

Diversos sistemas de expressão baseado em vírus também podem ser utilizados no contexto da invenção, derivados de uma variedade de vírus diferentes (por exemplo, retrovírus, adenovírus, AAV, poxvírus, vírus da herpes, vírus do sarampo, vírus espumoso, entre outros). Como utilizado neste documento, o termo “vetor viral” abrange o DNA de vetor, bem como partículas virais geradas a partir dele. Os vetores virais podem ser capazes de replicação, ou podem ser geneticamente desabilitados de modo a serem defectivos para replicação ou incapazes para replicação. O termo “capaz de replicação”,

conforme usado neste documento, abrange vetores virais de replicação seletiva e condicionalmente replicativos criados por engenharia genética para replicarem de forma melhor ou seletiva em células hospedeiras específicas (por exemplo, células tumorais).

Tais vetores incluem, por exemplo, vetores adenovirais que possuem uma série de vantagens bem documentadas para transferência gênica ou para produção recombinante (para obter uma visão geral, consulte "Adenoviral vectors for gene therapy", 2002, Ed D. Curiel and J. Douglas, Academic Press). Os vetores adenovirais para uso de acordo com a presente invenção podem ser derivados de uma variedade de fontes humanas ou animais. Qualquer serótipo pode ser empregado dentre os serótipos do adenovírus 1 a 51, com especial preferência pelos adenovírus humanos 2 (Ad2), 5 (Ad5), 6 (Ad6), 11 (Ad11), 24 (Ad24) e 35 (Ad35). Os adenovírus citados encontram-se disponíveis na American Type Culture Collection (ATCC, Rockville, Md.), e foram o objeto de estudo de numerosas publicações descrevendo sua seqüência, organização e métodos de produção, permitindo aos técnicos aplicá-los (veja, por exemplo, a US 6,133,028; US 6,110,735; WO 02/40665; WO 00/50573; EP 1016711; Vogels e col., 2003, J. Virol. 77, 8263-8271).

Em uma concretização, o vetor adenoviral da presente invenção é capaz de replicação. Exemplos de tais vetores adenovirais capazes de replicação são bem conhecidos na técnica

e encontram-se prontamente disponíveis aos versados na técnica (vide, por exemplo, Hernandez-Alcoceba e col., 2000, Human Gene Ther. 11, 2009-2024; Nemunaitis e col., 2001, Gene Ther. 8, 746- 759; Alemany e col., 2000, Nature Biotechnology 18, 723-727). Vetores adenovirais capazes de replicação adequados para uso na invenção podem ser criados por engenharia genética a partir de um genoma de adenovírus do tipo selvagem por deleção do domínio E1A CR2 para anular a ligação ao Rb (vide, por exemplo, a WO00/24408) e/ou por substituição dos promotores E1 e/ou E4 nativos por promotores específicos à condição do tecido, tumor ou célula (vide, por exemplo, a US5,998,205, WO99/25860, US5,698,443, WO00/46355, WO00/15820 e WO01/36650).

Em outra concretização, o vetor adenoviral da invenção é defectivo para replicação (vide, por exemplo, a WO94/28152; Lusky e col., 1998, J. Virol 72, 2022-2032).

Vetores adenovirais defectivos para replicação preferidos são o E1-defectivo (vide, por exemplo, a US 6,136,594 e a US 6,013,638), com uma deleção E1 estendendo-se aproximadamente das posições 459 a 3328 ou aproximadamente das posições 459 a 3510 (por referência à sequência do adenovírus humano tipo 5 revelado no GeneBank sob o número de acesso M 73260 e em Chroboczek e col., 1992, Virol. 186, 280-285). A capacidade de clonagem pode ser melhorada ainda mais deletando parte(s) adicionais do genoma adenoviral (toda ou

parte da região E3 não essencial ou das outras regiões E2, E4 essenciais).

A molécula de ácido nucléico da presente invenção pode ser inserida em qualquer localização do genoma adenoviral. De preferência, ela é inserida em substituição da
5 região E1. Ela pode ser posicionada na orientação no sentido ou anti-sentido em relação à direção transcricional natural da região em questão.

Outros vetores virais adequados são derivados
10 dos poxvírus (vide, por exemplo, Cox e col. Em "Viruses in Human Gene Therapy" Ed J. M. Hos, Carolina Academic Press). No contexto da presente invenção, um vetor poxviral pode ser obtido de qualquer membro da Poxviridae, em especial do vírus canarypox, fowlpox e da varíola, sendo preferido este último.
15 Vírus da Varíola adequados incluem, sem limitação, a linhagem Copenhagen (Goebel e col., 1990, Virol. 179, 247-266 e 517-563; Johnson e col., 1993, Virol. 196, 381-401), a linhagem Wyeth e a linhagem Ankara modificada (MVA) (Antoine e col., 1998, Virol. 244, 365- 396). As condições gerais para construir o poxvírus
20 recombinante são bem conhecidas na técnica (vide, por exemplo, a EP 206 920; Mayr e col., 1975, Infection 3, 6-14; Sutter and Moss, 1992, Proc. Natl. Acad. Sci. USA 89, 10847-10851; US 6,440,422). A molécula de ácido nucléico da presente invenção é, de preferência, inserida dentro do genoma poxviral em um locus
25 não essencial. O gene da timidina quinase é particularmente apropriado para inserção nos vetores da rubéola Copenhagen

(Hruby e col., 1983, Proc. Natl. Acad. Sci USA 80, 3411-3415; Weir e col., 1983, J. Virol. 46, 530-537) e deleção II ou III para inserção no vetor MVA (Meyer e col., 1991, J. Gen. Virol. 72, 1031-1038; Sutter e col., 1994, Vaccine 12, 1032-1040).

5 De acordo com uma concretização preferida, os vetores da invenção compreendem a molécula de ácido nucléico da invenção em uma forma adequada para sua expressão em uma célula ou organismo hospedeiro, o que significa que a molécula de ácido nucléico é colocada sob o controle de uma ou mais
10 seqüências regulatórias, apropriadas ao vetor e/ou à célula hospedeira. Como utilizado neste documento, o termo “seqüência reguladora” refere-se a qualquer seqüência que permita, contribua ou module a expressão de uma molécula de ácido nucléico em uma determinada célula hospedeira, inclusive replicação,
15 duplicação, transcrição, *splicing* (montagem de genes), tradução, estabilidade e/ou transporte do ácido nucléico ou um de seus derivados (isto é, o mRNA) na célula hospedeira. Será apreciado pelos versados na técnica que a escolha das seqüências reguladoras pode depender de fatores tais como a célula
20 hospedeira, o nível de expressão desejado, etc.

O promotor é de especial importância, e promotores adequados úteis no contexto da presente invenção incluem promotores constitutivos que direcionam a expressão da molécula de ácido nucléico em muitos tipos de células
25 hospedeiras e aqueles que direcionam a expressão da seqüência de nucleotídeos apenas em certas células hospedeiras (por exemplo,

seqüências reguladoras específicas do fígado) ou em resposta a eventos específicos ou fatores exógenos (por exemplo, por temperatura, aditivo nutricional, hormônio ou outro ligante).

Promotores adequados para expressão na célula hospedeira de *E. coli* incluem, sem a isto se limitar, o promotor pL lamba bacteriófago, os promotores lac, TRP e pTAC induzível por IPTG. Promotores adequados para expressão em levedura incluem os promotores TEF (Mumberg e col., 1995, *Gene* 156, 119-122), PGK (Hitzeman e col., 1983, *Science* 219, 620-625), MF alfa (Inokuchi e col., 1987, *Mol. Cell. Biol.* 7, 3185-3193), CYC-I (Guarente e col., 1981, *Proc. Natl. Acad. Sci. USA* 78, 2199), GAL-I, GAL4, GAL10, PHO5, gliceraldeído-3-fosfato desidrogenase (GAP ou GAPDH), e álcool desidrogenase (ADH) (Denis e col., 1983, *J. Biol. Chem.* 258, 1165). Para expressão em vetores poxvirais, pode-se utilizar o promotor 7,5K, H5R, TK, p28, p11 ou K1L da varíola, os promotores sintéticos descritos em Chakrabarti e col. (1997, *Biotechniques* 23, 1094-1097), Hammond e col. (1997, *J. Virological Methods* 66, 135-138) e Kumar and Boyle (1990, *Virology* 179, 151-158), bem como promotores quiméricos precoces/tardios. Promotores adequados para expressão constitutiva em células de mamíferos incluem o promotor precoce imediato do citomegalovírus (CMV) (Boshart e col., 1985, *Cell* 41, 521-530), o promotor tardio principal do adenovírus, o promotor da fosfoglicero quinase (PGK) (Adra e col., 1987, *Gene* 60, 65-74), e o promotor da timidina quinase (TK) do vírus da herpes simplex (HSV)-1. Promotores

eucarióticos induzíveis por compostos fornecidos de forma exógena incluem, sem a isto se limitar, o promotor da metalotioneína induzível por zinco (MT) (Mc Ivor e col., 1987, Mol. Cell Biol. 7, 838-848), o promotor do vírus do tumor mamário de camundongo induzível por dexametasona (Dex) (MMTV), o sistema promotor da polimerase T7 (WO 98/10088), o promotor de inseto ecdisona (No e col., 1996, Proc. Natl. Acad. Sci. USA 93, 3346-3351), o promotor reprimível por tetraciclina (Gossen e col., 1992, Proc. Natl. Acad. Sci. USA 89, 5547-5551), o promotor induzível por tetraciclina (Kim e col., 1995, J. Virol. 69, 2565-2573), o promotor induzível por RU486 (Wang e col., 1997, Nat. Biotech. 15, 239-243 e Wang e col., 1997, Gene Ther. 4, 432-441) e o promotor induzível por rapamicina (Magari e col., 1997, J. Clin. Invest. 100, 2865-2872).

Promotores específicos do tecido também podem ser usados, principalmente aqueles que permitem visar as células infectadas pelo VHC para as quais se deseja obter benefício terapêutico. Promotores adequados incluem promotores específicos do fígado, como os da HMG-CoA redutase (Luskey, 1987, Mol. Cell. Biol. 7, 1881-1893); elemento regulador do esterol 1 (SRE-I; Smith e col., 1990, J. Biol. Chem. 265, 2306-2310); albumina (Pinkert e col., 1987, Genes Dev. 1, 268-277); fosfoenol piruvato carboxi quinase (PEPCK) (Eisenberger e col., 1992, Mol. Cell Biol. 12, 1396-1403); proteína reativa-C humana (CRP) (Li e col., 1990, J. Biol. Chem. 265, 4136-4142); glucoquinase humana (Tanizawa e col., 1992, Mol.

Endocrinology 6, 1070-1081); colesterol 7-alfa hidrolase (CYP-7) (Lee e col., 1994, J. Biol. Chem. 269, 14681-14689); alfa-1 antitripsina (Ciliberto e col., 1985, Cell 41, 531-540); proteína de ligação ao fator de crescimento tipo insulina (IGFBP- 1) (Babajko e col., 1993, Biochem Biophys. Res. Comm. 196, 480-486); transferrina humana (Mendelzon e col., 1990, Nucleic Acids Res. 18, 5717-5721); colágeno tipo I (Houglum e col., 1994, J. Clin. Invest. 94, 808-814) e genes FIX (US 5,814,716).

Promotores adicionais adequados para uso nesta invenção podem ser obtidos de genes que são preferivelmente expressos em células tumorais proliferativas, como os promotores do gene da alfa-foetoproteína superexpresso em cânceres hepáticos (Kanai e col., 1997, Câncer Res. 57, 461-465), a transcriptase reversa telomerase (TERT) (WO99/27113, WO 02/053760 e Horikawa e col., 1999, Cancer Res. 59, 826), e elemento responsivo à hipoxia (HRE).

Os versados na técnica irão apreciar que os elementos reguladores controlando a expressão da molécula de ácido nucléico da invenção podem compreender ainda elementos adicionais para iniciação, regulação e/ou terminação adequada da transcrição (por exemplo, seqüências de sinal de localização poliA), transporte de mRNA (por exemplo, seqüências de sinal de localização nuclear), processamento (por exemplo, sinais de *splicing* (montagem de genes)), e estabilidade (por exemplo, íntrons e seqüências 5' e 3' não codificadoras), tradução (por exemplo, sinal peptídico, propeptídeo, seqüências líderes

tripartidas, sítios de ligação ribossomal, seqüências de Shine-Dalgarno, etc.) na célula ou organismo hospedeiro e etapas de purificação (por exemplo, uma marca).

Por exemplo, um peptídeo de sinal, eventualmente em combinação com um pró-peptídeo, pode ser usado para facilitar a secreção da proteína de fusão no meio de cultura. O peptídeo de sinal é geralmente inserido no N-terminal da proteína de fusão imediatamente após o iniciador Met, e, se necessário, o pró-peptídeo é inserido a jusante do peptídeo de sinal. A escolha dos peptídeos de sinal e/ou pró-peptídeos é ampla e acessível aos versados na técnica.

Exemplos de pró-peptídeos/peptídeos de sinal apropriados para a presente invenção incluem, sem a isto se limitar, as seqüências de peptídeos de sinal do fator correspondente (MF) alfa (Kurjan e Herskowitz, 1982, Cell 30, 933-934 e US 5,879,926); invertase (WO84/01153); PHO5 (DK 3614/83); YAP3 (protease aspártica de levedura 3; WO95/02059); e BAR1 (WO87/02670). Durante a entrada no ER, o peptídeo de sinal é clivado fora do polipeptídeo precursor em um sítio de processamento. O sítio de processamento pode compreender qualquer seqüência peptídica que seja reconhecida por uma enzima proteolítica da célula hospedeira. Exemplos de sítios de processamento preferidos incluem, sem a isto se limitar, qualquer combinação dos dois resíduos básicos Lys e Arg (com especial preferência por Lis-Arg), que são clivados pela endopeptidase codificada pelo gene KEX2 da *Saccharomyces cerevisiae* (vide

Fuller e col., 1986, Microbiology, 273-278) ou a protease equivalente de outras espécies de levedura (vide Julius e col., 1983, Cell 32, 839-852).

Uma marca peptídica (por exemplo, uma
5 seqüência curta de peptídeos capaz de ser reconhecida pelo anti-
soro disponível) pode ser utilizada para facilitar a purificação da
proteína de fusão recombinante a partir da célula hospedeira ou
do sobrenadante da cultura. De preferência, a marca é inserida no
C-terminal da proteína de fusão. Exemplos de marcas incluem,
10 sem limitação, a marca His (por exemplo, composta de 6 ou mais
resíduos de histidina), seqüência de peptídeos da glutathione-S-
transferase (GST) do *S. mansoni*, proteína de ligação à maltose
(MPB) da *E. coli*, fosfatase alcalina humana, o octapeptídeo
FLAG e a marca-e (US ,686,152).

15 Uma concretização preferida da invenção está
voltada para um plasmídeo para expressão em levedura,
compreendendo a molécula de ácido nucléico da invenção (por
exemplo, a seqüência de nucleotídeos do SEQ ID NO: 13 e 14)
colocada sob o controle do promotor TEF e uma seqüência
20 terminadora transcricional adequada (por exemplo: terminador
eye do citocromo) que adicionalmente compreende origens de
replicação para e células hospedeiras de levedura e *E. coli* (por
exemplo, ORI e 2 μ , respectivamente), marcadores de seleção
apropriados para expressão em *E. coli* (por exemplo, gene de
25 resistência à ampicilina) e levedura (por exemplo URA3 para
autotrofia e leu2-3).

Outra concretização preferida da invenção está voltada a um plasmídeo pGEX adequado para expressão em *E. coli*, compreendendo a molécula de ácido nucléico da invenção (por exemplo, a seqüência de nucleotídeos do SEQ ID NO: 15 e 16) colocada sob o controle do promotor pTAC induzível por IPTG e uma seqüência terminadora transcricional apropriada que adicionalmente compreende uma origem adequada de replicação para *E. coli* (por exemplo, ORI) e um marcador de seleção adequado (por exemplo, gene de resistência à ampicilina).

Ainda outra concretização preferida da invenção está voltada a um vetor adenoviral defectivo para replicação para expressão em células de mamíferos, compreendendo a molécula de ácido nucléico da invenção (por exemplo, a seqüência de nucleotídeos do SEQ ID NO: 11 e 12) colocada sob elementos reguladores apropriados. De preferência, o vetor adenoviral defectivo para replicação é um genoma Ad5 E1 e E3 deletados, com a deleção do E1 estendendo-se desde aproximadamente o nucleotídeo 459 até aproximadamente o nucleotídeo 3511 e a deleção do E3 desde aproximadamente o nucleotídeo 28592 até aproximadamente o nucleotídeo 30470. Um vetor adenoviral preferido compreende seqüências Ad5 desde aproximadamente o nucleotídeo 1 até aproximadamente o nucleotídeo 458, o cassete de expressão contendo, de 5' a 3', a seqüência reforçadora/promotor de resposta imediata CMV, um íntron quimérico da β -globina/IgG humana (como encontrado no vetor PCI disponível na Promega), a seqüência codificando qualquer

uma das duas formas de poliproteína (SEQ ID NO: 11 ou 12) e o sinal de poliadenilação tardio SV40 seguido da sequência Ad5 desde aproximadamente o nucleotídeo 3512 até aproximadamente os nucleotídeos 28592 e desde aproximadamente o nucleotídeo 30470 até aproximadamente o nucleotídeo 35935. Se necessário, o vetor da invenção pode adicionalmente compreender um ou mais transgene(s), por exemplo, um gene de interesse a ser expresso junto com a molécula de ácido nucléico da invenção em uma célula ou organismo hospedeiro. De preferência, a expressão do transgene tem uma atividade terapêutica ou protetora em uma doença ou distúrbio associado ao VHC. Transgenes adequados incluem, sem limitação, os genes de codificação da citocina (por exemplo, IL-2, IL-7, IL-15, IL-18, IL-21, IFN γ), genes de codificação da toxina (e.g. ricina, toxina da difteria, toxina da cólera), genes suicidas, e, principalmente, os genes codificando o TK HSV-1 (Caruso e col., 1993, Proc. Natl. Acad. Sci. USA 90, 7024-7028), que será utilizado em combinação com o pró-fármaco acyclovir ou ganciclovir, citosina deaminase (CDase), uracil fosforibosil transferase (UPRTase), que devem ser utilizadas em combinação com o prodrug 5-fluorocitosina (5-FC). O produto gênico *FCU-I* (descrito na WO 99/54481) é particularmente útil neste contexto. Se um transgene for utilizado, ele é colocado sob o controle de elementos reguladores apropriados e pode ser inserido em qualquer localização do vetor da invenção ou em um vetor independente que é utilizado em combinação com o vetor da invenção.

A inserção da molécula de ácido nucléico da invenção em um *backbone* de vetor pode ser realizada pela biologia molecular de rotina, por exemplo, como descrito em Sambrook e col. (2001, Molecular Cloning- A Laboratory Manual, Cold Spring Harbor Laboratory). A inserção em um
5 vetor adenoviral ou em um vetor poxviral pode ser realizada através da recombinação homóloga, como descrito respectivamente em Chartier e col. (1996, J. Virol. 70, 4805-4810) e Paul e col. (2002, Cancer gene Ther. 9, 470-477).

10 A presente invenção também abrange vetores (por exemplo, DNA de plasmídeo) complexados em lipídeos ou polímeros para formar estruturas particuladas, como lipossomas, lipoplexos ou nanopartículas. Tais tecnologias encontram-se disponíveis na técnica (vide, por exemplo, Arangoa e col., 2003,
15 Gene Ther. 10: 5-14; Eliaz e col., 2002, Gene Ther. 9, 1230-1237 e Betageri e col., 1993, "Liposome drug delivery systems", Technomic Publishing Company, Inc).

Em outra concretização, a presente invenção oferece partículas virais infecciosas compreendendo as moléculas
20 de ácido nucléico ou vetores previamente descritos da presente invenção.

Geralmente, tais partículas virais são produzidas por meio de um processo compreendendo as etapas de:

(a) introduzir o vetor viral da invenção em uma
25 linhagem celular adequada,

(b) cultivar a referida linhagem celular sob condições adequadas de modo a permitir a produção da referida partícula viral infecciosa,

(c) recuperar a partícula viral infecciosa
5 produzida da cultura da referida linhagem celular, e

(d) opcionalmente purificar a referida partícula viral infecciosa recuperada. Quando o vetor viral é defectivo, as partículas infecciosas são geralmente produzidas em uma linhagem celular de complementação ou pelo uso de um vírus
10 auxiliar, que fornece *in trans* os genes virais funcionais. Por exemplo, linhagens celulares adequadas para complementar os vetores adenovirais deletados em E1 incluem as células 293 (Graham e col., 1997, J. Gen. Virol. 36, 59-72), bem como as células PER-C6 (Fallaux e col., 1998, Human Gene Ther. 9,
15 1909-1917). Células apropriadas para a propagação dos vetores poxvirais são as células de aves, e mais preferivelmente, fibroblastos de embrião de galinha (CEF) primários preparados a partir de embriões de galinha obtidos de ovos fertilizados.

As partículas virais infecciosas podem ser
20 recuperadas do sobrenadante da cultura ou das células após a lise. Elas podem ser adicionalmente purificadas de acordo com técnicas convencionais (cromatografia, ultracentrifugação em um gradiente de cloreto de cézio, como descrito, por exemplo, na WO96/27677, WO98/00524, WO98/22588, WO98/26048,
25 WO00/40702, EP1016700 e na WO00/50573).

A presente invenção também engloba vetores ou partículas virais que foram modificados para permitir o direcionamento preferencial a uma célula-alvo específica (vide, por exemplo, Wickam e col., 1997, J. Virol. 71, 8221-8229; Arnberg e col., 1997, Virol. 227, 239-244; Michael e col., 1995, Gene Therapy 2, 660-668; WO94/10323; WO02/96939 e a EP 1 146 125). Um aspecto característico dos vetores direcionados e das partículas virais da invenção é a presença, em sua superfície, de um ligante capaz de reconhecer e se ligar a um componente celular e exposto à superfície, tal como um marcador de célula específica (por exemplo, uma célula infectada pelo VHC), um marcador de tecido específico (por exemplo, um marcador específico do fígado), bem como um antígeno viral (por exemplo, o VHC). Exemplos de ligantes adequados incluem anticorpos ou seus fragmentos direcionados a um domínio antigênico do VHC. O direcionamento celular pode ser realizado pela inserção genética do ligante em um polipeptídeo presente na superfície do vírus (por exemplo, fibra adenoviral, penton, pIX ou produto gênico pi 4 da varíola).

A invenção refere-se também a células hospedeiras que compreendem as moléculas de ácido nucléico, vetores ou partículas virais infecciosas da invenção descritas no presente documento. No contexto da presente invenção, o termo “célula hospedeira” deve ser interpretado de maneira ampla, sem limitação a uma organização específica no tecido, órgão ou nas células isoladas. Tais células podem ser de um tipo único de

células ou um grupo de tipos diferentes de células e abrangem linhagens celulares cultivadas, células primárias e células proliferativas.

No contexto da invenção, as células hospedeiras

5 incluem células procarióticas, células eucarióticas inferiores, como a levedura, e outras células eucarióticas, como as células de insetos, as células de plantas e células de mamíferos (por exemplo, humanos ou não humanos). Células hospedeiras de *E. coli* preferidas incluem, sem limitação, a BL21 de *E. coli*

10 (disponível na Amersham Biosciences). Células hospedeiras de levedura preferidas incluem, sem limitação, *S. cerevisiae*, *S. pombe*, *Pichia pastoris*, com especial preferência por KEX-2 expressando células de levedura, como a linhagem TGY47.1 (ou seu conversor Leu^+ isogênico TGY73.4) descrita na EP 607 080 e

15 aquelas descritas na US 5,879,926, US 5,162,208 e US 6,103,515. Células hospedeiras de mamíferos preferidas incluem, sem limitação, BHK (células renais de hamster), CV-1 (linhagem celular renal de Maçado Africano), células COS (por exemplo, COS-7), células do ovário de hamster chinês (CHO), células

20 NIH/3T3 de camundongo, células HeLa e células Vero. As células hospedeiras também englobam células de complementação, capazes de complementar pelo menos uma função defectiva de um vetor defectivo para replicação da invenção (por exemplo, vetor adenoviral), tal como as células 293

25 e PERC.6. De acordo com uma concretização específica da invenção, a célula hospedeira pode ser adicionalmente

encapsulada. A tecnologia de encapsulamento celular já foi descrita anteriormente (Tresco e col., 1992, ASAIO J. 38, 17-23; Aebischer e col., 1996, Human Gene Ther. 7, 851-860).

Ainda outro aspecto da presente invenção
5 consiste em um método para produzir por recombinação a proteína de fusão, empregando os vetores, partículas virais infecciosas e/ou células hospedeiras da invenção. O método para produção da proteína de fusão compreende (a) introduzir um vetor ou uma partícula viral infecciosa da invenção em uma célula
10 hospedeira adequada para produzir uma célula hospedeira transfectada ou infectada, (b) cultivar *in-vitro* a referida célula hospedeira transfectada ou infectada sob condições adequadas ao crescimento da célula hospedeira, (c) recuperar a proteína de fusão da cultura celular, e (d) opcionalmente, purificar a proteína
15 de fusão recuperada.

Espera-se que os versados na técnica conheçam os numerosos sistemas de expressão disponíveis para produzir as proteínas de fusão da invenção em células hospedeiras apropriadas e os métodos para introduzir um vetor ou uma
20 partícula viral infecciosa em uma célula hospedeira. Tais métodos incluem, sem a isto se limitar, microinjeção (Capechi e col., 1980, Cell 22, 479-488), transfecção mediada por CaPO_4 (Chen and Okayama, 1987, Mol. Cell Biol. 7, 2745-2752), transfecção mediada por DEAE-dextrano, eletroporação (Chu e col., 1987,
25 Nucleic Acid Res. 15, 1311-1326), fusão por lipofecção/lipossoma (Feigner e col., 1987, Proc. Natl. Acad. Sci.

USA 84, 7413-7417), bombardeio de partículas (Yang e col., 1990, Proc. Natl. Acad. Sci. USA 87, 9568-9572), biobalística, transdução, infecção viral, bem como administração direta em um organismo hospedeiro por variados meios.

5 Os vetores da invenção podem ser utilizados em associação com reagentes de transfecção a fim de facilitar a introdução do vetor na célula hospedeira, tais como polímeros policationicos (por exemplo, quitosana, polimetacrilato, PEI, etc.) e lipídeos cationicos (por exemplo, DC-Chol/DOPE, lipofectina
10 transfectam, que agora se encontra disponível através da Promega). Além do mais, como discutido acima, tecnologias de DNA recombinante podem ser utilizadas para melhorar a expressão da molécula de ácido nucléico na célula hospedeira, por exemplo, utilizando-se vetores de grande número de cópia,
15 substituindo ou modificando-se uma ou mais seqüências reguladoras transcricionais (por exemplo, promotor, seqüência reforçadora, entre outros), otimizando a utilização dos códons da molécula de ácido nucléico de codificação da fusão para a célula hospedeira, e suprimindo as seqüências negativas que podem
20 desestabilizar a transcrição.

As células hospedeiras da presente invenção podem ser cultivadas em bioreatores de fermentação, frascos e placas de Petri convencionais. A cultura pode ser realizada a uma temperatura, pH e teor de oxigênio apropriados para uma
25 determinada célula hospedeira. Não será feita uma tentativa de descrever em detalhes os vários métodos conhecidos para a

produção de proteínas em células procariotas e eucariotas. A produção da proteína de fusão pode ser periplásmica, intracelular ou secretada fora da célula hospedeira (por exemplo: no meio de cultura).

5 Quando a proteína de fusão não é secretada fora da célula hospedeira produtora, ou quando ela não é secretada completamente, ela pode ser recuperada por procedimentos de lise padrão, incluindo gelo-degelo, sonicação, rompimento mecânico, uso de agentes indutores da lise, entre outros. Caso seja secretada,
10 ela pode ser recuperada diretamente do meio de cultura. A proteína de fusão pode então ser purificada por métodos de purificação consolidados, inclusive precipitação por sulfato de amônio, extração de ácido, eletroforese em gel, filtração e métodos cromatográficos (por exemplo, fase reversa, exclusão por
15 tamanho, troca iônica, afinidade, fosfocelulose, interação hidrófoba, hidroxilapatita, ou cromatografia líquida de alta eficiência). As condições e tecnologia utilizadas para purificar uma proteína de fusão específica da invenção dependerão de fatores como a carga líquida, o peso molecular, a hidrofobia, a
20 hidrofília, e serão facilmente reconhecidas pelos versados na técnica. Além disso, o nível de purificação irá depender do fim a que se destina. Também é entendido que, dependendo da célula hospedeira produtora, as proteínas de fusão podem ter vários padrões de glicosilação, ou podem ser não glicosiladas (por
25 exemplo, quando produzidas em bactérias).

Em outro aspecto, a presente invenção oferece uma composição compreendendo a proteína de fusão, a molécula de ácido nucléico, o vetor, a partícula viral infecciosa, a célula hospedeira da invenção (também chamada aqui de “agente ativo”) ou qualquer combinação destes (por exemplo, combinação de proteínas de fusão ou vetores/partículas virais codificando proteínas de fusão de diferentes genótipos ou subtipos). De preferência, a composição é uma composição farmacêutica que compreende, além de uma quantidade eficaz do(s) agente(s) ativo(s) do ponto de vista terapêutico, um veículo aceitável do ponto de vista farmacêutico. Como usado neste documento, uma “quantidade eficaz do ponto de vista terapêutico” é uma dose suficiente para o alívio de um ou mais sintomas normalmente associados à doença ou condição que se deseja tratar. Com respeito ao uso profilático, esse termo significa uma dose suficiente para impedir ou retardar a consolidação de uma doença ou condição. Por exemplo, uma quantidade eficaz do ponto de vista terapêutico para induzir uma resposta imune poderia ser a quantidade necessária para causar a ativação do sistema imunológico (por exemplo, resultando no desenvolvimento de uma resposta anti-VHC).

Como usado neste documento, o termo “veículo aceitável do ponto de vista farmacêutico” pretende incluir todo e qualquer veículo, solvente, diluente, excipiente, adjuvante, meio de dispersão, revestimento, agente antibacteriano e antifúngico e

agente de retardo de absorção, entre outros, compatível com a administração farmacêutica.

De forma apropriada, a composição farmacêutica da invenção compreende um diluente apropriado para uso humano ou animal. De preferência, ele é isotônico, hipotônico ou fracamente hipertônico, e tem uma força iônica relativamente baixa. Exemplos representativos incluem água estéril, solução salina fisiológica (por exemplo, cloreto de sódio), solução de Ringer, glicose, trealose ou soluções de sacarose, solução de Hank, e outras soluções salinas fisiologicamente equilibradas (vide, por exemplo, a edição mais recente de Remington : The Science and Practice of Pharmacy, A. Gennaro, Lippincott, Williams&Wilkins). A composição da invenção é adequadamente tamponada de modo a ser apropriada para uso humano a um pH fisiológico ou ligeiramente básico (por exemplo, entre aproximadamente o pH 7 e pH 9). Tampões adequados incluem, sem limitação, tampão fosfato (por exemplo, PBS), tampão bicarbonato e/ou tampão Tris.

A composição também pode conter outros excipientes aceitáveis do ponto de vista farmacêutico para obter propriedades farmacêuticas ou farmacodinâmicas desejadas, incluindo, por exemplo, modificar ou manter o pH, osmolalidade, viscosidade, clareza, cor, esterilidade, estabilidade, coeficiente de dissolução da formulação, modificar ou manter a liberação ou absorção em um organismo humano ou animal, promover o transporte através da barreira sanguínea ou penetração em um

órgão específico (por exemplo, o fígado). Excipientes adequados incluem aminoácidos.

Veículos aceitáveis do ponto de vista farmacêutico incluídos na composição da invenção também
5 devem permitir a preservação de sua estabilidade sob as condições de fabricação e estocagem a longo prazo (isto é, de pelo menos um mês) sob congelamento (por exemplo, -70°C , -20°C), temperaturas refrigeradas (por exemplo, 4°C) ou à temperatura ambiente. Sob esse aspecto, formulações que são particularmente
10 adaptadas à composição da invenção incluem:

o 1M sacarose, 150 mM de NaCl, 1mM de MgCl_2 , 54 mg/l de Tween 80, 10 mM de Tris pH 8.5 (especialmente quando o agente ativo é um vetor adenoviral), e

o 10 mg/ml de manitol, 1 mg/ml de HSA, 20
15 mM de Tris, pH 7.2, e 150 mM de NaCl 25

o solução salina fisiológica

Além disso, a composição da invenção pode compreender um ou mais adjuvantes adequados para aplicação sistêmica ou mucosal em humanos. De preferência, o adjuvante é
20 capaz de estimular a imunidade para a composição da invenção, em especial uma imunidade mediada por célula T, por exemplo, através dos receptores do tipo “toll” (TLR), como o TLR-7, TLR-8 e o TLR-9. Exemplos representativos de adjuvantes úteis incluem, sem limitação, alume, emulsão de óleo mineral, tal como
25 os adjuvantes completos e incompletos de Freund (IFA), lipopolissacarídeo ou um derivado deste (Ribi e col., 1986,

Immunology and Immunopharmacology of Bacterial Endotoxins, Plenum Publ. Corp., NY₃ p407-419), saponinas como a QS21 (Sumino e col., 1998, J.Virol. 72, 4931-4939; WO 98/56415), compostos de imidazol-quinolina como Tmiquimod (Suader, 5 2000, J. Am Acad Dermatol. 43, S6-S11) e o composto relacionado S- 27609 (Smorlesi, 2005, Gene Ther. 12, 1324-1332), oligodeoxinucleotídeos de citosina fosfato guanosina como o CpG (Chu e col., 1997, J. Exp. Med. 186: 1623; Tritel e col., 2003, J. Immunol. 171 : 2358-2547) e peptídeos catiônicos como 10 o IC-31 (Kritsch e col., 2005, J. Chromatogr Anal. Technol Biomed Life Sci 822, 263-270).

A composição da presente invenção pode ser administrada por uma variedade de modos de administração, inclusive administração sistêmica, tópica e localizada. A injeção 15 pode ser realizada por qualquer meio, por exemplo, por injeção subcutânea, intradérmica, intramuscular, intravenosa, intraperitonal, intratumoral, intravascular, intra-arterial ou direta em uma artéria (por exemplo, por infusão na artéria hepática) ou em uma veia que alimenta o fígado (por exemplo, injeção na veia 20 porta). As injeções podem ser realizadas com seringas e agulhas convencionais, ou qualquer outro dispositivo apropriado disponível na técnica. Como alternativa, a composição da presente invenção pode ser administrada via uma rota mucosal, tal como a rota oral/alimentar, nasal, intratraqueana, intrapulmonar, 25 intravaginal ou intrarectal. A administração no trato respiratório pode ser realizada através da nebulização ou aerossolização de

gotículas, spray ou composições em pó seco usando um recipiente pressurizado (por exemplo, com um propelente adequado, como diclorodifluormetano, propano, nitrogênio, entre outros) ou em um dispensador não pressurizado. A administração tópica também
5 pode ser realizada utilizando-se meios transdérmicos (por exemplo, adesivo e similares). No contexto da invenção, as administrações intramuscular e subcutânea constituem as rotas preferidas.

A composição da invenção pode ser em várias
10 formas, por exemplo, sólida, líquida ou congelada. Composições sólidas (por exemplo, em pó seco ou liofilizadas) podem ser obtidas por um processo envolvendo secagem a vácuo ou secagem por congelamento. Para administração mucosal, as composições podem ser formadas como cápsulas e grânulos
15 gastrorresistentes para administração via oral, supositórios para administração retal ou vaginal, talvez com acentuadores de absorção úteis para aumentar o tamanho de poro das membranas mucosais. Tais acentuadores de absorção normalmente são substâncias que possuem similaridades estruturais com os
20 domínios de fosfolipídeo das membranas mucosais, como o deoxicholato de sódio, glicocholato de sódio, dimetil-beta-ciclodextrina, lauroil-1-lisofosfatidilcolina).

A dosagem apropriada pode ser adaptada em função de vários parâmetros, em especial do modo de
25 administração; da composição empregada; da idade, da saúde, e do peso do organismo hospedeiro; da natureza e do grau dos

sintomas; do tipo de tratamento simultâneo; da frequência do tratamento; e/ou da necessidade de prevenção ou terapia. Outros apuramentos sobre os cálculos necessários para determinar a dosagem apropriada para tratamento são efetuados rotineiramente por um profissional, considerando as circunstâncias relevantes. Para orientação geral, a dosagem adequada para uma composição compreendendo um vírus (por exemplo, partículas de adenovírus) varia de cerca de 10^5 a 10^{13} ui (unidades de infecção), de preferência cerca de 10^7 a 10^{12} ui. Uma composição baseada em plasmídeos de vetor pode ser administrada em doses entre 10 µg e 20 mg, de preferência entre 100 µg e 2 mg. Uma composição protéica pode ser administrada em uma ou mais doses entre 10 ng e 20 mg, com especial preferência por um dosagem de cerca de 0,1 µg a cerca de 2 mg da proteína terapêutica por kg de peso corporal. A administração pode ocorrer em uma única dose ou em uma dose repetida uma ou várias vezes após certo intervalo de tempo.

A composição farmacêutica da invenção pode ser empregada em métodos para tratar uma variedade de doenças e condições patológicas, principalmente aquelas associadas a uma infecção pelo VHC. Como utilizado neste documento, o termo “tratamento” abrange tanto profilaxia quanto terapia. Ele é de especial utilidade no tratamento de infecções persistentes pelo VHC e do câncer hepático em pacientes infectados pelo VHC. O termo “câncer” abrange quaisquer condições cancerosas, inclusive tumores difusos ou localizados, metástase, pólipos cancerosos,

bem como lesões preneoplásticas (por exemplo, cirrose). De preferência, quando da introdução em um organismo hospedeiro de acordo com as modalidades descritas neste documento, a composição da invenção proporciona um benefício terapêutico ao hospedeiro tratado. O termo “organismo hospedeiro” pretende incluir qualquer animal, principalmente mamíferos, assim como murinos, ratos, bovinos, suínos, caninos, felinos, eqüinos, macacos ou pacientes humanos, por exemplo, um humano infectado pelo VHC. O benefício terapêutico pode ser evidenciada por diversas formas, por exemplo, uma diminuição da viremia do VHC detectada no sangue, plasma ou soro de um indivíduo infectado, se comparado a antes do tratamento, e/ou pela detecção de uma resposta imune anti-VHC (por exemplo, produção de anticorpos anti-VHC e/ou imunidade mediada por células T) ou pelo retardo dos sintomas associados a uma infecção pelo VHC (por exemplo, retardo no desenvolvimento de uma cirrose ou câncer hepático), ou por uma redução dos sintomas de inflamação/esteatose/fibrose do fígado, geralmente associados à infecção pelo VHC ou por uma resposta melhor do indivíduo às terapias convencionais.

Sendo assim, a presente invenção também abrange o uso da proteína de fusão, molécula de ácido nucléico, vetor, partícula viral infecciosa, célula hospedeira ou composição da invenção para a preparação de um fármaco destinado ao tratamento ou prevenção de infecções pelo VHC, doenças

associadas ao VHC e condições patológicas, de acordo com as modalidades descritas acima.

A presente invenção também oferece um método para o tratamento ou prevenção de um organismo humano ou animal, compreendendo administrar ao referido organismo uma quantidade terapeuticamente eficaz da proteína de fusão, da molécula de ácido nucléico, do vetor, da partícula viral infecciosa, da célula hospedeira ou da composição da invenção.

Se desejado, o método da invenção pode ser realizado em combinação com uma ou mais modalidades terapêuticas convencionais (por exemplo, radiação, quimioterapia e/ou cirurgia). O uso de múltiplas abordagens terapêuticas oferece ao paciente uma intervenção de base mais extensa. Em uma concretização, o método da invenção pode ser precedido ou seguido de uma intervenção cirúrgica. Em outra concretização, ele pode ser precedido ou seguido de radioterapia (por exemplo, radiação gama). Os versados na técnica serão capazes de formular de imediato protocolos e parâmetros apropriados para a terapia por radiação para serem utilizados (vide, por exemplo, Perez and Brady, 1992, Principles and Practice of Radiation Oncology, 2nd Ed. JB Lippincott Co; usando adaptações e modificações apropriadas, que podem ser facilmente concebidas pelos versados na técnica).

Em ainda outra concretização, o método ou uso da invenção está associado à quimioterapia com um ou mais fármacos para o VHC convencionalmente utilizados para o

tratamento ou prevenção de infecções pelo VHC, doenças associadas ao VHC ou condições patológicas. Exemplos representativos de fármacos para o VHC incluem, sem limitação, inibidores da protease (por exemplo, inibidores da serina protease, como o VX950 da Vertex), inibidores da polimerase, inibidores da helicase, antifibróticos, análogos de nucleosídeos, agonistas de TLR, inibidores da N-glicosilação, siRNA, oligonucleotídeos anti-sentido, anticorpos anti-VHC, moduladores imunológicos, vacinas terapêuticas e agentes antitumorais geralmente utilizados no tratamento de hepatocarcinomas associados ao VHC (por exemplo, adriamicina ou uma mistura de adriamicina lipiodol e spongel, normalmente administrada por quimioembolização na artéria hepática). Por exemplo, vacinas terapêuticas incluem antígenos recombinantes, VLPs, vetores ou peptídeos sintéticos baseados em, ou codificando proteínas estruturais do VHC (Núcleo, envoltório E1 e/ou E2) que são particularmente adequadas para desencadear uma resposta humoral anti-VHC. Tais fármacos para o VHC podem ser fornecidos em uma dose única, ou, como alternativa, em múltiplas doses segundo os protocolos, dosagens e regimes convencionais, ao longo de várias horas, dias e/ou semanas. A administração desses fármacos pode ocorrer antes, durante ou depois da administração da composição da invenção. Uma combinação particularmente adequada inclui o tratamento com IFN- α 2a peguillado (por exemplo, a uma dose de 10 μ g/semana) e/ou ribavirina (por exemplo, de 800 a 1200

mg/dia) durante 24 a 48 semanas, antes, durante ou depois do método da invenção.

Em outra concretização, o método ou uso da invenção é realizado de acordo com uma modalidade terapêutica de indução e reforço, que compreende a administração sequencial de uma ou mais composições iniciadoras e uma ou mais composições reforçadoras. Geralmente, as composições iniciadoras e reforçadoras utilizam diferentes veículos, que compreendem ou codificam pelo menos um domínio antigênico em comum. A composição iniciadora é inicialmente administrada ao organismo hospedeiro e a composição reforçadora é subsequentemente administrada ao mesmo organismo hospedeiro após um período, que pode variar de um dia a doze meses. O método da invenção pode compreender de uma a dez administrações sequências da composição iniciadora, seguido de uma a dez administrações sequenciais da composição reforçadora. De preferência, os intervalos de injeção são uma questão de uma semana a seis meses. Além do mais, as composições iniciadoras e reforçadoras podem ser administradas no mesmo local, ou em locais alternativos, pela mesma via ou por vias de administração diferentes. Por exemplo, composições baseadas em polipeptídeo podem ser administradas por uma via mucosal, ao passo que composições baseadas em vetores são preferivelmente injetadas, por exemplo, injeção subcutânea para um vetor MVA, injeção intramuscular para um plasmídeo de DNA e para um vetor adenoviral.

No contexto da invenção, uma composição da invenção é utilizada tanto para induzir quanto para reforçar, ou então para induzir e reforçar uma resposta imunológica anti-VHC. As composições iniciadoras preferidas da invenção são, de preferência, aquelas que compreendem um vetor/vírus adenoviral de codificação da fusão e um DNA de plasmídeo de codificação de fusão. De preferência, utilizam-se vetores/vírus MVA de codificação da fusão para o reforço. Entretanto, é possível utilizar a proteína de fusão produzida por recombinação de forma independente para a iniciação ou reforço. Por exemplo, pode-se realizar a iniciação com uma proteína de fusão produzida por recombinação contendo a sequência de aminoácidos apresentada no SEQ ID NO: 9 ou no SEQ ID NO: 10 e o reforço com um vetor adenoviral codificando tal proteína de fusão. Como outro exemplo, pode-se realizar a iniciação com uma um plasmídeo de DNA codificando uma proteína de fusão contendo a sequência de aminoácidos apresentada no SEQ ID NO: 9 ou no SEQ ID NO: 10 e o reforço com partículas de vírus MVA codificando tal proteína de fusão. Também é possível utilizar a composição da invenção em combinação com qualquer material do estado da técnica codificando ou compreendendo um domínio antigênico em comum com a composição da invenção (por exemplo, um domínio antigênico do polipeptídeo NS3, NS4B e/ou NS5B). A origem desse material é vasta, e inclui, sem limitação, peptídeos, proteínas, vetor viral de uma variedade de vírus, DNA de plasmídeo, partícula proteináceas, como partículas tipo vírus

(Burns e col., 1994, Mol. Biol. Tech no. 1, 137-145), materiais celulares, como células irradiadas, partículas de vírus, etc. Por exemplo, vetores particularmente apropriados úteis neste contexto são os vetores adenovirais e MVA descritos na WO2004/111082 que codificando o NS3, NS4 e NS5B. Como outro exemplo, os peptídeos descritos na WO03/0976t77 compreendendo domínios antigênicos de NS3, NS4B e NS5B também são úteis no contexto da invenção. Para fins ilustrativos, pode-se realizar a iniciação com as partículas adenovirais da WO2004/111082 e depois o reforço com a proteína de fusão produzida de forma recombinante do SEQ ID NO: 9 ou SEQ ID NO: 10 ou com o vetor/vírus MVA codificando a proteína de fusão do SEQ ID NO: 9 ou SEQ ID NO: 10 conforme definida acima. Como outro exemplo, pode-se realizar a iniciação com um plasmídeo de DNA codificando a proteína de fusão do SEQ ID NO: 9 ou SEQ ID NO: 10, e reforçar com o vetor MVA da WO2004/111082.

No entanto, outras combinações de indutor e reforço também são plausíveis no contexto da invenção.

A presente invenção também oferece um método para induzir ou estimular uma resposta imunológica contra o VHC em um organismo hospedeiro, compreendendo administrar, ao referido organismo, a proteína de fusão, a molécula de ácido nucléico, o vetor, a partícula viral infecciosa, a célula hospedeira ou composição da invenção de modo a induzir ou estimular a referida resposta imunológica. A resposta imunológica pode ser uma resposta específica e/ou não específica,

humoral e/ou mediada por célula. A resposta imunológica é, de preferência, uma resposta de células T direcionada a um antígeno do VHC.

A capacidade do método da invenção em
5 induzir ou estimular uma resposta imunológica anti-VHC quando da administração em um organismo animal ou humano pode ser avaliada tanto *in vitro* quanto *in vivo* usando uma variedade de análises convencionadas na técnica. Para uma descrição geral das técnicas disponíveis para avaliar o princípio e a ativação de uma
10 resposta imunológica, consulte, por exemplo, Coligan e col. (1992 e 1994, Current Protocols in Immunology ; ed J Wiley & Sons Inc, National Institute of Health). A medição da imunidade celular pode ser realizada medindo-se os perfis de citocina secretados pelas células efetoras ativadas incluindo os derivados das células-
15 T CD4+ e CD8+ (por exemplo, quantificação das células produzindo IL-10 ou IFN γ por ELISPOT), pela determinação da condição de ativação das células efetoras imunológicas (por exemplo, ensaios de proliferação de células-T por uma absorção de timidina [3 H] clássica), analisando a presença de linfócitos-T
20 específicos do antígeno em um paciente sensibilizado (por exemplo, lise específica do peptídeo em uma análise de citotoxicidade). A capacidade de estimular uma resposta humoral pode ser determinada pela ligação de anticorpos e/ou competição na ligação (refira-se, por exemplo, a Harlow, 1989, Antibodies,
25 Cold Spring Harbor Press). O método da invenção também pode ser validado em modelos animais desafiados com um agente

infeccioso ou indutor de tumor apropriado (por exemplo, um vírus da varíola expressando genes NS) para determinar a neutralização do agente infeccioso ou indutor de tumor e, por fim, a resistência parcial aos sintomas associados, refletindo uma indução ou uma
5 melhora de uma resposta imunológica anti-VHC. O teste e validação das composições da invenção também são ilustrados na seção de Exemplos anexa.

A invenção também oferece anticorpos que se ligam seletivamente à proteína de fusão da presente invenção ou a
10 fragmentos peptídeos dela. Como usado neste documento, um anticorpo se liga seletivamente a um peptídeo-alvo quando ele se liga ao peptídeo-alvo e não se liga significativamente a proteínas não relacionadas. Em certos casos, entende-se que o anticorpo ligando-se ao peptídeo ainda é seletivo apesar de certo nível de
15 reatividade cruzada. Contudo, prefere-se que o anticorpo da invenção não se ligue com alta afinidade ou alta seletividade a polipeptídeos NS nativos individuais e fusões dos polipeptídeos NS configuradas em uma configuração nativa.

Como usado neste documento, um anticorpo
20 define-se em termos condizentes com o reconhecido na esfera técnica. Os anticorpos da presente invenção incluem anticorpos policlonais e anticorpos monoclonais, bem como fragmentos de tais anticorpos, incluindo, sem a isto se limitar, fragmentos Fav ou F(ab').sub.2, e Fv. Os anticorpos da presente invenção podem
25 ser produzidos usando técnicas convencionais, por exemplo, após a administração, a um animal, de uma quantidade eficaz de uma

proteína de fusão da presente invenção e/ou de um fragmento peptídico dela. Os anticorpos são preferivelmente preparados a partir de regiões ou fragmentos distintos da proteína de fusão da invenção compreendendo seqüências únicas, como as que
5 sobrepõe o sítio de fusão entre os polipeptídeos NS4A e NS3 e os polipeptídeos NS3 e NS5B.

Os anticorpos da presente invenção possuem uma variedade de usos possíveis, todos os quais se enquadram no âmbito da presente invenção. Por exemplo, tais anticorpos podem
10 ser usados (a) como reagentes em ensaios para detectar uma proteína de fusão da presente invenção, (b) como reagentes em ensaios para detectar a presença de um vírus VHC em uma amostra biológica e/ou (c) como ferramentas para recuperar a proteína de fusão produzida de forma recombinante da presente
15 invenção a partir de uma mistura de proteínas e outros contaminantes (por exemplo, permitindo a purificação pela cromatografia por afinidade ou imunoprecipitação a partir de células hospedeiras cultivadas).

A presente invenção também se refere a um
20 método para a detecção e/ou quantificação de um vírus VHC ou um anticorpo anti-VHC em uma amostra biológica (por exemplo, plasma, soro, tecido) coletada de um indivíduo suscetível a ser infectado pelo referido vírus VHC usando a proteína de fusão, molécula de ácido nucléico, vetor, partícula viral infecciosa,
25 célula hospedeira, composição ou anticorpo da invenção.

Em uma concretização, o método é mais particularmente adequado para a detecção e/ou quantificação de um vírus VHC em uma amostra biológica e compreende pelo menos as etapas de colocar a referida amostra biológica em
5 contato com pelo menos um dos anticorpos da invenção sob condições que permitem a formação de um complexo entre o vírus e o anticorpo, e detectar e/ou quantificar a formação do referido complexo por qualquer meio apropriado.

Em outra concretização, o método é mais
10 particularmente adequado à detecção e/ou quantificação de um anticorpo anti-VHC em uma amostra biológica e compreende pelo menos as etapas de colocar a referida amostra biológica em contato com pelo menos um dentre a proteína de fusão, a molécula de ácido nucléico, o vetor, a partícula viral infecciosa, a
15 célula hospedeira, a composição da invenção sob condições que permitem a formação de um complexo entre o anticorpo anti-VHC e a proteína de fusão, a molécula de ácido nucléico, o vetor, a partícula viral infecciosa, a célula hospedeira, a composição da invenção, e detectar e/ou quantificar a formação do referido
20 complexo por qualquer meio apropriado.

Os versados na técnica serão capazes de determinar facilmente a quantidade de anticorpos, proteínas de fusão, moléculas de ácido nucléico, vetores, partículas virais infecciosas, células hospedeiras e composições a serem utilizados
25 nos métodos da invenção. Os meios de detecção e/ou quantificação do vírus são comuns e conhecidos pelos versados na

técnica. A título de ilustração, podemos mencionar “blots”, ELISA, as chamadas técnicas “sanduíche”, técnicas de competição e técnicas PCR, em especial as chamadas técnicas “em tempo real”. O uso de um anticorpo, proteína de fusão, molécula de ácido nucléico, vetor, partícula viral infecciosa, célula hospedeira, ou composição da presente invenção como reagente pode ser facilitado pelo acoplamento (isto é, ligação física) a uma substância detectável. Exemplos de substâncias detectáveis incluem várias enzimas (por exemplo, peroxidase de raiz-forte, fosfata alcalina, beta-galactosidase ou acetilcolinesterase), grupos protéticos (por exemplo, estreptavidina/biotina, ou avidina/biotina), materiais fluorescentes (por exemplo, umbeliferona, fluoresceína ou derivados da fluoresceína), materiais luminescentes, materiais bioluminescentes (por exemplo, luciferase, luciferina ou aequorina) e materiais radioativos (por exemplo ^{125}I , ^{131}I , ^{35}S ou ^3H).

Finalmente, a invenção refere-se ao uso da proteína de fusão, molécula de ácido nucléico, vetor, partícula viral infecciosa, célula hospedeira, composição ou anticorpo da invenção para o diagnóstico *in vitro* de uma infecção pelo VHC em uma amostra biológica.

A invenção foi descrita de maneira ilustrativa, e deve ser compreendido que a terminologia utilizada é de caráter descritivo em vez de limitativo. Obviamente, muitas modificações e variações da presente invenção são possíveis à luz dos

ensinamentos acima. Portanto, deve-se entender que dentro do âmbito das reivindicações apenas, a invenção pode ser praticada de forma diferente da descrita especificamente neste documento.

Todas as revelações de patentes, publicações e
5 entradas de base de dados citadas acima são incorporadas neste por referência em sua totalidade da mesma forma que se cada patente, publicação ou entrada individual fosse indicada de forma específica e individual para ser incorporada por referência.

Legendas das Figuras

10 A Figura 1 ilustra a seqüência de aminoácidos de uma proteína de fusão NS4A-3-5B. Resíduos de aminoácidos adicionados ou modificados encontram-se sublinhados, ao passo que os resíduos nativos substituídos são apresentados abaixo da seqüência.

A Figura 2 ilustra a seqüência de aminoácidos
15 de uma proteína de fusão NS4A-3-4B-5B. Resíduos de aminoácidos adicionados ou modificados encontram-se sublinhados, ao passo que os resíduos nativos substituídos são apresentados abaixo da seqüência. A Figura 3 ilustra a seqüência de nucleotídeos codificando a proteína de fusão NS4A-3-5B
20 otimizada para expressão em células de mamíferos.

A Figura 4 ilustra a seqüência de nucleotídeos codificando a proteína de fusão NS4A-3-4B-5B otimizada para expressão em células de mamíferos.

A Figura 5 ilustra a seqüência de nucleotídeos
25 codificando a proteína de fusão NS4A-3-5B otimizada para expressão em células de levedura.

A Figura 6 ilustra a sequência de nucleotídeos codificando a proteína de fusão NS4A-3-4B-5B otimizada para expressão em células de levedura.

5 A Figura 7 ilustra a sequência de nucleotídeos codificando a proteína de fusão NS4A-3-5B otimizada para expressão em células procarióticas.

A Figura 8 ilustra a sequência de nucleotídeos codificando a proteína de fusão NS4A-3-4B-5B otimizada para expressão em células procarióticas, e a Figura 9 ilustra o desenho
10 esquemático das proteínas de fusão NS4A-3-5B e NS4A-3-4B-5B ilustradas nos Exemplos a seguir. A, B, C, D e E representam domínios antigênicos que podem ser prognosticados nos polipeptídeos NS3, NS4B e NS5B (vide a WO03/097677).

A Figura 10 ilustra a construção esquemática
15 dos vetores de expressão gWiz NS4A-3-5B Hs e gWiz NS4A-3-4B-5B Hs.

A Figura 11 ilustra a expressão *in vitro* por *Western blot* após a transfecção das células Huh-7 com os plasmídeos gWiz NS4A-3-5B (A pista 1 e C pista 3), gWiz
20 NS4A-3-4B-5B (B pista 1 e C pista 2) e gWiz (A pista 2, B pista 2 e C pista 1). As células Huh-7 transfectadas foram submetidas à lise 48h após a transfecção e as proteínas de fusão foram detectadas pela análise “Western blot” usando anticorpos anti-NS3 monoclonais de camundongo (A e B) ou anti-NS5B
25 monoclonais de camundongo (C).

A Figura 12 ilustra a avaliação das respostas das células CD8⁺T específicas do NS3 pelo ensaio IFN γ ELISPOT após a injeção intramuscular de gWiz NS4A-3-5B Hs. Os esplenócitos de camundongos individuais foram cultivados na presença do peptídeo específico do NS3 DLL ou de um peptídeo irrelevante por 40 h. Os resultados são ilustrados como o valor médio do número de manchas detectadas para 10⁶ esplenócitos obtidos para poços em triplicata. M1, M2, M3 e M4 representam 4 camundongos imunizados com o gWiz NS4A-3-5B Hs. Mc é um camundongo controle.

A Figura 13 ilustra a expressão das proteínas de fusão a partir de células infectadas pelo adenovírus.

Células A549 (1,5x10⁶ células) foram infectadas em um MOI de 10 com diferentes adenovírus, respectivamente, os adenovírus de expressão de fusão AdtG17476 e AdtG17477, bem como um adenovírus vazio e um adenovírus recombinante irrelevante como controles negativos. As células foram cultivadas por 48h com 10% de soro antes de serem submetidas à lise com tampão Laemmli. Amostras de proteínas (1/20 do volume) foram carregadas em um gel de eletroforese SDS PAGE. As proteínas foram detectadas pelo azul de Coomassie. Pista 1: marcadores de peso molecular; Pista 2: células A549 não infectadas; Pista 3: células A 549 infectadas com adenovírus vazio; Pista 4: células A 549 infectadas com um adenovírus irrelevante; Pista 5: células A 549 infectadas com AdTG17476; e Pista 6: células A 549 infectadas com AdTG17477.

A Figura 14 ilustra respostas IFN γ Elispot induzidas com proteínas de fusão do VHC Ad em camundongos transgênicos HLA-A2.1. O AdTG17476, AdTG17477 ou Ad Vazio foram injetados pela via intramuscular no músculo tibial anterior 1 vez a uma dose de 10^9 IU/camundongo. As respostas imunológicas celulares foram investigadas 2 semanas após a injeção por IFN γ ELISPOT para cada grupo animal (incluindo 4 camundongos para os grupos de teste e 2 camundongos para o grupo controle). As respostas dos camundongos individuais (M1 a 4) contra os epítomos restritos em HLA-A2 específicos ao NS3 GLL (Figura 14A) e KLT (Figura 14B) e o epítomo restrito em HLA-A2 específico ao NS5B KLQ (Figura 14C) são apresentadas. O peptídeo irrelevante DLM é utilizado como controle negativo. Uma resposta ELISPOT é considerada positiva se o número de manchas para 10^6 esplenócitos for maior do que 50 (linha de corte).

Os exemplos a seguir servem para ilustrar a presente invenção.

EXEMPLOS

Materiais e Métodos

As construções descritas abaixo são preparadas de acordo com as técnicas gerais de engenharia genética e de clonagem molecular, como detalhado em Sambrook e col. (2001, Molecular Cloning ; A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY) ou segundo as recomendações do fabricante quando um kit comercial é utilizado.

As técnicas de amplificação PCT são conhecidas pelos versados na técnica (vide, por exemplo, PCR Protocols - A guide to Methods and applications, 1990; Ed Innis, Gelfand, Sninsky and White, Academic Press Inc). A linhagem celular do hepatoma Huh-7 humano foi utilizada para avaliar a expressão correta das proteínas de fusão a partir das construções descritas nos exemplos que seguem. As condições de cultura são convencionais na técnica. Para fins ilustrativos, as células são cultivadas a 37°C em atmosfera de CO₂ 5% em meio DMEM completo contendo meio de Eagles modificado por Dulbecco suplementado com Soro Fetal Bovino a 10%, 2mM de L-glutamina e 100IU/ml de penicilina/estreptomicinas (Sigma). As células são transfectadas usando o reagente Lipofectamine Plus (Invitrogen) e segundo as instruções do fabricante. Anticorpos monoclonais murinos direcionados ao NS3 e ao NS5B podem ser adquiridos comercialmente ou preparados segundo as técnicas convencionadas.

Construções dos plasmídeos

Oligonucleotídeos sobrepostos foram utilizados para gerar genes sintéticos (GeneART, Regensburg, Alemanha) codificando, respectivamente, as proteínas de fusão NS4A-NS3-NS5B (3666 nucleotídeos) e NS4A-NS3-NS4B-NS5B (3747 nucleotídeos). O desenho das fusões é ilustrado na Figura 9. As seqüências de nucleotídeos foram otimizadas para expressão elevada em homo sapiens (SEQ ID NO: 11 e 12 e Figuras 3 e 4), Pichia (SEQ ID NO: 13 e 14 e Figuras 5 e 6) e E. coli (SEQ ID

NO: 15 e 16 e Figuras 7 e 8). A sequência de aminoácidos das fusões resultantes é ilustrada no SEQ ID NO: 9 para NS4A-NS3-NS5B e no SEQ ID NO: 10 para NS4A-NS3-NS4B-NS5B. Cada uma das seqüências de nucleotídeos sintéticos foi clonada nos
 5 sítios de restrição *SacI* e *KpnI* do plasmídeo pPCR-Script (Stratagene).

Para expressão temporária em células de mamíferos, genes sintéticos de codificação de fusão otimizados para expressão em homo sapiens (SEQ ID NO: 11-12 e Figuras 3
 10 e 4) foram inseridos nos sítios de restrição *SacI* e *NotI* do vetor de expressão gWiz (Gene Therapy Systems) sob o controle transcricional do promotor do citomegalovírus, para gerar os vetores gWiz NS4A-3-5B Hs e gWiz NS4A-3-4B-5B Hs (Figura 10). Em seguida, a seqüência foi modificada para aperfeiçoar a
 15 seqüência consenso Kozak (sublinhada) usando o oligonucleotídeo apresentado no SEQ ID NO: 17 (5'-ATG TAC GTC GAC CACC ATG GGC AGC GTG GTG ATT GTG GGC CGG ATC 3') e para criar códons de parada (em negrito) usando o oligonucleotídeo apresentado no SEQ ID NO: 18 (5'-CAT GTA
 20 GC GGC CGC **TCA TCA** TCT AGG CCT GGC CCT GGA CAG-3'). Os plasmídeos AU foram verificados por sequenciamento.

Estudos de expressão *in vitro*

Análise Western blot

25 Um dia antes da transfecção, Huh-7 (5 x 10⁵ por poço) foram distribuídas em uma placa de 6 poços. As células

foram transfectadas com o vetor gWiz NS4A-3-5B Hs, gWiz NS4A-3-4B-5B Hs ou com o plasmídeo gWiz como controle negativo. Passadas 48 h da transfecção, as células foram submetidas à lise em 50mM de Tris-HCl, 150mM de NaCl, 0,1% de SDS, 1% de NP40 e 0,5% de Na-deoxicholato. Amostras de lisados foram aquecidas a 95°C por 5 minutos e as proteínas foram separadas por eletroforese em SDS-8% poliacrilamida. O “immunoblotting” foi realizado usando anticorpos monoclonais murinos específicos do NS3 (por exemplo, 8G4C3, bioMerieux) ou NS5B (por exemplo, 11F6C8, bioMerieux) como anticorpo primário seguido de um anticorpo anti-IgG-peroxidase de camundongo produzido em cabra (Sigma) como anticorpo secundário. Os sinais foram revelados usando o kit ECL comercial (Amersham Biosciences).

Análise de imunofluorescência

Um dia antes da transfecção, lamínulas foram colocadas numa placa de 6 poços e tratadas com Gelatina a 0,2% por 10 min e depois 5×10^5 células Huh-7 foram distribuídas em cada poço da placa. As monocamadas celulares foram transfectadas com o vetor gWiz NS4A-3-5B Hs, gWiz NS4A-3-4B-5B Hs ou com o plasmídeo gWiz como controle negativo. Passadas 48 h da transfecção, as lamínulas foram lavadas duas vezes em PBS e as células foram fixadas por 10 min com 4% PFA e então lavadas e permeabilizadas com Triton X-100 a 0,1% em PBS por 10min. As lamínulas foram lavadas e tratadas com anticorpos monoclonais murinos específicos do NS3 (por

exemplo, 8D8E1, bioMerieux) ou NS5B (por exemplo, 5B-12B7, fornecido pelo D. Moradpour) como anticorpo primário por 1 h à temperatura ambiente. Após a lavagem, adicionou-se anti-IgG de camundongo TRITC (Dako) por 30 min. Após a lavagem, as
 5 lamínulas foram montadas em glicerol a 80%. As pistas foram observadas com um microscópio Carl Zeiss Axioplan, sendo obtidas imagens com a câmera digital AxioCam Color.

Estudos de expressão *in vivo*

Camundongos transgênicos HLA-A2.1
 10 produzidos como descrito por Pascolo e col. (1997, J. of Experimental Medicine 185, 2043-2051) foram procriados. Esses camundongos têm os genes H-2D^b e murine β_2 -microglobulina murina neutralizados e expressam uma molécula de classe I de histocompatibilidade de monocadeia transgênica na qual o C-
 15 terminal do β_2 m humano é ligado de forma covalente ao N-terminal de uma cadeia pesada quimérica (domínios transmembrana e intracitoplásmicos HLA-A2.1 $\alpha 1$ - $\alpha 2$, H-2D^b $\alpha 3$)

Protocolos de Imunização

Camundongos transgênicos HLA-A2 receberam
 20 um pré-tratamento com 10 μ M de cardiotoxina (Latoxan) cinco dias antes da injeção dos plasmídeos. Os plasmídeos gWiz NS4A-3-5B, gWIZNS4A-3-4B-5B e gWiz (controle negativo) foram injetados por via intramuscular 2 vezes em intervalos de duas semanas a uma dose de 100 μ g (em 100 μ l de IX PBS). As
 25 respostas das células CD8⁺-T restritas em HLA-A2 foram

investigadas 2 semanas após a 2^a imunização, tal como descrito em Himoudi e col. (2002, J. Virol. 76, 12735-46).

Monitoramento da resposta imunológica por ensaios ELISPOT

5 Os esplenócitos (2×10^5) foram cultivados em poços em triplicata por 40 h em placas Multiscreen (Millipore) revestidas com anticorpo monoclonal anti- IFN γ de camundongo (Pharmingen; 10 μ g/ml) em meio de cultura α MEM completo (Invitrogen) na presença de 10 unidades/ml de IL-2 recombinante
10 (ReproTech Inc, England) sozinha como controle negativo, ou 5 μ g/ml de Conavalina A como controle positivo ou com 10 μ M de peptídeo (peptídeo irrelevante DLM, peptídeo GLL localizado na proteína NS3 descrito em Martin e col., 2004, J. Med. Virol. 74, 397-405 e Himoudi e col., 2002. J. Virol. 76. 12735-46). As
15 células produtoras de IFN γ foram quantificadas pelo ensaio ELISPOT (*enzyme linked immunospot assay*) específico à citocina, conforme descrito anteriormente (Himoudi e col., 2002, J. Virol. 76, 12735-46). O número de manchas (representando as células produtoras de IFN γ individuais) nos poços de controle
20 negativos foi subtraído do número nos poços de teste contendo peptídeos. Os resultados são apresentados como o valor médio obtido para poços em triplicata.

EXEMPLO 1: Expressão *in vitro* das poliproteínas NS4A-3-5B e NS4A-3-4B-5B

25 As seqüências codificando as proteínas de fusão NS4A-3-5B e NS4A-3-4B-5B foram construídas como delineado

na Figura 9. As seqüências foram otimizadas para expressão em três hospedeiros diferentes (GeneART, Regensburg, Alemanha), *Pichia pastoris* (Pp) humana (Hs) e *E. coli* (Ec), respectivamente, como listado a seguir:

5 NS4A-3-5B Hs otimizado para expressão em
homo sapiens (SEQ ID NO: 11 e Figura 3)

 NS4A-3-5B Pp otimizado para expressão em
Pichia (SEQ ID NO: 13 e Figura 5)

 NS4A-3-5B Ec otimizado para expressão em *E.*
10 *coli* (SEQ ID NO: 15 e Figura 7)

 NS4A-3-4B-5B Hs otimizado para expressão
em homo sapiens (SEQ ID NO: 12 e Figura 4)

 NS4A-3-4B-5B Pp otimizado para expressão
em *Pichia* (SEQ ID NO: 14 e Figura 6)

15 NS4A-3-4B-5B Ec otimizado para expressão
em *E. coli* (SEQ ID NO: 16 e Figura 8)

 As seqüências otimizadas para expressão em
homo sapiens codificando as fusões NS4A-NS3-NS4B-NS5B e
NS4A-NS3-NS5B foram clonadas no vetor de expressão gWiz
20 (GeneTherapy Systems) sob o controle transcricional do promotor
do citomegalovírus (Figura 10). A expressão das proteínas de
fusão NS foi avaliada por Western blot e a imunofluorescência
em células Huh-7 transfectadas com os vetores gWiz NS4A-3-5B
Hs, gWiz NS4A-3-4B-5B Hs ou o plasmídeo gWiz como controle
25 negativo.

Como ilustrado na Figura 11, a análise Western Blot revelou uma banda única contendo o peso molecular esperado de aproximadamente 135 kDa para cada fusão individual (Figura 11A pista 1, e Figura 11C pista 3 para a fusão NS4A-3-5B curta e Figura 11B pista 1, e Figura 11C pista 2 para a fusão NS4A-3-4B-5B longa) após a detecção com anticorpos específicos anti-NS3 (Figuras 11A e 7B) ou anti-NS5B (Figura 11C). Como esperado, nenhuma proteína foi detectada nas amostras obtidas das células Huh-7 transfectadas com o plasmídeo gWiz (Figura 11A pista 2, Figura 11B pista 2 Figura 11C pista 1). Esses resultados confirmam que a fusão curta NS4A-NS3-NS5B e a fusão longa adicionalmente incorporando NS4B são expressas nas células Huh-7.

A análise de imunofluorescência confirmou a expressão das duas proteínas de fusão nas células transfectadas Huh-7. Curiosamente, ambas as proteínas de fusão apresentaram uma localização citoplásmica indicando o processamento apropriado como resultado da deleção da maioria dos domínios de ancoragem à membrana hidrófobos. As células transfectadas com o plasmídeo gWiz não apresentaram nenhum sinal.

EXEMPLO 2: Avaliação *in vivo* da imunogenicidade do gWIZ NS4A-3-5B Hs e do NS4A-3-4B-5B Hs

Para avaliar a capacidade dos plasmídeos gWiz NS4A-3-5B Hs e NS4A-3-4B-5B Hs em induzir respostas das células CD8⁺-T em camundongos transgênicos HLA-A2, um

estudo de imunogenicidade foi realizado seguindo os camundongos transgênicos HLA-A2 como descritos em Material e Métodos. As respostas das células CD8⁺-T restritas por HLA-A2 foram investigadas pelo ensaio ELISPOT IFN γ 2 semanas após a 2^a imunização. Como ilustrado na Figura 12, o pgWiz NS4A-3-5B foi capaz de induzir células-T produtoras de IFN γ específicas do peptídeo GLL bem conhecido e imunodominante restrito em HLA-A2 (proteína NS3, Martin e col., 2004, J. Med. Virol. 74, 397- 405), evidenciando a apresentação correta deste epítipo. Também é possível utilizar ensaios de desafio murino pré-clínicos a fim de avaliar as proteínas de fusão quanto à imunidade protetora. Uma vez que os camundongos não podem ser infectados diretamente pelos vírus VHC, esses ensaios empregam ou vírus recombinantes da varíola (linhagem WR) ou Listéria expressando os antígenos NS do VHC. Após a injeção do vírus da varíola ou Listéria, os camundongos desenvolvem uma infecção que alcança seu ápice 2 a 6 dias após a injeção, resultando em vírus ou bactérias detectados nos ovários (modelo de varíola) ou fígados (modelo de Listéria) dos camundongos injetados. A avaliação pré-clínica envolve a vacinação dos camundongos com a vacina candidata (administração da proteína de fusão recombinante ou de partículas de adenovírus expressando a proteína de fusão), controles (por exemplo, administração da fusão de polipeptídeos NS na configuração nativa) e construções negativas (por exemplo: antígenos não-VHC ou adenovírus vazio). Os animais são subsequentemente

desafiados com um vírus recombinante da varíola expressando todos os antígenos NS do VHC (NS2 a NS5) ou Listéria expressando um dos antígenos NS incluídos no candidato à vacina. Se a vacina iniciar com êxito células-T específicas, os

5 camundongos desafiados desenvolvem resposta imunológica ao vetor da varíola ou Listéria e ao(s) antígeno(s) NS. A atividade protetora da vacina candidate irá resultar em títulos de varíola ou Listéria reduzidos, conforme medido nos ovários ou fígados dos camundongos vacinados, se comparado aos títulos de varíola ou

10 Listéria medidos nos ovários ou fígados dos camundongos não vacinados.

EXEMPLO 3: Expressão da proteína de fusão NS4A-3-5B e NS4A-3-4B-5B em células de mamíferos.

Os genes sintéticos de codificação da fusão

15 otimizados para expressão em homo sapiens (SEQ ID NO: 11-12), correspondendo a NS4A-NS3-NS5B (3666 nucleotídeos) e NS4A-NS3-NS4B-NS5B (3747 nucleotídeos) foram amplificados pela PCR a partir dos plasmídeos pPCR-Script correspondentes, utilizando-se os seguintes oligonucleotídeos iniciadores

20 OTG18033

(GGGGGGCTAGCGCCACCATGGGCAGCGTGGTGATTG;
SEQ ID NO: 19) e OTG18036
(GGGGGGGTACCCTCATCATCTAGGCCTGGCCCTG; SEQ
ID NO: 20). Ambos os fragmentos foram inseridos nos sítios de

25 restrição *NheI* e *KpnI* de um plasmídeo “lançadeira” adenoviral contendo um cassete de expressão induzido pelo CMV

circundado por seqüências adenovirais (nucleotídeos adenovirais 1-458 e nucleotídeos 3328-5788, respectivamente) para permitir a geração do genoma vetorial por recombinação homóloga (Chartier e col., 1996, J. Virol. 70, 4805- 4810). Os vetores
5 adenovirais resultantes são o E3 (nucleotídeos 28592-30470) e E1 (nucleotídeos 459-3511) deletados, com a região E1 substituída pelo cassete de expressão contendo, de 5' a 3', a seqüência reforçadora/promotor de resposta imediata CMV, um íntron quimérico da β -globina/IgG humana (como encontrado no vetor
10 pCI disponível na Promega), a seqüência codificando qualquer uma das duas formas de poliproteína e o sinal de poliadenilação tardio SV40. Os vetores adenovirais resultantes foram chamados de pTG17476 (NS4A-3-4B-5B) e pTG17477 (NS4A-3-5B). Os adenovírus recombinantes, AdTG17476 e AdTG17477, foram
15 gerados transfectando os genomas virais linearizados *PacI* em uma linhagem celular de complementação E1 convencional. A propagação do vírus, a purificação e a titulação foram realizadas conforme descrito anteriormente (Erbs, 2000, Cancer Res. 60, 3813-3822).

20 A expressão das proteínas de fusão foi avaliada nas células infectadas pelo adenovírus pela eletroforese SDS PAGE. As células A549 ($1,5 \times 10^6$ células) (ATCC CCL-185) foram infectadas a um MOI de 10 ou 100 com os adenovírus de expressão da fusão AdTG17476 e AdTG1 7477, assim como com
25 um adenovírus vazio e um adenovírus recombinante irrelevante como controles negativos. As células foram cultivadas por 48h

com soro a 10% antes de serem submetidas à lise com tampão Laemmli. Amostras de proteínas (1/20 do volume) foram carregadas em um gel de eletroforese SDS PAGE e as proteínas foram coradas por Azul de Coomassie. Como ilustrado na Figura 13, uma banda única contendo o peso molecular esperado de aproximadamente 135 kDa foi claramente observada nas células infectadas com AdTG17476 e AdTG17477 (Pistas 5 e 6), refletindo os altos níveis de expressão nas células infectadas com os vetores adenovirais de codificação da fusão. No entanto, essa banda está ausente nas amostras coletadas das células infectadas com os controles negativos (Pistas 3 e 4), bem como nas células A549 não infectadas (Pista T). Os níveis de expressão foram analisados por Western Blot nas células A549 infectadas com as construções virais ao MOI de 10 por 48 horas. Os precipitados de células foram coletados e sondados com um anticorpo monoclonal murino anti-NS3 (por exemplo, 1B6, bioMerieux) e um anticorpo policlonal de Coelho anti-NS4 (por exemplo, RB, bioMerieux). Uma banda principal com o peso molecular esperado foi revelada nas amostras coletadas de células infectadas com o AdTG17476 e o AdTG17477 após a detecção com os anticorpos específicos anti-NS3 ou anti-NS4, confirmando os altos níveis de expressão detectados após corar com Azul de Coomassie.

EXEMPLO 4: Imunogenicidade dos Adenovírus expressando as poliproteínas do VHC em camundongos transgênicos HLA A2. Modelo animal

Os camundongos transgênicos HLA-A2.1 foram produzidos como descrito por Pascolo e col. (1997, J Ex Med). Esses camundongos têm os genes H-2D^b e murine β_2 -microglobulina murina neutralizados e expressam uma molécula de classe I de histocompatibilidade de monocadeia transgênica na qual o C-terminal do β_2m humano é ligado de forma covalente ao N-terminal de uma cadeia pesada quimérica (domínios transmembrana e intracitoplásmicos HLA-A2.1 $\alpha 1$ - $\alpha 2$, H-2D^b $\alpha 3$)

Protocolo de Imunização

Três grupos de camundongos foram incluídos no protocolo, respectivamente, quatro camundongos injetados com AdTG17476 (Ad NS4A-3-4B-5B), quatro camundongos injetados com AdTG17477 (Ad NS4A-3-5B) e dois camundongos controle injetados com um Ad vazio (Ad1514). Os adenovírus foram injetados por via intramuscular no músculo anterior tibial 1 vez a uma dose de 10^9 em 100 μ l de IX PBS. As respostas imunológicas celulares foram investigadas 2 semanas após a injeção.

Monitoramento da resposta imunológica

Ensaio ELISPOT

Os esplenócitos de camundongos de cada grupo foram coletados e os glóbulos vermelhos do sangue foram submetidos à lise. $2,10^5$ células foram cultivadas em poços em triplicata por 40 h em placas Multiscreen (Millipore) revestidas com anticorpo monoclonal anti-IFN γ de camundongo (Pharmlingen; 10 μ g/ml) em meio de cultura α MEM complete

(GIBCO BRL) diante da presença de 10 unidades/ml de UL-2 murina recombinante (PeproTech Inc, Inglaterra), sozinho como controle negativo, ou com 10 μ M de peptídeo (peptídeo irrelevante DLM, peptídeos GLL e KLT localizados na proteína NS3, peptídeo KLQ localizado na proteína NS5B (Martin e col., 2004, J Med Virol. 74, 397-405) e (Himoudi col., 2002, J Virol., 76, 12735-12746), ou 5 μ g/ml de Concanavalina A como controle positivo. As células-T produtoras de IFN γ foram quantificadas pelo ensaio ELISPOT (*enzyme linked immunospot assay*) conforme descrito anteriormente (Himoudi e col., J Virol 2002). O número de manchas (representando as células-T produtoras de IFN γ individuais) nos poços de controle negativos foi subtraído do número nos poços de teste contendo peptídeos. Os resultados são apresentados como o valor médio obtido para poços em triplicata.

Resultados

As respostas das células-T restritas em HLA-A2 foram investigadas pelo ensaio ELISPOT IFN γ 2 semanas após a imunização contra o adenovírus. Como ilustrado na Figura 14, tanto o AdTG17476 como o AdTG17477 foram capazes de induzir, em todos os animais vacinados, altas freqüências de células-T produtoras de IFN γ específicas do epítipo GLL restrito em HLA-A2 (Figura 14A; valor mediano: 1210 e 868 manchas, respectivamente) com uma eficiência significativamente superior do ponto de vista estatístico para o AdTG17476 (teste de Mann-Whitney: P=0,0209). Ambos os vírus também induziram células-

T produtoras de IFN γ específicas do epítipo KLT subdominante (Figura 14B) com frequências inferiores (valor mediano: 187 e 90 manchas). Além disso, 2:4 camundongos vacinados com AdTG17476 e 1:4 camundongos vacinados com AdTG17477 desenvolveram respostas direcionadas ao epítipo KLQ da proteína NS5B (Figura 14C).

Isso demonstra que as proteínas de fusão de acordo com a invenção podem ser produzidas em níveis elevados nos sistemas eucarióticos e que as proteínas resultantes são imunogênicas nos animais do modelo convencional.

. 10 20 30 40 50 60
 | | | | |
. |-> NS3
MGSVVIVGRIILSGSGSAPITAYSQQTRGLLGCIITSLSLTGRDKNQVDGEVQVLSTATQS
peptídeo de ligação 'domínio antigênico B, NS3 protease)
domínio de interação do NS4B

 70 80 90 100 110 120
LATCVNGVCWTVYAAGAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSMTPCTCGSSD
H (domínio antigênico E, NS3 protease)

 130 140 150 160 170 180
LYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHVVGIFRAAVCTRGV

 190 200 210 220 230 240
AKAVDFIPVESMETTMRSPVFETDNSSPPAVPQTTFQVAHLHAPTGS GKSTKVPAAYAAQGY

 250 260 270 280 290 300
KVLVLNPSVAATLGF GAYMSKAHGIEPNIRTGVRTITTGGPITYSAYGKFLADGGCSGGA
(domínio antigênico A, NS3 helicase) T

 310 320 330 340 350 360
YDIICDECHSTDWTTILGIGTVLDQAETAGARLVVLATATPPGSITVPHPNIEEVALSN

 370 380 390 400 410 420
TGEIPFYGKAIP IEAIKGRRLIFCHSKKKCELA AKLTGLGLNAVAYYRG LDVSVIPTS.

 430 440 450 460 470 480
GDVVVVATDALMTGFTGDFDSVIDCNTCVTQT VDFS LDPFTTIETTVPQDAVSR SQRRG

 490 500 510 520 530 540
ATGRGRSGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYLNTPGLP
R

 550 560 570 580 590 600
VCQDHLEFWE SVFTGLTHIDAHFLSQTKQAGDNFP YLVAYQATVCARAQAPPPSWDQMWK

 610 620 630 640 650 660
CLIRLKPTLHGPTPLLYRLGAVQNEITLTHPI TKFVMACMSADLEVVGSGSGSMSXTWTG
NS3 <- | | -> NS5B
peptídeo de ligação

 670 680 690 700 710 720
ALITPCAAEESKL PINPLSNSLLRHHS MVYSTTSRSASLRQKKVT FDR LQVLDDHYRDVL

 730 740 750 760 770 780
KEMKAKASTVKARLLSIEEACKLTPPHSAKSKFGYGA KDVRSLSSRAVNHIRSVWEDLLE

```

      790      800      810      820      830      840
      |      |      |      |      |      |
DTETPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMG
      ( domínio antigênico D, NS5B)

      850      860      870      880      890      900
      |      |      |      |      |      |
PSYGFQYSPGQRVFLVNTWKSCKCPMGFSYNTRCFDSTVTENDIRTEESIQCCDLAPE
      |
      D

      910      920      930      940      950      960
      |      |      |      |      |      |
ARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCTYLKATAACRAAKL

      970      980      990      1000      1010      1020
      |      |      |      |      |      |
QDCTMLVNGNDLVVICESAGTQEDAASLRVFTEAMTRYSAAPPDPPQPEYDLELITSCSS
      |
      D

      1030      1040      1050      1060      1070      1080
      |      |      |      |      |      |
NVSVAHDASGKRVYYLTRDPTTPLARAAWETVRHTPVNSWLGNIIIMYAPTLWARMILMTH

      1090      1100      1110      1120      1130      1140
      |      |      |      |      |      |
FFSILLAEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVASC

      1150      1160      1170      1180      1190      1200
      |      |      |      |      |      |
LRKLGVPPLRVWRHRARSVRAKLLSQGGAATCGKYLFNWAVRTKLKLTPIFAASQLDLS

      1210      1220
      |      |
GWFFVAGYNGGDIYHSLSRARPR

```

. 10 20 30 40 50 60
 . | | | | | |
 . |-> NS3
 MGSVVIVGRIILSGSGSAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQSF
 peptideo de ligação (domínio antigênico. B, NS3 protease)
 domínio de interação do NS4B
 . 70 80 90 100 110 120
 . | | | | | |
 LATCVNGVCWTVYAGAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSMTPCTCGSSD
 . H (domínio antigênico: E, NS3 Protease)
 . 130 140 150 160 170 180
 . | | | | | |
 LYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKSSGGPLLCPSGHVVGIFRAAVCTRGV
 . 190 200 210 220 230 240
 . | | | | | |
 AKAVDFIPVESMETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGBKSTKVPAAAYAAQGY
 . 250 260 270 280 290 300
 . | | | | | |
 KVLVLNPSVAATLGFGAYMSKAHGIEPNIRTGVRTITTGGPITYSA^TYGKFLADGGCSGGA
 (.domínio antigênico A, NS3 helicase)
 . 310 320 330 340 350 360
 . | | | | | |
 YDIIICDECHSTDWTTILGIGTVLDQAETAGARLVVLATATPPGSITVPHPNIEEVALSN
 . 370 380 390 400 410 420
 . | | | | | |
 TGEIPFYGKAIP^EIAIKGGRHLIFCHSKKKCDELA^AKLTLGLGLNAVAYYRGLDVSVIPTS
 . 430 440 450 460 470 480
 . | | | | | |
 GDVVVVATDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSR^SQRRG
 . 490 500 510 520 530 540
 . | | | | | |
 ATGRGRSGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYLNTPLP
 R
 . 550 560 570 580 590 600
 . | | | | | |
 VCQDHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVA^YQATVCARAQAPPSWDQMWK
 . 610 620 630 640 650 660
 . | | | | | |
 CLIRLKPTLHGPTPLLYRLGAVQNEITLTHPITKFVMACMSADLEV^VSLMAFTASITSPL
 . NS3 <-!-> NS4B epitopo
 (domínio
 . 670 680 690 700 710 720
 . | | | | | |
 . NS4B epitope <-!-> NS5B
 TTQNTLLFNILGGWVAAQLSMSYTW^TGTALITPCAAEESKLPINPLSNSLLRHHSMVYSTT
 antígeno C, NS4B)

. 730 740 750 760 770 780
 . | | | | |
 SRSASLRQKKVTFDRLQVLDHRYRDVLKEMKAKASTVKARLLSIEEACKLTPPHSAKSKF

 . 790 800 810 820 830 840
 . | | | | |
 GYGAKDVRSLSSRAVNHIRSVWEDLLEDTEPIDTTIMAKNEVFCVQPEKGGRKPARLIV

 . 850 860 870 880 890 900
 . | | | | |
 FPDLGVRVCEKMALYDVVSTLPQAVMGPSYGFQYSPGQQRVEFLVNTWKS~~KK~~CPMGFSYNT
 (.domínio antigênico. D, NS5B) D

 . 910 920 930 940 950 960
 . | | | | |
 RCFDSTVTENDIRTEESIQCCDLAPPEARQAIKSLTERLYIGGPLTNSKGQNCGYRR CRA

 . 970 980 990 1000 1010 1020
 . | | | | |
 SGVLTTS~~CG~~NTLTTCYLKATAACRAAKLQDCTMLVNGNDLVVICESAGTQEDAASLRVFTE
 D

 . 1030 1040 1050 1060 1070 1080
 . | | | | |
 AMTRYSA~~PPG~~DPPQPEYDLELITSCSSNVSV~~AHD~~ASGKRVYYLTRDPTT~~PLARA~~AWETVR

 . 1090 1100 1110 1120 1130 1140
 . | | | | |
 HTPVNSWLGNII~~MY~~APTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDL~~PQ~~

 . 1150 1160 1170 1180 1190 1200
 . | | | | |
 IIERLHGLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVR~~AKLLS~~QGGRATC

 . 1210 1220 1230 1240
 . | | | |
 GKYLFWAVRTKLKLTPIPAASQLDL~~SGW~~FVAGYNGGDIYHSLSRARPR

```

1  CGAATTGGGTACCGCCACCATGGGCAGCGTGGTGATTGTGGGCCGGATCATCCTGAGCGG
   -----+-----+-----+-----+-----+
   GCTTAACCCATGGCGGTGGTACCCGTCGCACCACTAACACCCGGCCTAGTAGGACTCGCC
       M _ G _ S _ V _ V _ I _ V _ G _ R _ I _ I _ L _ S _ G _

61  CAGCGGCAGCGCCCCCATCACCGCCTACAGCCAGCAGACCAGAGGCCTGCTGGGCTGTAT
   -----+-----+-----+-----+-----+
   GTCGCCGTCGCGGGGGTAGTGGCGGATGTCGGTCGTCTGGTCTCCGGACGACCCGACATA
   S _ G _ S _ A _ P _ I _ T _ A _ Y _ S _ Q _ Q _ T _ R _ G _ L _ L _ G _ C _ I _

121 CATCACCAGCCTGACCGGCAGAGACAAGAATCAGGTGGACGGCGAGGTGCAGGTGCTGTG
   -----+-----+-----+-----+-----+
   GTAGTGGTCGGA CTGGCCGTCTCTGTTCTTAGTCCACCTGCCGCTCCACGTCCACGACAG
   I _ T _ S _ L _ T _ G _ R _ D _ K _ N _ Q _ V _ D _ G _ E _ V _ Q _ V _ L _ S _

181 CACCGCCACCCAGAGCTTCCTGGCCACCTGTGTGAATGGCGTGTGTTGGACCGTGTACGC
   -----+-----+-----+-----+-----+
   GTGGCGGTGGGTCTCGAAGGACCGGTGGACACACTTACCGCACACAACCTGGCACATGCG
   T _ A _ T _ Q _ S _ F _ L _ A _ T _ C _ V _ N _ G _ V _ C _ W _ T _ V _ Y _ A _

241 CGGAGCCGGCAGCAAGACCCTGGCCGGACCCAAGGGCCCCATCACCCAGATGTACACCAA
   -----+-----+-----+-----+-----+
   GCCTCGGCCGTCGTTCTGGGACCGGCCTGGGTTCCCGGGGTAGTGGGTCTACATGTGGTT
   G _ A _ G _ S _ K _ T _ L _ A _ G _ P _ K _ G _ P _ I _ T _ Q _ M _ Y _ T _ N _

301 CGTGGACCAGGACCTGGTGGGCTGGCCTGCCCCTCCTGGCGCCAGAAGCATGACCCCTTG
   -----+-----+-----+-----+-----+
   GCACCTGGTCCTGGACCACCCGACCGGACGGGGAGGACCGCGGTCTTCGTACTGGGGAAC
   V _ D _ Q _ D _ L _ V _ G _ W _ P _ A _ P _ P _ G _ A _ R _ S _ M _ T _ P _ C _

361 TACCTGTGGCAGCAGCGACCTGTACCTGGTGACCAGACACGCCGATGTGATCCCTGTGAG
   -----+-----+-----+-----+-----+
   ATGGACACCGTCGTCGCTGGACATGGACCACTGGTCTGTGCGGCTACACTAGGGACACTC
   T _ C _ G _ S _ S _ D _ L _ Y _ L _ V _ T _ R _ H _ A _ D _ V _ I _ P _ V _ R _

421 GAGGAGAGGCGATAGCAGAGGCAGCCTGCTGTCTCCTAGACCCGTGTCCTACCTGAAGGG
   -----+-----+-----+-----+-----+
   CTCCTCTCCGCTATCGTCTCCGTCGGACGACAGAGGATCTGGGCACAGGATGGACTTCCC
   R _ R _ G _ D _ S _ R _ G _ S _ L _ L _ S _ P _ R _ P _ V _ S _ Y _ L _ K _ G _

481 CAGCAGCGGCGGACCCCTGCTGTGCCCCAGCGGCCACGTGGTGGGCATCTTCAGAGCCGC
   -----+-----+-----+-----+-----+
   GTCGTGCGCCGCTGGGGACGACACGGGGTCGCCGCTGCACCACCCGTAGAAGTCTCGGCG
   S _ S _ G _ G _ P _ L _ L _ C _ P _ S _ G _ H _ V _ V _ G _ I _ F _ R _ A _ A _

541 CGTGTGTACCAGAGGCGTGGCCAAGGCCGTGGATTTTCATCCCCGTGGAGAGCATGGAGAC
   -----+-----+-----+-----+-----+
   GCACACATGGTCTCCGCACCGGTTCCGGCACCTAAAGTAGGGGCACCTCTCGTACCTCTG
   V _ C _ T _ R _ G _ V _ A _ K _ A _ V _ D _ F _ I _ P _ V _ E _ S _ M _ E _ T _

601 CACCATGAGGAGCCCCGTGTTACCGACAATAGCAGCCCCCTGCCGTGCCTCAGACCTT
   -----+-----+-----+-----+-----+
   GTGGTACTCCTCGGGGCACAAGTGGCTGTTATCGTCGGGGGACGGCACGGAGTCTGGAA
   T _ M _ R _ S _ P _ V _ F _ T _ D _ N _ S _ S _ P _ P _ A _ V _ P _ Q _ T _ F _

CCAGGTGGCCCCACCTGCACGCCCCACCGGCTCCGGCAAGAGCACCAAGGTGCCAGCCGC

```

661 -----+-----+-----+-----+-----+
 GGTCCACCGGGTGGACGTGCGGGGGTGGCCGAGGCCGTTCTCGTGGTTCCACGGTCGGCG
Q V A H L H A P T G S G K S T K V P A A
 CTACGCCGCCAGGGCTACAAGGTGCTGGTGTGAATCCCAGCGTGGCCGCCACCTGGG
 721 -----+-----+-----+-----+-----+
 GATGCGGCGGGTCCCGATGTTCCACGACCACGACTTAGGGTCGCACCGGCGGTGGGACCC
Y A A Q G Y K V L V L N P S V A A T L G
 CTTGCGCGCTACATGAGCAAGGCCACGGCATCGAGCCCAATATCCGGACCGGAGTGAG
 781 -----+-----+-----+-----+-----+
 GAAGCCGCGGATGTACTCGTTCCGGGTGCGGTAGCTCGGGTTATAGGCCTGGCCTCACTC
F G A Y M S K A H G I E P N I R T G V R
 GACCATCACCACAGGCGGGCCCTATCACCTACAGCGCCTACGGCAAGTTCCTGGCCGACGG
 841 -----+-----+-----+-----+-----+
 CTGGTAGTGGTGTCCGCCGGATAGTGGATGTCGCGGATGCCGTTCAAGGACCGGTGCC
T I T T G G P I T Y S A Y G K F L A D G
 CGGCTGTAGCGGCGGAGCCTACGACATCATCTGTGACGAGTGCCACAGCACCGATTG
 901 -----+-----+-----+-----+-----+
 GCCGACATCGCCGCTCGGATGCTGTAGTAGTAGACACTGCTCACGGTGTCTGGCTAAC
G C S G G A Y D I I I C D E C H S T D W
 GACCACCATCCTGGGCATCGGCACCGTGCTGGACCAGGCCGAGACCGCCGGAGCCAGACT
 961 -----+-----+-----+-----+-----+
 CTGGTGGTAGGACCCGTAGCCGTGGCACGACCTGGTCCGGCTCTGGCGGCCTCGGTCTGA
T T I L G I G T V L D Q A E T A G A R L
 GGTGGTGTGGCCACAGCCACACCCCTGGCAGCATCACCGTGCCCCACCCCAACATCGA
 1021 -----+-----+-----+-----+-----+
 CCACCACGACCGGTGTGCGGTGTGGGGGACCGTCGTAGTGGCACGGGGTGGGGTTGTAGCT
V V L A T A T P P G S I T V P H P N I E
 GGAGGTGGCCCTGAGCAACACCGGCGAGATCCCCTTCTACGGCAAGGCCATCCCTATCGA
 1081 -----+-----+-----+-----+-----+
 CCTCCACCGGGACTCGTTGTGGCCGCTCTAGGGGAAGATGCCGTTCCGGTAGGGATAGCT
E V A L S N T G E I P F Y G K A I P I E
 GGCCATCAAGGGCGGCAGACACCTGATCTTCTGCCACAGCAAGAAGAAGTGTGACGAGCT
 1141 -----+-----+-----+-----+-----+
 CCGGTAGTTCGCCGCTCTGTGGACTAGAAGACGGTGTGTTCTTCTTCACTGCTCGA
A I K G G R H L I F C H S K K K C D E L
 GGCCGCCAAGCTGACCGGCCTGGGCCTGAACGCGTGGCCTACTACAGAGGCCTGGACGT
 1201 -----+-----+-----+-----+-----+
 CCGGCGGTTCGACTGGCCGGACCCGGAAGTTCGCGGCACCGGATGATGTCTCCGGACCTGCA
A A K L T G L G L N A V A Y Y R G L D V
 GTCCGTGATCCCTACCAGCGGCGATGTGGTGGTGGTGGCCACCGACGCCCTGATGACCGG
 1261 -----+-----+-----+-----+-----+
 CAGGCACTAGGGATGGTTCGCCGCTACACCACCACCACCGGTGGCTGCGGGACTACTGGCC
S V I P T S G D V V V V A T D A L M T G
 CTTACCGGCGATTTCGACAGCGTGATCGACTGTAATACCTGTGTGACCCAGACCGTGGA
 1321 -----+-----+-----+-----+-----+
 GAAGTGGCCGCTAAAGCTGTGCACTAGCTGACATTATGGACACACTGGGTCTGGCACCT
F T G D F D S V I D C N T C V T Q T V D
 CTTACGCCTGGACCCACCTTCACCATCGAGACCACCACCGTGGCCACAGGATGCCGTGTC
 1381 -----+-----+-----+-----+-----+
 GAAGTCGGACCTGGGGTGGAGTGGTAGCTCTGGTGGTGGCACGGTGTCTACGGCACAG
F S L D P T F T I E T T T V P Q D A V S

CAGAAGCCAGAGAAGAGGGCGCCACCGGCAGAGGCAGAAGCGGCATCTACAGATTTCGTGAC
 1441-----+-----+-----+-----+-----+-----+
 GTCTTCGGTCTCTTCTCCGCGGTGGCCGTCTCCGTCTTCGCCGTAGATGTCTAAGCACTG
R S Q R R G A T G R G R S G I Y R F V T
 CCCTGGCGAGAGACCCAGCGGCATGTTTCGATAGCAGCGTGTGTGTGAGTGCTACGACGC
 1501-----+-----+-----+-----+-----+-----+
 GGGACCGCTCTCTGGGTGCGCGTACAAGCTATCGTCGCACGACACACTCACGATGCTGCG
P G E R P S G M F D S S V L C E C Y D A
 CGGCTGTGCCTGGTACGAGCTGACCCCAGCCGAGACCACAGTGAGGCTGAGGGCCTACCT
 1561-----+-----+-----+-----+-----+-----+
 GCCGACACGGACCATGCTCGACTGGGGTTCGGTCTGCTGTCACCTCCGACTCCCGGATGGA
G C A W Y E L T P A E T T V R L R A Y L
 GAACACCCCTGGCCTGCCTGTGTGCCAGGATCACCTGGAGTTCTGGGAGAGCGTGTTTAC
 1621-----+-----+-----+-----+-----+-----+
 CTTGTGGGGACCGGACGGACACACGGTCTTAGTGGACCTCAAGACCCTCTCGCACAAATG
N T P G L P V C Q D H L E F W E S V F T
 CGGCCTGACCCACATCGATGCCCCACTTTCTGAGCCAGACCAAGCAGGCCGGCGACAACCT
 1681-----+-----+-----+-----+-----+-----+
 GCCGGACTGGGTGTAGCTACGGGTGAAAGACTCGGTCTGGTTTCGTCCGGCCGCTGTTGAA
G L T H I D A H F L S Q T K Q A G D N F
 CCCCTACCTGGTGGCCTACCAGGCCACCGTGTGTGCCAGAGCCCAGGCCCTCCCCCAG
 1741-----+-----+-----+-----+-----+-----+
 GGGGATGGACCACCGGATGGTCCGGTGGCACACACGGTCTCGGGTCCGGGGAGGGGGGTC
P Y L V A Y Q A T V C A R A Q A P P P S
 CTGGGACCAGATGTGGAAGTGCCTGATCAGGCTGAAGCCCACCCTGCACGGCCCTACCCC
 1801-----+-----+-----+-----+-----+-----+
 GACCCTGGTCTACACCTTCACGGACTAGTCCGACTTCGGGTGGGACGTGCCGGGATGGGG
W D Q M W K C L I R L K P T L H G P T P
 CCTGCTGTACAGACTGGGCGCCGTGCAGAATGAGATCACCTGACCCACCCTATCACCAA
 1861-----+-----+-----+-----+-----+-----+
 GGACGACATGTCTGACCCGCGGCACGTCTTACTCTAGTGGGACTGGGTGGGATAGTGGTT
L L Y R L G A V Q N E I T L T H P I T K
 GTTCGTGATGGCCTGTATGAGCGCCGACCTGGAGGTGGTGGGCAGCGGCTCCGGCTCCAT
 1921-----+-----+-----+-----+-----+-----+
 CAAGCACTACCGGACATACTCGCGGCTGGACCTCCACCACCCGTCGCCGAGGCCGAGGTA
F V M A C M S A D L E V V G S G S G S M
 GAGCTACACCTGGACAGGCGCCCTGATCACACCCTGTGCCGCCGAGGAGAGCAAGCTGCC
 1981-----+-----+-----+-----+-----+-----+
 CTCGATGTGGACCTGTCCGCGGGACTAGTGTGGGACACGGCGGCTCCTCTCGTTCGACGG
S Y T W T G A L I T P C A A E E S K L P
 CATCAACCCCCTGAGCAATAGCCTGCTGAGGCACCCACAGCATGGTGTACAGCACCACTC
 2041-----+-----+-----+-----+-----+-----+
 GTAGTTGGGGGACTCGTTATCGGACGACTCCGTGGTGTCTGACCATGTCTGTTGGGAG
I N P L S N S L L R H H S M V Y S T T S
 CAGAAGCGCCAGCCTGAGGCAGAAGAAGGTGACCTTCGACAGGCTGCAGGTGCTGGACGA
 2101-----+-----+-----+-----+-----+-----+
 GTCTTCGCGGTGCGACTCCGTCTTCTTCCACTGGAAGCTGTCCGACGTCCACGACCTGCT
R S A S L R Q K K V T F D R L Q V L D D
 CCACTACAGGGACGTGCTGAAGGAGATGAAGGCCAAGGCCAGCACCGTGAAGGCCAGACT
 2161-----+-----+-----+-----+-----+-----+

GGTGATGTCCCTGCACGACTTCCTCTACTTCCGGTTCCGGTCGTGGCACTTCCGGTCTGA
H Y R D V L K E M K A K A S T V K A R L
GCTGTCTATCGAGGAGGCTGTAAGCTGACCCCCCTCACAGCGCCAAGAGCAAGTTCCG
2221-----+-----+-----+-----+-----+-----+
CGACAGATAGCTCCTCCGGACATTCGACTGGGGGGGAGTGTGCGGGTCTCGTTCAAGCC
L S I E E A C K L T P P H S A K S K F G
CTACGGCGCCAAGGATGTGAGAAGCCTGAGCAGCAGAGCCGTGAACCACATCCGGTCTGT
2281-----+-----+-----+-----+-----+-----+
GATGCCGCGGTTCTTACACTCTTCGGA CTGTCGTCTCGGCACTTGGTGTAGGCCAGACA
Y G A K D V R S L S S R A V N H I R S V
GTGGGAGGATCTGCTGGAGGATACCGAGACCCCCATCGACACCACAATCATGGCCAAGAA
2341-----+-----+-----+-----+-----+-----+
CACCTCCTAGACGACCTCCTATGGCTCTGGGGGTAGCTGTGGTGTAGTACCGGTTCTT
W E D L L E D T E T P I D T T I M A K N
CGAGGTGTTCTGCGTGCAGCCCGAGAAGGGCGGAAGAAAGCCCGCCAGGCTGATCGTGT
2401-----+-----+-----+-----+-----+-----+
GCTCCACAAGACGCACGTGCGGGCTCTTCCCGCCTTCTTTTCGGGCGGTCCGACTAGCACAA
E V F C V Q P E K G G R K P A R L I V F
CCCTGACCTGGGAGTGAGAGTGTGTGAGAAGATGGCCCTGTACGACGTGGTGTCCACCCT
2461-----+-----+-----+-----+-----+-----+
GGGACTGGACCCTCACTCTCACACACTCTTCTACCGGGACATGCTGCACCACAGGTGGGA
P D L G V R V C E K M A L Y D V V S T L
GCCCCAGGCCGTGATGGGCCCCAGCTACGGCTTCCAGTACAGCCCTGGCCAGAGAGTGGA
2521-----+-----+-----+-----+-----+-----+
CGGGGTCCGGCACTACCCGGGGTCGATGCCGAAGGTCATGTGCGGACCGGTCTCTCACCT
P Q A V M G P S Y G F Q Y S P G Q R V E
GTTCTTGGTGAACACCTGGAAGAGCAAGAAATGCCCCATGGGCTTCAGCTACAACACCCG
2581-----+-----+-----+-----+-----+-----+
CAAGGACCACTTGTGGACCTTCTCGTTCTTTACGGGGTACCCGAAGTCGATGTTGTGGGC
F L V N T W K S K K C P M G F S Y N T R
GTGCTTCGACAGCACAGTGACCGAGAACGACATCAGGACCGAGGAGTCCATCTACCAGTG
2641-----+-----+-----+-----+-----+-----+
CACGAAGCTGTGTCGTCCTGCTCTTGTCTGTAGTCCTGGCTCCTCAGGTAGATGGTCAC
C F D S T V T E N D I R T E E S I Y Q C
CTGTGACCTGGCCCCCGAGGCCAGACAGGCCATCAAAAGCCTGACCGAGCGGCTGTACAT
2701-----+-----+-----+-----+-----+-----+
GACACTGGACCGGGGGCTCCGGTCTGTCCGGTAGTTTTTCGGA CTGCTCGCCGACATGTA
C D L A P E A R Q A I K S L T E R L Y I
CGGCGGACCTCTGACCAACAGCAAGGGCCAGAACTGTGGCTACAGAAGATGTAGGGCCAG
2761-----+-----+-----+-----+-----+-----+
GCCGCCTGGAGACTGGTTGTGCTGTTCCCGTCTTGACACCGATGTCTTCTACATCCCGGTC
G G P L T N S K G Q N C G Y R R C R A S
CGGCGTGCTGACCACCTCTTGTGGCAACACCCTGACCTGTTACCTGAAGGCCACCGCCGC
2821-----+-----+-----+-----+-----+-----+
GCCGCACGACTGGTGGAGAACACCGTTGTGGGACTGGACAATGGACTTCCGGTGGCGGCG
G V L T T S C G N T L T C Y L K A T A A
CTGTAGAGCCGCCAAACTGCAGGACTGTACCATGCTGCTGAACGGCAACGACCTGGTGGT
2881-----+-----+-----+-----+-----+-----+
GACATCTCGGCGGTTTGACGTCTTGACATGGTACGACCACTTGCCGTTGCTGGACCACCA
C R A A K L Q D C T M L V N G N D L V V

GATCTGTGAGAGCGCCGGCACCCAGGAGGATGCCGCTCCCTGAGAGTGTTTACCGAGGC
 2941-----+-----+-----+-----+-----+-----+-----+
 CTAGACACTCTCGCGGCCGTGGGTCTCTACGGCGGAGGGACTCTCACAAATGGCTCCG
I C E S A G T Q E D A A S L R V F T E A
 CATGACCAGATACAGCGCCCCCTCCCGCGACCCCTCCCCAGCCCCAGTACGATCTGGAGCT
 3001-----+-----+-----+-----+-----+-----+-----+
 GTACTGGTCTATGTGCGGGGAGGGCCGCTGGGAGGGGTGCGGCTCATGCTAGACCTCGA
M T R Y S A P P G D P P Q P E Y D L E L
 GATCACCAGCTGCTCCAGCAATGTGGCGGTGGCTCAGCAGCCAGCGGCAAGAGAGTGTA
 3061-----+-----+-----+-----+-----+-----+-----+
 CTAGTGGTCGACGAGGTCGTTACACCCGACCCAGTGCTGCGGTGCGCGTTCTCTCACAT
I T S C S S N V G V A H D A S G K R V Y
 CTACCTGACCAGGGACCCACCCCTCTGGCCAGAGCCGCTGGGAGACAGTGAGACA
 3121-----+-----+-----+-----+-----+-----+-----+
 GATGGACTGGTCCCTGGGGTGGTGGGAGACCGGTCTCGCGGACCCCTCTGTCACTCTGT
Y L T R D P T T P L A R A A W E T V R H
 CACCCCGTGAACAGCTGGCTGGGCAACATCATCATGTACGCCCTACCCTGTGGGCCAG
 3181-----+-----+-----+-----+-----+-----+-----+
 GTGGGGGCACTTGTGACCGACCCGTTGTAGTAGTACATGCGGGGATGGGACACCCGGTC
T P V N S W L G N I I M Y A P T L W A R
 AATGATCCTGATGACCCACTTCTTCAGCATCCTGCTGGCCCAGGAGCAGCTGGAGAAGGC
 3241-----+-----+-----+-----+-----+-----+-----+
 TTACTAGGACTACTGGGTGAAGAAGTCGTAGGACGACCGGCTCCTCGTACCTCTTCCG
M I L M T H F F S I L L A Q E Q L E K A
 CCTGGACTGCCAGATCTACGGCGCCTGCTACAGCATCGAGGCTCTGGACCTGCCTCAGAT
 3301-----+-----+-----+-----+-----+-----+-----+
 GGACCTGACGGTCTAGATGCCGCGGACGATGTCTGTAGCTCGGAGACCTGGACGGAGTCTA
L D C Q I Y G A C Y S I E P L D L P Q I
 CATCGAGAGACTGCACGGCCTGAGCGCCTTCAGCCTGCACAGCTACAGCCCAGGCGAGAT
 3361-----+-----+-----+-----+-----+-----+-----+
 GTAGCTCTCTGACGTGCCGGAAGTCGGACGTGTGATGTGCGGTCCGCTCTA
I E R L H G L S A F S L H S Y S P G E I
 CAATAGAGTGGCCAGCTGCCTGAGAAAGCTGGGCGTGCCACCTCTGAGAGTGTGGCGGCA
 3421-----+-----+-----+-----+-----+-----+-----+
 GTTATCTCACCGGTGACGGACTCTTTCGACCCGCACGGTGGAGACTCTCACACCGCCGT
N R V A S C L R K L G V P P L R V W R H
 CAGAGCCAGATCTGTGAGGGCCAAGCTGCTGTCCCAGGGCGGCAGGGCCGCCACCTGTGG
 3481-----+-----+-----+-----+-----+-----+-----+
 GTCTCGGTCTAGACACTCCCGGTTGACGACAGGGTCCCGCCGTCCCGGCGGTGGACACC
R A R S V R A K L L S Q G G R A A T C G
 CAAGTACCTGTTCAACTGGGCCGTGAGGACAAAGCTGAAGCTGACACCCATCCCTGCCGC
 3541-----+-----+-----+-----+-----+-----+-----+
 GTTCATGGACAAGTTGACCCGGCACTCCTGTTTCGACTTCGACTGTGGGTAGGGACGGCG
K Y L F N W A V R T K L K L T P I P A A
 CAGCCAGCTGGACCTGAGCGGCTGGTTCGTGGCCGGCTACAATGGCGGCGACATCTACCA
 3601-----+-----+-----+-----+-----+-----+-----+
 GTCGGTGCACCTGGACTCGCCGACCAAGCACCAGGCGGATGTTACCGCGCTGTAGATGGT
S Q L D L S G W F V A G Y N G G D I Y H

10/45

CAGCCTGTCCAGGGCCAGGCCTAGATGATGAGGAGCTCCAGCTTT
3661-----+-----+-----+-----+-----
GTCGGACAGGTCCCGGTCCGGATCTACTACTCCTCGAGGTCGAAA
S L S R A R P R * *

Proteína de fusão 4A-3-4B-5B (Hs) otimizada para Homo sapiens

KpnI *NcoI*

```

1  GGGCGAATTGGGTACCGCCACCATGGGCAGCGTGGTGATTGTGGGCCGGATCATCCTGAG
   -----+-----+-----+-----+-----+-----+-----+
   CCCGCTTAACCCATGGCGGTGGTACCCGTCGCACCACTAACACCCGGCCTAGTAGGACTC
               M  G  S  V  V  I  V  G  R  I  I  L  S
61  CGGCAGCGGCAGCGCCCCATCACCGCCTACAGCCAGCAGACCAGAGGCCTGCTGGGCTG
   -----+-----+-----+-----+-----+-----+-----+
   GCCGTCGCCGTCGCGGGGGTAGTGGCGGATGTGCGTCTGCTGGTCTCCGGACGACCCGAC
   G  S  G  S  A  P  I  T  A  Y  S  Q  Q  T  R  G  L  L  G  C
121  TATCATCACCAGCCTGACCGGCAGAGACAAGAATCAGGTGGACGGCGAGGTGCAGGTGCT
   -----+-----+-----+-----+-----+-----+-----+
   ATAGTAGTGGTGGACTGGCCGTCCTCTGTTCTTAGTCCACCTGCCGCTCCACGTCCACGA
   I  I  T  S  L  T  G  R  D  K  N  Q  V  D  G  E  V  Q  V  L
181  GTCCACCGCCACCCAGAGCTTCCTGGCCACCTGTGTGAATGGCGTGTGTTGGACCGTGTA
   -----+-----+-----+-----+-----+-----+-----+
   CAGGTGGCGGTGGGTCTCGAAGGACCGGTGGACACACTTACCGCACACAACCTGGCACAT
   S  T  A  T  Q  S  F  L  A  T  C  V  N  G  V  C  W  T  V  Y
241  CGCCGGAGCCGGCAGCAAGACCCTGGCCGGACCCAAGGGCCCCATCACCCAGATGTACAC
   -----+-----+-----+-----+-----+-----+-----+
   GCGCCCTCGGCCGTCGTTCTGGGACCGCCCTGGGTTCCCGGGCTAGTGGGTCTACATGTG
   A  G  A  G  S  K  T  L  A  G  P  K  G  P  I  T  Q  M  Y  T
301  CAACGTGGACCAGGACCTGGTGGGCTGGCCTGCCCCTCCTGGCGCCAGAAGCATGACCCC
   -----+-----+-----+-----+-----+-----+-----+
   GTTGACCTGGTCTGACCAACCCGACCGGACGGGAGGACCGCGTCTTCGTACTGGGG
   N  V  D  Q  D  L  V  G  W  P  A  P  P  G  A  R  S  M  T  P
361  TTGTACCTGTGGCAGCAGCGACCTGTACCTGGTGACAGACACGCCGATGTGATCCCTGT
   -----+-----+-----+-----+-----+-----+-----+
   AACATGGACACCGTCGTCGCTGGACATGGACCACTGGTCTGTGCGGCTACACTAGGGACA
   C  T  C  G  S  S  D  L  Y  L  V  T  R  H  A  D  V  I  P  V
421  GAGGAGGAGAGGCGATAGCAGAGGCAGCCTGCTGTCTCCTAGACCCGTGTCCTACCTGAA
   -----+-----+-----+-----+-----+-----+-----+
   CTCCTCCTCTCCGCTATCGTCTCCGTCGGACGACAGAGGATCTGGGCACAGGATGGACTT
   R  R  R  G  D  S  R  G  S  L  L  S  P  R  P  V  S  Y  L  K
481  GGGCAGCAGCGGCGGACCCCTGCTGTGCCCCAGCGGCCACGTGGTGGGCATCTTCAGAGC
   -----+-----+-----+-----+-----+-----+-----+
   CCCGTCGTCGCCGCTGGGGACGACACGGGGTCGCCGCTGCACCACCCGTAGAAGTCTCG
   G  S  S  G  G  P  L  I  C  P  S  G  H  V  V  G  I  F  R  A
541  CGCCGTGTGTACCAGAGGCGTGGCCAAGGCCGTGGATTTCATCCCCGTGGAGAGCATGGA
   -----+-----+-----+-----+-----+-----+-----+
   GCGGCACACATGGTCTCCGCACCGGTTCCGGCACCTAAAGTAGGGGCACCTCTCGTACCT
   A  V  C  T  R  G  V  A  K  A  V  D  F  I  P  V  E  S  M  E
601  GACCACCATGAGGAGCCCCGTGTTACCGACAAATAGCAGCCCCCTGCCGTGCCTCAGAC
   -----+-----+-----+-----+-----+-----+-----+
   CTGGTGGTACTCCTCGGGGCACAAGTGGCTGTTATCGTCGGGGGACGGCACGGAGTCTG
   T  T  M  R  S  P  V  F  T  D  N  S  S  P  P  A  V  P  Q  T
661  CTTCCAGGTGGCCACCTGCACGCCCCCACCGGCTCCGGCAAGAGCACCAAGGTGCCAGC
   -----+-----+-----+-----+-----+-----+-----+
   GAAGGTCCACCGGTGGACGTGCGGGGGTGGCCGAGGCCGTTCTCGTGGTTCCACGGTCG

```

F Q V A H L H A P T G S G K S T K V P A
 CGCCTACGCCGCCAGGGCTACAAGGTGCTGGTGAATCCCAGCGTGGCCGCCACCCT
 721 -----+-----+-----+-----+-----+-----+
 GCGGATGCGGCGGGTCCCGATGTTCCACGACCACGACTTAGGGTCGCACCGGCGGTGGGA
A Y A A Q G Y K V L V L N P S V A A T L
 GGGCTTCGGCGCCTACATGAGCAAGGCCACGGCATCGAGCCCAATATCCGGACCGGAGT
 781 -----+-----+-----+-----+-----+-----+
 CCCGAAGCCGCGGATGTACTCGTTCCGGGTGCCGTAGCTCGGGTTATAGGCCTGGCCTCA
G F G A Y M S K A H G I E P N I R T G V
 GAGGACCATCACACAGGCGGCCCTATCACCTACAGCGCCTACGGCAAGTTCCTGGCCGA
 841 -----+-----+-----+-----+-----+-----+
 CTCCTGGTAGTGGTGTCCGCCGGGATAGTGGATGTCGCGGATGCCGTTCAAGGACCGGCT
R T I T T G G P I T Y S A Y G K F L A D
 CGGCGGCTGTAGCGGCGGAGCCTACGACATCATCTGTGACGAGTGCCACAGCACCGA
 901 -----+-----+-----+-----+-----+-----+
 GCGCGCGACATCGCCGCCTCGGATGCTGTAGTAGACACTGCTCACGGTGTCTGGTGGCT
G G C S G G A Y D I I I C D E C H S T D
 TTGGACCACCATCCTGGGCATCGGCACCGTGCTGGACCAGGCCGAGACCGCCGGAGCCAG
 961 -----+-----+-----+-----+-----+-----+
 AACCTGGTGGTAGGACCCGTAGCCGTGGCAGCACCTGGTCCGGCTCTGGCGGCCTCGGTC
W T T I L G I G T V L D Q A E T A G A R
 ACTGGTGGTGGTGGCCACAGCCACACCCCTGGCAGCATCACCGTGCCCCACCCCAACAT
 1021 -----+-----+-----+-----+-----+-----+
 TGACCACCACGACCGGTGTCTGGTGTGGGGGACCGTCGTAGTGGCACGGGTGGGGTTGTA
L V V L A T A T P P G S I T V P H P N I
 CGAGGAGGTGGCCCTGAGCAACACCGGCGAGATCCCCTTCTACGGCAAGGCCATCCCTAT
 1081 -----+-----+-----+-----+-----+-----+
 GCTCCTCCACCGGGACTCGTTGTGGCCGCTCTAGGGGAAGATGCCGTTCCGGTAGGGATA
E E V A L S N T G E I P F Y G K A I P I
 CGAGGCCATCAAGGGCGGCAGACACCTGATCTTCTGCCACAGCAAGAAGAAGTGTGACGA
 1141 -----+-----+-----+-----+-----+-----+
 GCTCCGGTAGTTCCCGCCGTCTGTGGACTAGAAGACGGTGTCTTCTTTCACACTGCT
E A I K G C R H L I F U H S K K K C D E
 GCTGGCCGCCAAGCTGACCGGCCTGGGCCTGAACGCCGTGGCCTACTACAGAGGCCTGGA
 1201 -----+-----+-----+-----+-----+-----+
 CGACCGGCGGTTGACTGGCCGGACCCGACTTGCGGCACCGGATGATGTCTCCGGACCT
L A A K L T G L G L N A V A Y Y R G L D
 CGTGTCCGTGATCCCTACCAGCGCGATGTGGTGGTGGTGGCCACCGACGCCCTGATGAC
 1261 -----+-----+-----+-----+-----+-----+
 GCACAGGCACCTAGGGATGGTCGCCGCTACACCACCACCACCGGTGGCTGCGGGACTACTG
V S V I P T S G D V V V V A T D A L M T
 CGGCTTACCGGCGGATTTCCGACAGCGTGATCGACTGTAATACCTGTGTGACCCAGACCGT
 1321 -----+-----+-----+-----+-----+-----+
 GCCGAAGTGGCCGCTAAAGCTGTGCACTAGCTGACATTATGGACACACTGGGTCTGGCA
G F T G D F D S V I D C N T C V T Q T V
 GGACTTCAGCCTGGACCCACCTTCACCATCGAGACCACCACCGTGCCACAGGATGCCGT
 1381 -----+-----+-----+-----+-----+-----+
 CCTGAAGTCCGACCTGGGGTGGAAAGTGGTAGCTCTGGTGGTGGCACGGTGTCTACGGCA
D F S L D P T F T I E T T T V P Q D A V
 GTCCAGAGCCAGAGAGGAGGCGCACCGCTAGAGCCAGAAGCGGCATCTACAGATTCGT
 1441 -----+-----+-----+-----+-----+-----+

CAGGTCTTCGGTCTCTTCTCCGCGGTGGCCGTCTCCGTCTTCGCCGTAGATGTCTAAGCA
S R S Q R R G A T G R G R S G I Y R F V
 GACCCCTGGCGAGAGACCCAGCGGCATGTTTCGATAGCAGCGTGCTGTGTGAGTGCTACGA
 1501-----+-----+-----+-----+-----+-----+-----+
 CTGGGGACCGCTCTCTGGGTTCGCGGTACAAGCTATCGTCGCACGACACACTCACGATGCT
T P G E R P S G M F D S S V L C E C Y D
 CGCCGGCTGTGCCTGGTACGAGCTGACCCAGCCGAGACCACAGTGAGGCTGAGGGCCTA
 1561-----+-----+-----+-----+-----+-----+-----+
 GCGGCCGACACGGACCATGCTCGACTGGGGTCGGCTCTGGTGTCCTCCGACTCCCGGAT
A G C A W Y E L T P A E T T V R L R A Y
 CCTGAACACCCCTGGCCTGCCTGTGTGCCAGGATCACCTGGAGTTCTGGGAGAGCGTGTT
 1621-----+-----+-----+-----+-----+-----+-----+
 GGACTTGTGGGGACCGGACGACACACGGTCCTAGTGGACCTCAAGACCCTCTCGCACAA
L N T P G L P V C Q D H L E F W E S V F
 TACCGGCCTGACCCACATCGATGCCCACTTTCTGAGCCAGACCAAGCAGGCCGGCGACAA
 1681-----+-----+-----+-----+-----+-----+-----+
 ATGGCCGGACTGGGTGTAGCTACGGGTGAAGACTCGGTCTGGTTCTCGTCCGGCCCGCTGTT
T G L T H I D A H F L S Q T K Q A G D N
 CTTCCCTACCTGGTGGCCTACCAGGCCACCGTGTGTGCCAGAGCCCAGGCCCTCCCCC
 1741-----+-----+-----+-----+-----+-----+-----+
 GAAGGGGATGGACCACCGGATGGTCCGGTGGCACACACGGTCTCGGGTCCGGGGAGGGGG
F P Y L V A Y Q A T V C A R A Q A P P P
 CAGCTGGGACCAGATGTGGAAGTGCCTGATCAGGCTGAAGCCCACCCTGCACGGCCCTAC
 1801-----+-----+-----+-----+-----+-----+-----+
 GTCGACCCTGGTCTACACCTTCACGGACTAGTCCGACTTCGGGTGGGACGTGCCGGGATG
S W D Q M W K C L I R L K P T L H G P T
 CCCCCTGCTGTACAGACTGGGCGCCGTGCAGAATGAGATCACCCCTGACCCACCCTATCAC
 1861-----+-----+-----+-----+-----+-----+-----+
 GGGGGACGACATGTCTGACCCGCGGCACGTCTTACTCTAGTGGGACTGGGTGGGATAGTG
P L L Y R L G A V Q N E I T L T H P I T
 CAAGTTTCGTGATGGCCTGTATGAGCGCCGACCTGGAGGTGGTGTCCCTGATGGCCTTCAC
 1921-----+-----+-----+-----+-----+-----+-----+
 GTTCAAGCACTACCGGACATACTCGCGGTGGACCTCCACCACAGGGACTACCGGAAGTG
K F V M A C M S A D L E V V S L M A F T
 CGCCAGCATACAAGCCCCCTGACCACCCAGAATACCCTGCTGTTCAACATCCTGGGCGG
 1981-----+-----+-----+-----+-----+-----+-----+
 GCGGTCGTAGTGTTCGGGGGACTGGTGGGTCTTATGGGACGACAAGTTGTAGGACCCGCC
A S I T S P L T T Q N T L L F N I L G G
 CTGGGTGGCCGCCCAGCTGTCCATGAGCTACACCTGGACAGGCGCCCTGATCACACCCTG
 2041-----+-----+-----+-----+-----+-----+-----+
 GACCCACCGGCGGGTCGACAGGTACTCGATGTGGACCTGTCCGCGGGACTAGTGTGGGAC
W V A A Q L S M S Y T W T G A L I T P C
 TGCCGCCGAGGAGAGCAAGCTGCCCATCAACCCCTGAGCAATAGCCTGCTGAGGCACCA
 2101-----+-----+-----+-----+-----+-----+-----+
 ACGGCGGCTCCTCTCGTTTCGACGGGTAGTTGGGGGACTCGTTATCGGACGACTCCGTGGT
A A E E S K L P I N P L S N S L L R H H
 CAGCATGGTGTACAGCACCACTCCAGAAGCGCCAGCCTGAGGCAGAAGAAGGTGACCTT
 2161-----+-----+-----+-----+-----+-----+-----+
 GTCGTACCACATGTCGTGGTGGAGGTCTTCGCGGTTCGACTCCGTCTTCTTCCACTGGAA
S M V Y S T T S R S A S L R Q K K V T F

CGACAGGCTGCAGGTGCTGGACGACCACTACAGGGACGTGCTGAAGGAGATGAAGGCCAA
 2221-----+-----+-----+-----+-----+-----+-----+
 GCTGTCCGACGTCCACGACCTGCTGGTGATGTCCCTGCACGACTTCTCTACTTCCGGTT
D R L Q V L D D H Y R D V L K E M K A K

 GGCCAGCACCGTGAAGGCCAGACTGCTGTCTATCGAGGAGGCCTGTAAGCTGACCCCCC
 2281-----+-----+-----+-----+-----+-----+-----+
 CCGGTCGTGGCACTTCCGGTCTGACGACAGATAGCTCCTCCGGACATTGACTGGGGGGG
A S T V K A R L L S I E E A C K L T P P

 TCACAGCGCCAAGAGCAAGTTCGGCTACGGCGCCAAGGATGTGAGAAGCCTGAGCAGCAG
 2341-----+-----+-----+-----+-----+-----+-----+
 AGTGTGCGGGTTCTCGTTCAAGCCGATGCCGCGGTTCTTACACTCTTCGGACTCGTCGTC
H S A K S K F G Y G A K D V R S L S S R

 AGCCGTGAACCACATCCGGTCTGTGTGGGAGGATCTGCTGGAGGATACCGAGACCCCCAT
 2401-----+-----+-----+-----+-----+-----+-----+
 TCGGCACTTGGTGTAGGCCAGACACACCCTCCTAGACGACCTCCTATGGCTCTGGGGGTA
A V N H I R S V W E D L L E D T E T P I

 CGACACCACAATCATGGCCAAGAACGAGGTGTTCTGCGTGCAGCCCGAGAAGGGCGGAAG
 2461-----+-----+-----+-----+-----+-----+-----+
 GCTGTGGTGTAGTACCGGTTCTTGCTCCACAAGACGCACGTCGGGCTCTTCCCGCCTTC
D T T I M A K N E V F C V Q P E K G G R

 AAAGCCCGCCAGGCTGATCGTGTTCCTGACCTGGGAGTGAGAGTGTGTGAGAAGATGGC
 2521-----+-----+-----+-----+-----+-----+-----+
 TTTCCGGGCGGTCCGACTAGCACAAAGGGACTGGACCCTCACTCTCACACACTCTTCTACCG
K P A R L I V F P D L G V R V C E K M A

 CCTGTACGACGTGGTGTCCACCCTGCCCCAGGCCGTGATGGGCCCCAGCTACGGCTTCCA
 2581-----+-----+-----+-----+-----+-----+-----+
 GGACATGCTGCACCACAGGTGGGACGGGGTCCGGCACTACCCGGGGTTCGATGCCGAAGGT
L Y D V V S T L P Q A V M G P S Y G F Q

 GTACAGCCCTGGCCAGAGAGTGGAGTTCCTGGTGAACACCTGGAAGAGCAAGAAATGCCC
 2641-----+-----+-----+-----+-----+-----+-----+
 CATGTGCGGACCGGTCTCTACCTCAAGGACCACTTGTGGACCTTCTCGTTCTTTACGGG
Y S P G Q R V E F L V N T W K S K K C P
 CATGGGCTTCAGCTACAACACCCGGTGCTTCGACAGCACAGTGACCGAGAACGACATCAG
 2701-----+-----+-----+-----+-----+-----+-----+
 GTACCCGAAGTCGATGTTGTGGGCCACGAAGCTGTGCTGCTACTGGCTCTTGCTGTAGTC
M G F S Y N T R C F D S T V T E N D I R

 GACCGAGGAGTCCATCTACCAGTGCTGTGACCTGGCCCCGAGGCCAGACAGGCCATCAA
 2761-----+-----+-----+-----+-----+-----+-----+
 CTGGCTCCTCAGGTAGATGGTCACGACACTGGACCGGGGGCTCCGGTCTGTCCGGTAGTT
T E E S I Y Q C C D L A P E A R Q A I K

 AAGCCTGACCGAGCGGCTGTACATCGGCGGACCTCTGACCAACAGCAAGGGCCAGAACTG
 2821-----+-----+-----+-----+-----+-----+-----+
 TTCGGACTGGCTCGCCGACATGTAGCCGCCTGGAGACTGGTTGTGCGTTCCCGGTCTTGAC
S L T E R L Y I G G P L T N S K G Q N C

 TGGCTACAGAAGATGTAGGGCCAGCGCGTGCTGACCACCTCTGTGGCAACACCCCTGAC
 2881-----+-----+-----+-----+-----+-----+-----+
 ACCGATGTCTTCTACATCCCGGTGCGCGCACGACTGGTGGAGAACACCGTTGTGGGACTG
G Y R R C R A S G V L T T S C G N T L T

 CTGTTACCTGAAGGCCACCGCCGCTGTAGAGCCGCCAAACTGCAGGACTGTACCATGCT
 2941-----+-----+-----+-----+-----+-----+-----+
 GACAAATGGAATTCGGTGGCGGGGACATCTCGGCGGTTTGACGTCCTGACATGGTACGA
C Y L K A T A A C R A A K L Q D C T M L

GGTGAACGGCAACGACCTGGTGGTGATCTGTGAGAGCGCCGGCACCCAGGAGGATGCCGC
 3001-----+-----+-----+-----+-----+-----+
 CCACTTGCCGTTGCTGGACCACCACTAGACACTCTCGCGGCCGTGGGTCTCTCTACGGCG
V N G N D L V V I C E S A G T Q E D A A
 CTCCCTGAGAGTGTTTACCGAGGCCATGACCAGATACAGCGCCCCTCCCGGCGACCCTCC
 3061-----+-----+-----+-----+-----+-----+
 GAGGGACTCTCACAAATGGCTCCGGTACTGGTCTATGTGCGGGGAGGGCCGCTGGGAGG
S L R V F T E A M T R Y S A P P G D P P
 CCAGCCCGAGTACGATCTGGAGCTGATCACCAGCTGCTCCAGCAATGTGGGCGTGGCTCA
 3121-----+-----+-----+-----+-----+-----+
 GGTCGGGCTCATGCTAGACCTCGACTAGTGGTCGACGAGGTCGTTACACCCGCACCGAGT
Q P E Y D L E L I T S C S S N V G V A H
 CGACGCCAGCGGCAAGAGAGTGTAACCTGACCAGGGACCCACCCCTCTGGCCAG
 3181-----+-----+-----+-----+-----+-----+
 GCTGCGGTGCGCGTTCTCTCACATGATGGACTGGTCCCTGGGGTGGTGGGGAGACCGGTC
D A S G K R V Y Y L T R D P T T P L A R
 AGCCGCCTGGGAGACAGTGAGACACACCCCGTGAACAGCTGGCTGGGCAACATCATCAT
 3241-----+-----+-----+-----+-----+-----+
 TCGGCGGACCCCTCTGTCACTCTGTGTGGGGGCACTTGTCGACCGACCCGTTGTAGTAGTA
A A W E T V R H T P V N S W L G N I I M
 GTACGCCCTACCCTGTGGGCCAGAATGATCCTGATGACCCACTTCTTCAGCATCCTGCT
 3301-----+-----+-----+-----+-----+-----+
 CATGCGGGGATGGGACACCCGGTCTTACTAGGACTACTGGGTGAAGAAGTCGTAGGACGA
Y A P T L W A R M I L M T H F F S I L L
 GGCCAGGAGCAGCTGGAGAAGGCCCTGGACTGCCAGATCTACGGCGCCTGCTACAGCAT
 3361-----+-----+-----+-----+-----+-----+
 CCGGGTCTCTCGTCGACCTCTTCCGGGACCTGACGGTCTAGATGCCGCGGACGATGTCGTA
A Q E Q L E K A L D C Q I Y G A C Y S I
 CGAGCCTCTGGACCTGCCTCAGATCATCGAGAGACTGCACGGCCTGAGCGCCTTCAGCCT
 3421-----+-----+-----+-----+-----+-----+
 GCTCGGAGACCTGGACGGAGTCTAGTAGCTCTCTGACGTGCCGGACTCGCGGAAGTCGGA
E P L D L P Q I I E R L H G L S A F S L
 GCACAGCTACAGCCAGGCGAGATCAATAGAGTGGCCAGCTGCCTGAGAAAGCTGGGCGT
 3481-----+-----+-----+-----+-----+-----+
 CGGTGAGATGTCGGGTCCGCTCTAGTTATCTCACCGGTGACCGACTCTTTCGACCCGCA
H S Y S P G E I N R V A S C L R K L G V
 GCCACCTCTGAGAGTGTTGGCGGCACAGAGCCAGATCTGTGAGGGCCAAGCTGCTGTCCCA
 3541-----+-----+-----+-----+-----+-----+
 CGGTGGAGACTCTCACACCGCCGTGTCTCGGTCTAGACACTCCCGGTTTCGACGACAGGGT
P P L R V W R H R A R S V R A K L L S Q
 GGGCGGCAGGGCCGCCACCTGTGGCAAGTACCTGTTCAACTGGGCGGTGAGGACAAAGCT
 3601-----+-----+-----+-----+-----+-----+
 CCCGCCGTCCCGGCGGTGGACACCGTTTCATGGACAAGTTGACCCGGCACTCCTGTTTCGA
G G R A A T C G K Y L F N W A V R T K L
 GAAGCTGACACCCATCCCTGCCGCCAGCCAGCTGGACCTGAGCGGCTGGTTTCGTGGCCGG
 3661-----+-----+-----+-----+-----+-----+
 CTTGCACTGTGGGTAGGGACGGCGGTGCGTTCGACCTGGACTCGCCGACCAAGCACCAGGCC
K L T P I P A A S Q L D L S G W F V A G
 CTACAATGGCGGCGACATCTACCACAGCCTGTCCAGGGCCAGGCCTAGATGATGAGGAGC
 3721-----+-----+-----+-----+-----+-----+

16/45

GATGTTACCGCCGCTGTAGATGGTGTCTGGACAGGTCCCGGTCCGGATCTACTACTCCTCG

_Y_N_G_G_D_I_Y_H_S_L_S_R_A_R_P_R_*_*_

TCCAGCTTTTGTTCCC

3781-----+-----

AGGTCGAAAACAAGGG

Proteína de fusão 4A-3-5B
Otimizada para expressão em: *Pichia pastoris*

```

1  ATGGGGTCCGTTGTTATCGTTGGTAGAATCATCTTGTCTGGTTCTGGTTCCGCTCCAATT
   +-----+-----+-----+-----+-----+-----+
   TACCCCAGGCAACAATAGCAACCATCTTAGTAGAACAGACCAAGACCAAGGCGAGGTTAA
   M _ G _ S _ V _ V _ I _ V _ G _ R _ I _ I _ L _ S _ G _ S _ G _ S _ A _ P _ I _

61  ACTGCTTACTCCCAGCAGACTAGAGGATTGTTGGGTTGTATCATCACTTCCTTGACTGGT
   +-----+-----+-----+-----+-----+-----+
   TGACGAATGAGGGTCGTCTGATCTCCTAACAACCAACATAGTAGTGAAGGAACTGACCA
   T _ A _ Y _ S _ Q _ Q _ T _ R _ G _ L _ L _ G _ C _ I _ I _ T _ S _ L _ T _ G _

121  AGAGACAAGAACCAAGTTGACGGAGAGGTTTCAGGTTTTGTCCACTGCTACTCAGTCTTTC
   +-----+-----+-----+-----+-----+-----+
   TCCTCTGTTCTTGGTTCAACTGCCTCTCCAAGTCCAAAACAGGTGACGATGAGTCAGAAAG
   R _ D _ K _ N _ Q _ V _ D _ G _ E _ V _ Q _ V _ L _ S _ T _ A _ T _ Q _ S _ F _

181  TTGGCTACTTGTGTTAACGGTGTTTGTGGACTGTTTACGCTGGTGCTGGTTCTAAAACT
   +-----+-----+-----+-----+-----+-----+
   AACCGATGAACACAATTGCCACAAACAACCTGACAAATGCGACCACGACCAAGATTTTGA
   L _ A _ T _ C _ V _ N _ G _ V _ C _ W _ T _ V _ Y _ A _ G _ A _ G _ S _ K _ T _

241  TTGGCTGGTCCAAAGGGTCCAATCACTCAGATGTACACAAACGTTGACCAGGATTTGGTT
   +-----+-----+-----+-----+-----+-----+
   AACCGACCAGGTTTCCCAGGTTAGTGAGTCTACATGTGTTTGCAACTGGTCCTAAACCAA
   L _ A _ G _ P _ K _ G _ P _ I _ T _ Q _ M _ Y _ T _ N _ V _ D _ Q _ D _ L _ V _

301  GGTGGCCAGCTCCACCAGGTGCTAGATCCATGACTCCATGTACTTGTGGTTCTCCGAC
   +-----+-----+-----+-----+-----+-----+
   CCAACCGGTGCGAGGTGGTCCACGATCTACGTACTGAGGTACATGAACACCAAGGAGGCTG
   G _ W _ P _ A _ P _ P _ G _ A _ R _ S _ M _ T _ P _ C _ T _ C _ G _ S _ S _ D _

361  TTGTACTTGGTTACTAGACACGCTGACGTTATCCCAGTTAGAAGAAGAGGAGACTCCAGA
   +-----+-----+-----+-----+-----+-----+
   AACATGAACCAATGATCTGTGCGACTGCAATAGGGTCAATCTTCTCTCTCTGAGGTCT
   L _ Y _ L _ V _ T _ R _ H _ A _ D _ V _ I _ E _ V _ R _ R _ R _ G _ D _ S _ R _

421  GGATCTTTGTTGTCCCCAAGACCAGTTTCTTACTTGAAGGGATCTTCCGGTGGTCCATTG
   +-----+-----+-----+-----+-----+-----+
   CCTAGAAACAACAGGGGTTCTGGTCAAAGAATGAACCTCCCTAGAAGGCCACCAGGTAAC
   G _ S _ L _ L _ S _ P _ R _ P _ V _ S _ Y _ L _ K _ G _ S _ S _ G _ G _ P _ L _

481  TTGTGTCCATCCGGTCACGTTGTTGGTATTTTCAGAGCTGCTGTTTGTACTAGAGGTGTT
   +-----+-----+-----+-----+-----+-----+
   AACACAGGTAGGCCAGTGCAACAACCATAAAGTCTCGACGACAAACATGATCTCCACAA
   L _ C _ P _ S _ G _ H _ V _ V _ G _ I _ F _ R _ A _ A _ V _ C _ T _ R _ G _ V _

541  GCTAAGGCTGTTGACTTCATCCCAGTTGAGTCCATGGAGACTACTATGAGATCCCCAGTT
   +-----+-----+-----+-----+-----+-----+
   CGATTCCGACAACCTGAAGTAGGGTCAACTCAGGTACCTCTGATGATACTCTAGGGGTCAA
   A _ K _ A _ V _ D _ F _ I _ P _ V _ E _ S _ K _ E _ T _ T _ M _ R _ S _ P _ V _

601  TTCCTGACAACCTCTTCCCCACCTGCTGTTCACAAACTTTCCAAGTTGCTCACTTGCAAT
   +-----+-----+-----+-----+-----+-----+
   AAGTGAAGTGTGAGAAGGGTGGACGACAAGGTGTTTGAAGGTTCAACGAGTGAACGTA
   F _ T _ D _ N _ S _ S _ P _ P _ A _ V _ E _ Q _ T _ P _ Q _ V _ A _ H _ L _ H _

GCTCCAAGTGGTTCTGGTGAAGTCCACTAACGTTCCAGCTCCTTACGCTGCTCAAGGTTAC

```

661 -----+-----+-----+-----+-----+-----+
 CGAGGTTGACCAAGACCATTGAGGTGATTCCAAGGTCGACGAATGCGACGAGTTCCAATG
 A P T G S G K S T K V P A A Y A A Q G Y

 AAGGTTTTGGTTTTGAACCCATCCGTTGCTGCTACTTTGGGTTTCGGTGCTTACATGTCT
 721 -----+-----+-----+-----+-----+-----+
 TTCCAAAACCAAACCTTGGGTAGGCAACGACGATGAAACCCAAAGCCACGAATGTACAGA
 K V L V L N P S V A A T L G F G A Y M S

 AAGGCTCACGGTATTGAGCCAAACATCAGAACTGGTGTTAGAACTATCACTACTGGTGGT
 781 -----+-----+-----+-----+-----+-----+
 TTCCGAGTGCCATAACTCGGTTTGTAGTCTTGACCACAATCTTGATAGTGATGACCACCA
 K A H G I E P N I R T S V R T I T T G G

 CCTATTACTTACTCCGCTTACGGAAAGTTTTTGGCTGACGGTGGTTGTTCTGGTGGTGCT
 841 -----+-----+-----+-----+-----+-----+
 GGATAATGAATGAGGCGAATGCCTTTCAAAAACCGACTGCCACCAACAAGACCACCACGA
 P I T Y S A Y G K F L A D G G C S G G A

 TACGACATCATCATCTGTGACGAGTGTCCTCTACTGACTGGACTACTATCTTGGGTATC
 901 -----+-----+-----+-----+-----+-----+
 ATGCTGTAGTAGTAGACACTGCTCAGCTGAGATGACTGACCTGATGATAGAACCCATAG
 Y D I I I C D E C H S T D W T T I L G I

 GGTACTGTTTTGGACCAAGCTGAAACTGCTGGTGCTAGATTGGTTGTTTTGGCTACTGCT
 961 -----+-----+-----+-----+-----+-----+
 CCATGACAAAACCTGGTTCGACTTTGACGACCACGATCTAACCAACAAAACCGATGACGA
 G T V L D Q A E T A G A R L V V L A T A

 ACTCCACCAGGTTCCATTACTGTTCCACACCCCAAACATCGAGGAAGTTGCTTTGTCTAAC
 1021 -----+-----+-----+-----+-----+-----+
 TGAGGTGGTCCAAGGTAATGACAAGGTGTGGGTTGTAGCTCCTTCAACGAAACAGATTG
 T P P G S I T V P H P N I E E V A L S N

 ACTGGAGAGATCCCATTCTACGGAAAGGCTATCCCAATTGAGGCTATCAAGGGTGGTAGA
 1081 -----+-----+-----+-----+-----+-----+
 TGACCTCTCTAGCGTAAGATCCCTTTCCCATAGGGTTAACTCCGATAGTCCCACCATCT
 T G E I P F Y G K P I F I E A I K G C R

 CACTTGATTTTCTGTCACTCCAAGAAGAAGTGTGACGAGTTCGCTGCTAAGTTGACTGGA
 1141 -----+-----+-----+-----+-----+-----+
 GTGAACTAAAAGACAGTGAGGTTCTTCTTCACACTGCTCAACCGACGATTCAACTGACCT

 TTGGGATTGAACGCTGTTGCTTACTACAGAGGATTGGACGTTTCCGTTATCCCAACTTCC
 1201 -----+-----+-----+-----+-----+-----+
 AACCCTAACCTTGCACAAACGAATGATGTCTCTAACCTGCAAAGGCAATAGGGTTGAAGG
 L G L N A V A Y Y R G L D V S V I P T S

 GGTGATGTTGTTGTTGTTGCTACTGACGCTTTGATGACTGGTTTCACTGGTGACTTCGAC
 1261 -----+-----+-----+-----+-----+-----+
 CCACTACAACAACAACAACGATGACTGCGAAACTACTGACCAAAGTGACCACTGAAGCTG
 G D V V V V A T D A L N T G F T G D F D

 TCCGTTATCGACTGTAACACTTGTGTTACTCAGACTGTTGACTTCTCCTTGGACCCAACT
 1321 -----+-----+-----+-----+-----+-----+
 AGGCAATAGCTGACATTGTGACACAATGAGTCTGACAACCTGAAGAGGAACCTGGGTTGA
 S V I D C N T C V T Q T V D F S L D P T

 TTCCTATCGAGACTACTACTGTTCCTCAAGACGCTGTTTCCAGATCCCAAGAAGAGGT
 1381 -----+-----+-----+-----+-----+-----+
 AAGTGATAGCTCTGATGATGACAAGGAGTTCTGCGACAAAGGTCTAGGGTTTCTTCTCCA
 F T I E T T T V P Q D A V S R S Q R R G

GCTACTGGTAGAGGAAGATCCGGTATCTACAGATTCGTTACTCCAGGTGAAAGACCATCT
 1441-----+-----+-----+-----+-----+-----+
 CGATGACCATCTCCTTCTAGGCCATAGATGTCTAAGCAATGAGGTCCACTTTCTGGTAGA
 A T G R G R S G I Y R F V T P G E R P S

 GGAATGTTTCGACTCCTCCGTTTTGTGTGAATGTTACGACGCTGGTTGTGCTTGGTACGAA
 1501-----+-----+-----+-----+-----+-----+
 CCTTACAAGCTGAGGAGGCAAAACACACTTACAATGCTGCGACCAACACGAACCATGCTT
 G M F D S S V L C E C Y D A G C A W Y E

 TTGACTCCAGCTGAGACTACTGTTAGATTGAGAGCTTACTTGAACACTCCAGGATTGCCA
 1561-----+-----+-----+-----+-----+-----+
 AACTGAGGTGACTCTGATGACAATCTAATCTCGAATGAACCTGTGAGGTCCTAACGGT
 L T F A E T T V R L R A Y L N T P G L P

 GTTTGTCAAGACCACTTGGAAATCTGGGAGTCCGTTTTCACTGGATTGACTCACATTGAC
 1621-----+-----+-----+-----+-----+-----+
 CAAACAGTTCTGGTGAACCTTAAGACCCTCAGGCAAAAGTGACCTAACTGAGTGTAACGT
 V C Q D H L E F W E S V F T G L T H I D

 GCTCACTTTTTGTCCCAAATAAGCAGGCTGGTGACAACTTTCCATACTTGGTTGCTTAC
 1681-----+-----+-----+-----+-----+-----+
 CGAGTGAAAAACAGGGTTTGATTTCGTCGACCACTGTTGAAAGGTATGAACCAACGAATG
 A H F L S Q T K Q A G D N F P Y L V A Y

 CAGGCTACTGTTTGTGCTAGAGCACAAGCTCCACCACCATCTTGGGATCAGATGTGGAAG
 1741-----+-----+-----+-----+-----+-----+
 GTCCGATGACAAACACGATCTCGTGTTTCGAGGTGGTGGTAGAACCCCTAGTCTACACCTTC
 Q A T V C A R A Q A P P P S W D Q M W K

 TGTTTGATCAGATTGAAGCCAACTTTGCACGGTCCAACCTCCATTGTTGTACAGATTGGGT
 1801-----+-----+-----+-----+-----+-----+
 ACAAACCTAGTCTAACTTCGGTTGAAACGTGCCAGGTTGAGGTAACAACATGTCTAACCCA
 C L I R L K P T L H G P T P L L Y R L G

 GCTGTTTCAGAACGAGATCACTTTGACTCACCCAATCACTAAGTTCGTTATGGCTTGCATG
 1861-----+-----+-----+-----+-----+-----+
 CGACAAGTCTTGCTCTAGTGAAACTGAGTGGGTTAGTGATTCAAGCAATACCGAACGTAC
 A V Q N E I T L T H P I T K F V M A C M

 TCTGCTGACTTGGAAGTTGTTGGTTCCGGATCTGGTTCCATGTCCTACACTTGGACTGGT
 1921-----+-----+-----+-----+-----+-----+
 AGACGACTGAACCTTCAACAACCAAGGCCTAGACCAAGGTACAGGATGTGAACCTGACCA
 S A D L E V V G S G S G S M S Y T W T G

 GCTTTGATCACTCCATGTGCTGCTGAAGAATCCAAGTTGCCAATCAACCCATTGTCCAAC
 1981-----+-----+-----+-----+-----+-----+
 CGAAACTAGTGAGGTACACGACGACTTCTTAGGTTCAACGGTTAGTTGGGTAACAGGTTG
 A L I T P C A A E E S K L P I N P L S N

 TCTTTGTTGAGACACCACTCCATGGTTTACTCCACTACTTCCAGATCCGCTTCCTTGAGA
 2041-----+-----+-----+-----+-----+-----+
 AGAAACAACCTCTGTGGTGAGGTACCAAATGAGGTGATGAAGGTCTAGGCGAAGGAACCTCT
 S L L R H H S M V Y S T T S R S A S L R

 CAGAAGAAGGTTACATTGACAGATTGCAGGTTTTGACGACCACTACAGAGATGTTTTG
 2101-----+-----+-----+-----+-----+-----+
 GTCTTCTTCCAATGTAAGCTGTCTAACGTCCAAAACCTGCTGGTGATGTCTCTACAAAAC
 Q K K V T F D R L Q V L D D H Y R D V L

 AAGGAGATGAAGGCTAAGGCTTCCACTGTTAAGGCTAGATTGTTGTCCATTGAGGAGGCT
 2161-----+-----+-----+-----+-----+-----+
 TTCTCTACTTCCGATTCCGAAGGTGACAATTCCGATCTAACAACAGGTAACCTCCTCCGA

K _ E _ M _ K _ A _ K _ A _ S _ T _ V _ K _ A _ R _ L _ L _ S _ I _ E _ E _ A _
 TGTAAGTTGACTCCACCACACTCTGCTAAGTCCAAGTTTGGTTACGGTGCTAAGGATGTT
 2221-----+-----+-----+-----+-----+-----+-----+
 ACATTCAACTGAGGTGGTGTGAGACGATTGAGGTTCAAACCAATGCCACGATTCTACAA
 C _ K _ L _ T _ P _ P _ H _ S _ A _ K _ S _ K _ F _ G _ Y _ G _ A _ K _ D _ V _
 AGATCCTTGTCTCCAGAGCTGTTAACCACATCAGATCCGTTTGGGAGGATTTGTTGGAG
 2281-----+-----+-----+-----+-----+-----+-----+
 TCTAGGAACAGGAGGTCTCGACAATTGGTGTAGTCTAGGCAAACCCTCCTAAACAACCTC
 R _ S _ L _ S _ S _ R _ A _ V _ N _ H _ I _ R _ S _ V _ W _ E _ D _ L _ L _ E _
 GACACTGAGACTCCAATCGACACTACTATCATGGCTAAGAACGAGGTTTTCTGTGTTCAA
 2341-----+-----+-----+-----+-----+-----+-----+
 CTGTGACTCTGAGGTGAGGTTAGCTGTGATGATAGTACCGATTCTTGCTCCAAAAGACACAAGTT
 D _ T _ E _ T _ P _ I _ D _ T _ T _ I _ M _ A _ K _ N _ E _ V _ F _ C _ V _ Q _
 CCAGAGAAGGGTGAAGAAAGCCAGCTAGATTGATCGTTTTCCAGACTTGGGTGTTAGA
 2401-----+-----+-----+-----+-----+-----+-----+
 GGTCTCTTCCACCTTCTTTTCGGTTCGATCTAACTAGCAAAGGGTCTGAACCCACAATCT
 P _ E _ K _ G _ G _ R _ K _ P _ A _ R _ L _ I _ V _ F _ P _ D _ L _ G _ V _ R _
 GTTGTGAGAAGATGGCTTTGTACGACGTTGTTTCCACTTTGCCACAGGCTGTTATGGGA
 2461-----+-----+-----+-----+-----+-----+-----+
 CAAACACTCTTCTACCGAAACATGCTGCAACAAAGGTGAAACGGTGTCCGACAATACCTT
 V _ C _ E _ K _ M _ A _ L _ Y _ D _ V _ V _ S _ T _ L _ P _ Q _ A _ V _ M _ G _
 CCATCTTACGGTTTCCAATACTCCCCAGGACAAAGAGTTGAGTTCTTGGTTAACACTTGG
 2521-----+-----+-----+-----+-----+-----+-----+
 GGTAGAATGCCAAAGGTTATGAGGGGTCCTGTTTCTCAACTCAAGAACCAATGTGAACC
 P _ S _ Y _ G _ F _ Q _ Y _ S _ P _ G _ Q _ R _ V _ E _ F _ L _ V _ N _ T _ W _
 AAGTCCAAGAAATGTCCAATGGGATTCTCCTACAACACTAGATGTTTCGACTCCACTGTT
 2581-----+-----+-----+-----+-----+-----+-----+
 TTCAGGTTCTTTACAGGTTACCCTAAGAGGATGTTGTGATCTACAAAGCTGAGGTGACAA
 K _ S _ K _ K _ C _ P _ M _ G _ F _ S _ Y _ N _ T _ R _ C _ F _ D _ S _ T _ V _
 ACTGAGAACGACATCAGAACAGAGGAGTCCATCTACCAGTGTGTGACTTGGCTCCAGAA
 2641-----+-----+-----+-----+-----+-----+-----+
 TGACTCTTGCTGTAGTCTTGTCTCCTCAGGTAGATGGTCACAACACTGAACCGAGGTCTT
 T _ E _ N _ D _ I _ R _ T _ E _ E _ S _ I _ Y _ Q _ C _ C _ D _ L _ A _ P _ E _
 GCTAGACAAGCTATCAAGTCCTTGACTGAGAGATTGTACATCGGTGGTCTTTGACTAACC
 2701-----+-----+-----+-----+-----+-----+-----+
 CGATCTGTTTCGATAGTTTCAGGAAGTGAAGTCTCTAACATGTAGCCACCAGGAAACTGATTG
 A _ R _ Q _ A _ I _ K _ S _ L _ T _ E _ R _ L _ Y _ I _ G _ G _ P _ L _ T _ N _
 TCCAAGGGACAGAACTGTGGTTACAGAAGATGTAGAGCTTCCGGTGTGTTTACTACTTCC
 2761-----+-----+-----+-----+-----+-----+-----+
 AGGTTCCCTGTCTTGACACCAATGTCTTCTACATCTCGAAGGCCACAAAAGTATGAAGG
 S _ K _ G _ Q _ N _ C _ G _ Y _ R _ R _ C _ R _ A _ S _ G _ V _ L _ T _ T _ S _
 TGTGGTAACACTTTGACTTGTTACTTGAAGGCTACTGCTGCTTGTAGAGCTGCTAAATTG
 2821-----+-----+-----+-----+-----+-----+-----+
 ACACCATTGTGAAACTGAACAATGAAGTTCGGATGACGACGAACATCTCGACGATTTAAC
 C _ G _ N _ T _ L _ T _ C _ Y _ L _ K _ A _ T _ A _ A _ C _ R _ A _ A _ K _ L _
 CAGGACTGTACTATGTTGGTTAACGGTAACGACTTGGTTGTTATCTGTGAGTCCGCTGGT
 2881-----+-----+-----+-----+-----+-----+-----+
 GTCCTGACATGATACAACCAATTGCCATTGCTGAACCAACAATAGACACTCAGGCGACCA
 Q _ D _ C _ T _ M _ L _ V _ N _ G _ N _ D _ L _ V _ V _ I _ C _ E _ S _ A _ G _
 ACTCAAGAAGATGCTGCTTCTTTGAGAGTTTTACAGAGGCTATGACTAGATACTCTGCT

2941-----+-----+-----+-----+-----+
 TGAGTTCTTCTACGACGAAGAACTCTCAAAGTGTCTCCGATACTGATCTATGAGACGA
 T_Q_E_D_A_A_S_L_R_V_F_T_E_A_M_T_R_Y_S_A_

CCACCTGGTGATCCACCACAACCAGAATACGACTTGGAGTTGATCACTTCCTGTTCTCC
 3001-----+-----+-----+-----+-----+
 GGTGGACCACTAGGTGGTGTGGTCTTATGCTGAACCTCAACTAGTGAAGGACAAGGAGG
 P_P_G_D_P_P_Q_P_E_Y_D_L_E_L_I_T_S_C_S_S_

AATGTTGGTGTGCTCAGGATGCTTCCGAAAGAGAGTTTACTACTTGACTAGAGACCA
 3061-----+-----+-----+-----+-----+
 TTACAACCACAACGAGTGCTACGAAGGCCTTTCTCTCAAATGATGAACCTGATCTCTGGGT
 N_V_G_V_A_H_D_A_S_G_K_R_V_Y_Y_L_T_R_D_P_

ACTACTCCATTGGCTAGAGCTGCTTGGGAAACTGTTAGACACACTCCAGTTAACTCCTGG
 3121-----+-----+-----+-----+-----+
 TGATGAGGTAACCGATCTCGACGAACCCTTTGACAATCTGTGTGAGGTCAATTGAGGACC
 T_T_P_L_A_R_A_A_W_E_T_V_R_H_T_P_V_N_S_W_

TTGGGTAACATCATCATGTACGCTCCAACCTTTGTGGGCTAGAATGATCTTGATGACTCAC
 3181-----+-----+-----+-----+-----+
 AACCCATTGTAGTAGTACATGCGAGGTTGAAACACCCGATCTTACTAGAACTACTGAGTG
 L_G_N_I_I_M_Y_A_P_T_L_W_A_R_M_I_L_M_T_H_

TTCTTCTCCATCTTGTGGCTCAAGAGCAATTGGAAAAGGCTTTGGACTGTCAGATTTAC
 3241-----+-----+-----+-----+-----+
 AAGAAGAGGTAGAACAACCGAGTTCTCGTTAACCTTTTCCGAAACCTGACAGTCTAAATG
 F_F_S_I_L_L_A_Q_E_Q_L_E_K_A_L_D_C_Q_I_Y_

GGTGCTTGTTACTCCATTGAGCCATTGGACTTGCCACAGATCATTGAGAGATTGCACGGT
 3301-----+-----+-----+-----+-----+
 CCACGAACAATGAGGTAACCTCGGTAACCTGAACGGTGTCTAGTAACCTCTTAACGTGCCA
 G_A_C_Y_S_I_E_P_L_D_L_P_Q_I_I_E_R_L_H_G_

TTGTCTGCTTTCTCTTTGCACTCTTACTCCCCGGTGAAATCAACAGAGTTGCTTCCTGT
 3361-----+-----+-----+-----+-----+
 AACAGACGAAAGAGAAACGTGAGAATGAGGGGACCACTTTAGTTGTCTCAACGAAGGACA
 L_S_A_F_S_L_H_S_Y_S_P_G_E_I_N_R_V_A_S_C_

TTGAGAAAGTTGGGTGTTCCACCATTGAGAGTTTGGAGACACAGAGCTAGATCCGTTAGA
 3421-----+-----+-----+-----+-----+
 AACTCTTTCAACCCACAAGGTGGTAACTCTCAAACCTCTGTGTCTCGATCTAGGCAATCT
 L_R_K_L_G_V_P_P_L_R_V_W_R_H_R_A_R_S_V_R_

GCTAAGTTGTTGTCCCAAGGTGGAAGAGCTGCTACTTGTGGTAAGTACTTGTCAACTGG
 3481-----+-----+-----+-----+-----+
 CGATTCAACAACAGGGTCCACCTTCTCGACGATGAACACCATTGATGAACAAGTTGACC
 A_K_L_L_S_Q_G_G_R_A_A_T_C_G_K_Y_L_F_N_W_

GCTGTTAGAACAAGTTGAAGTTGACTCCTATTCTGCTGCTTCCCAATTGGATTGTCC
 3541-----+-----+-----+-----+-----+
 CGACAATCTTGTTCAACTTCAACTGAGGATAAGGACGACGAAGGGTTAACCTAAACAGG
 A_V_R_T_K_L_K_L_T_P_I_P_A_A_S_Q_L_D_L_S_

GGTGGTTTGTGCTGTTTACAACGGTGGTGACATCTACCACTCTTTGTCCAGAGCTAGA
 3601-----+-----+-----+-----+-----+
 CCAACCAAAACAACGACCAATGTTGCCACCACTGTAGATGGTGAGAAACAGGTCTCGATCT
 G_W_F_V_A_G_Y_N_G_G_D_I_Y_H_S_L_S_R_A_R_

CCAAGATAATAGACTAGTGGATCC
 3661-----+-----+-----+-----+-----+
 GGTCTATTATCTGATCACCTAGG
 P_R_*_*_

Proteína de fusão 4A-3-4B-5B
Otimizada para expressão em *Pichia pastoris*

```

1  ATGGGGTCCGTTGTTATCGTTGGTAGAATCATCTTGTCTGGTTCTGGTTCCGCTCCAATT
   -----+-----+-----+-----+-----+-----+-----+
61  TACCCCAGGCAACAATAGCAACCATCTTAGTAGAACAGACCAAGACCAAGGCGAGGTTAA
   M _ G _ S _ V _ V _ I _ V _ G _ R _ I _ I _ L _ S _ G _ S _ G _ S _ A _ P _ I _
   ACTGCTTACTCCCAGCAGACTAGAGGATTGTTGGGTTGTATCATCACTTCCTTGACTGGT
   -----+-----+-----+-----+-----+-----+-----+
121  TGACGAATGAGGGTCGTCTGATCTCCTAACCAACCAACATAGTAGTGAAGGAAGTACCA
   T _ A _ Y _ S _ Q _ Q _ T _ R _ G _ L _ L _ G _ C _ I _ I _ T _ S _ L _ T _ G _
   AGAGACAAGAACCAAGTTGACGGAGAGGTTGAGGTTTTGTCCACTGCTACTCAGTCTTTC
   -----+-----+-----+-----+-----+-----+-----+
181  TCTCTGTTCTTGGTTCAACTGCCTCTCCAAGTCCAAAACAGGTGACGATGAGTCAGAAAG
   R _ D _ K _ N _ Q _ V _ D _ G _ E _ V _ Q _ V _ L _ S _ T _ A _ T _ Q _ S _ F _
   TTGGCTACTTGTGTTAACGGTGTGTTGTTGGACTGTTTACGCTGGTGTGGTTCTAAAACT
   -----+-----+-----+-----+-----+-----+-----+
241  AACCGATGAACACAATTGCCACAAACAACCTGACAAATGCGACCACGACCAAGATTTTGA
   L _ A _ T _ C _ V _ N _ G _ V _ C _ W _ T _ V _ Y _ A _ G _ A _ G _ S _ K _ T _
   TTGGCTGGTCCAAAGGGTCCAATCACTCAGATGTACACAAACGTTGACCAGGATTTGGTT
   -----+-----+-----+-----+-----+-----+-----+
301  AACCGACCAGGTTTCCCAGGTTAGTGAGTCTACATGTGTTTGCAACTGGTCTAAACCAA
   L _ A _ G _ P _ K _ G _ P _ I _ T _ Q _ M _ Y _ T _ N _ V _ D _ Q _ D _ L _ V _
   GGTTGGCCAGCTCCACCAGGTGCTAGATCCATGACTCCATGTACTTGTGGTTCCCTCCGAC
   -----+-----+-----+-----+-----+-----+-----+
361  CCAACCGGTCGAGGTGGTCCACGATCTAGGTACTGAGGTACATGAACACCAAGGAGGCTG
   G _ W _ P _ A _ P _ P _ G _ A _ R _ S _ M _ T _ P _ C _ T _ C _ G _ S _ S _ D _
   TTGTACTTGGTTACTAGACACGCTGACGTTATCCCAGTTAGAAGAAGAGGAGACTCCAGA
   -----+-----+-----+-----+-----+-----+-----+
421  AACATGAACCAATGATCTGTGCGACTGCAATAGGGTCAATCTTCTCTCCTCTGAGGTCT
   L _ Y _ L _ V _ T _ R _ H _ A _ D _ V _ I _ P _ V _ R _ R _ R _ G _ D _ S _ R _
   GGATCTTTGTTGTCCCCAAGACCAGTTTCTTACTTGAAGGGATCTTCCGGTGGTCCATTG
   -----+-----+-----+-----+-----+-----+-----+
481  CCTAGAAACAACAGGGGTTCTGGTCAAAGAATGAACTTCCCTAGAAGGCCACCAGGTAAC
   G _ S _ L _ L _ S _ P _ R _ P _ V _ S _ Y _ L _ K _ G _ S _ S _ G _ G _ P _ L _
   TTGTGTCCATCCGGTCAAGTTGTTGGTATTTTCAGAGCTGCTGTTTGTACTAGAGGTGTT
   -----+-----+-----+-----+-----+-----+-----+
541  AACACAGGTAGGCCAGTGCAACAACCATAAAAGTCTCGACGACAAACATGATCTCCACAA
   L _ C _ P _ S _ G _ H _ V _ V _ G _ I _ F _ R _ A _ A _ V _ C _ T _ R _ G _ V _
   GCTAAGGCTGTTGACTTTCATCCCAGTTGAGTCCATGGAGACTACTATGAGATCCCCAGTT
   -----+-----+-----+-----+-----+-----+-----+
601  CGATTCGACAACCTGAAGTAGGGTCAACTCAGGTACCTCTGATGATACTCTAGGGGTCAA
   A _ K _ A _ V _ D _ F _ I _ P _ V _ E _ S _ M _ E _ T _ T _ M _ R _ S _ P _ V _
   TTTCACTGACAACCTCTTCCCCACCTGCTGTTCCACAACTTTCCAAGTTGCTCACTTGCAT
   -----+-----+-----+-----+-----+-----+-----+
661  AAGTGACTGTTGAGAAGGGGTGGACGACAAGGTGTTTGAAGGTTCAACGAGTGAACGTA
   F _ T _ D _ N _ S _ S _ P _ P _ A _ V _ P _ Q _ T _ F _ Q _ V _ A _ H _ L _ H _
   GCTCCAAGTGGTTCTGGTAAGTCCACTAAGGTTCCAGCTGCTTACGCTGCTCAAGGTTAC
   -----+-----+-----+-----+-----+-----+-----+
   CGAGGTTGACCAAGACCATTTCAGGTGATTCCAAGGTCGACGAATGCGACGAGTTCCAATG

```

A P T G S G K S T K V P A A Y A A Q G Y
 AAGGTTTTGGTTTTGAACCCATCCGTTGCTGCTACTTTGGGTTTCGGTGCTTACATGTCT
 721 -----+-----+-----+-----+-----+
 TTCCAAAACCAAACTTGGGTAGGCAACGACGATGAAACCCAAAGCCACGAATGTACAGA
 K V L V L N P S V A A T L G F G A Y M S
 AAGGCTCACGGTATTGAGCCAAACATCAGAACTGGTGTTAGAACTATCACTACTGGTGGT
 781 -----+-----+-----+-----+-----+
 TTCCGAGTGCCATAACTCGGTTTGTAGTCTTGACCACAATCTTGATAGTGATGACCACCA
 K A H G I E P N I R T G V R T I T T G G
 CCTATTACTTACTCCGCTTACGGAAAGTTTTTGGCTGACGGTGGTTGTTCTGGTGGTGGT
 841 -----+-----+-----+-----+-----+
 GGATAATGAATGAGGCGAATGCCTTTCAAAAACCGACTGCCACCAACAAGACCACCACGA
 P I T Y S A Y G K F L A D G G C S G G A
 TACGACATCATCATCTGTGACGAGTGTCACTCTACTGACTGGACTACTATCTTGGGTATC
 901 -----+-----+-----+-----+-----+
 ATGTGTAGTAGTAGACACTGCTCACAGTGAGATGACTGACCTGATGATAGAACCCATAG
 Y D I I I C D E C H S T D W T T I L G I
 GGTACTGTTTTGGACCAAGCTGAAACTGCTGGTGCTAGATTGGTTGTTTTGGCTACTGCT
 961 -----+-----+-----+-----+-----+
 CCATGACAAAACCTGGTTTCGACTTTGACGACCACGATCTAACCAACAAAACCGATGACGA
 G T V L D Q A E T A G A R L V V L A T A
 ACTCCACCAGGTTCCATTACTGTTCCACACCCAAACATCGAGGAAGTTGCTTTGTCTAAC
 1021 -----+-----+-----+-----+-----+
 TGAGGTGGTCCAAGGTAATGACAAGGTCTGGGTTGTAGCTCTTCAACGAAACAGATTG
 T P P G S I T V P H P N I E E V A L S N
 ACTGGAGAGATCCCATTCTACGGAAAGGCTATCCCAATTGAGGCTATCAAGGGTGGTAGA
 1081 -----+-----+-----+-----+-----+
 TGACCTCTCTAGGGTAAGATGCCTTTCCGATAGGGTTAACTCCGATAGTTCCCACCATCT
 T G E I P F Y G K A I P I E A I K G G R
 CACTTGATTTTTCTGTCACTCCAAGAAGAAGTGTGACGAGTTGGCTGCTAAGTTGACTGGA
 1141 -----+-----+-----+-----+-----+
 GTGAATAAAAGACAGTGAGGTTCTTCTTCACTGCTCAACCGACGATTCAACTGACCT
 H L I F C H S K K K C D E L A A K L T G
 TTGGGATTGAACGCTGTTGCTTACTACAGAGGATTGGACGTTTCCGTTATCCCAACTTCC
 1201 -----+-----+-----+-----+-----+
 AACCTAACTTGGGACAACGAATGATGTCTCTAACCTGCAAAGGCAATAGGGTTGAAGG
 L G L N A V A Y Y R G L D V S V I P T S
 GGTGATGTTGTTGTTGTTGCTACTGACGCTTTGATGACTGGTTTCACTGGTGACTTCGAC
 1261 -----+-----+-----+-----+-----+
 CCACTACAACAACAACAACGATGACTGCGAAACTACTGACCAAAAGTGACCACTGAAGCTG
 G D V V V V A T D A L M T G F T G D F D
 TCCGTTATCGACTGTAACACTTGTGTTACTCAGACTGTTGACTTCTCCTTGGACCCAACT
 1321 -----+-----+-----+-----+-----+
 AGGCAATAGCTGACATTGTGAACACAATGAGTCTGACAACCTGAAGAGGAACCTGGGTTGA
 S V I D C N T C V T Q T V D F S L D P T
 TTCCTATCGAGACTACTACTGTTCTCAAGACGCTGTTTCCAGATCCCAAAGAAGAGGT
 1381 -----+-----+-----+-----+-----+
 AAGTGATAGCTCTGATGATGACAAGGAGTTCTGCGACAAAGGTCTAGGGTTTCTTCTCCA
 F T I E T T T V P Q D A V S R S Q R R G
 GCTACTGGTAGAGGAAGATCCGGTATCTACAGATTTCGTTACTCCAGGTGAAAGACCATCT

1441-----+-----+-----+-----+-----+
 CGATGACCATCTCCTTCTAGGCCATAGATGTCTAAGCAATGAGGTCCACTTTCTGGTAGA
 A T G R G R S G I Y R F V T P G E R P S

 GGAATGTTGACTCCTCCGTTTTGTGTGAATGTTACGACGCTGGTTGTGCTTGGTACGAA
 1501-----+-----+-----+-----+-----+
 CCTTACAAGCTGAGGAGGCAAAACACACTTACAATGCTGCGACCAACACGAACCATGCTT
 G M F D S S V L C E C Y D A G C A W Y E

 TTGACTCCAGCTGAGACTACTGTTAGATTGAGAGCTTACTTGAACACTCCAGGATTGCCA
 1561-----+-----+-----+-----+-----+
 AACTGAGGTGCGACTCTGATGACAATCTAACTCTCGAATGAACTTGTGAGGTCTAACGGT
 L T P A E T T V R L R A Y L N T P G L P

 GTTTGTCAAGACCACTTGGGAATTCTGGGAGTCCGTTTTCACTGGATTGACTCACATTGAC
 1621-----+-----+-----+-----+-----+
 CAAACAGTTCTGGTGAACCTTAAGACCCTCAGGCAAAAGTGACCTAACTGAGTGTAACTG
 V C Q D H L E F W E S V F T G L T H I D

 GCTCACTTTTTGTCCCAAACCTAAGCAGGCTGGTGACAACTTCCATACTTGGTTGCTTAC
 1681-----+-----+-----+-----+-----+
 CGAGTGAAAAACAGGGTTTGATTTCGTCCGACCACTGTTGAAAGGTATGAACCAACGAATG
 A H F L S Q T K Q A G D N F P Y L V A Y

 CAGGCTACTGTTTGTGCTAGAGCACAAAGCTCCACCACCATCTTGGGATCAGATGTGGAAG
 1741-----+-----+-----+-----+-----+
 GTCCGATGACAAACACGATCTCGTGTTCGAGGTGGTGGTAGAACCTAGTCTACACCTTC
 Q A T V C A R A Q A P P P S W D Q M W K

 TGTTTGATCAGATTGAAGCCAACTTTGCACGGTCCAACCTCATTGTTGTACAGATTGGGT
 1801-----+-----+-----+-----+-----+
 ACAAACCTAGTCTAACTTCGGTTGAAACGTGCCAGGTTGAGGTAACAACATGTCTAACCCA
 C L I R L K P T L H G P T P L L Y R L G

 GCTGTTCAGAACGAGATCACTTTGACTCACCCAATCACTAAGTTTCGTTATGGCTTGCATG
 1861-----+-----+-----+-----+-----+
 CGACAAGTCTTGCTCTAGTGAAACTGAGTGGGTTAGTGATTCAAGCAATACCGAACGTAC
 A V Q N E I T L T H P I T K F V M A C M

 TCTGCTGACTTGGAAAGTTGTTTCCTTGATGGCTTTCACTGCTTCCATTACTTCCCCATTG
 1921-----+-----+-----+-----+-----+
 AGACGACTGAACCTTCAACAAAGGAACCTACCGAAAGTGACGAAGGTAATGAAGGGGTAAC
 S A D L E V V S L M A F T A S I T S P L

 ACTACTCAGAACACTTTGTTGTTCAACATCTTGGGAGGATGGGTTGCAGCTCAATTGTCC
 1981-----+-----+-----+-----+-----+
 TGATGAGTCTTGTGAAACAACAAGTTGTAGAACCCTCCTACCCAACGTCGAGTTAACAGG
 T T Q N T L L F N I L G G W V A A Q L S

 ATGTCCTACACTTGGACTGGTGCTTTGATCACTCCATGTGCTGCTGAAGAATCCAAGTTG
 2041-----+-----+-----+-----+-----+
 TACAGGATGTGAACCTGACCACGAAACTAGTGAGGTACACGACGACTTCTTAGGTTCAAC
 M S Y T W T G A L I T P C A A E E S K L

 CCAATCAACCCATTGTCCAACCTCTTTGTTGAGACACCACTCCATGGTTTACTCCACTACT
 2101-----+-----+-----+-----+-----+
 GGTTAGTTGGGTAACAGGTTGAGAAACAACCTCTGTGGTGAGGTACCAATGAGGTGATGA
 P I N P L S N S L L R H H S M V Y S T T

 TCCAGATCCGCTTCCTTGAGACAGAAGAAGGTTACATTGACAGATTGCAGGTTTTGGAC
 2161-----+-----+-----+-----+-----+
 AGGTCTAGGCGAAGGAACCTCTGTCTTCTTCCAATGTAAGCTGTCTAACGTCCAAAACCTG
 S R S A S L R Q K K V T F D R L Q V L D

GACCACTACAGAGATGTTTTGAAGGAGATGAAGGCTAAGGCTTCCACTGTTAAGGCTAGA
 2221-----+-----+-----+-----+-----+-----+-----+
 CTGGTGATGTCTCTACAAAACCTTCCTCTACTTCCGATTCCGAAGGTGACAATTCGGATCT
 D H Y R D V L K E M K A K A S T V K A R
 TTGTTGTCCATTGAGGAGGCTTGTAAGTTGACTCCACCACACTCTGCTAAGTCCAAGTTT
 2281-----+-----+-----+-----+-----+-----+-----+
 APCACAGGTAACCTCCTCCGAACATTCAACTGAGGTGGTGTGAGACGATTACAGGTTCAAA
 L L S I E E A C K L P P H S A K S K F
 GGTTACGGTGCTAAGGATGTTAGATCCTTGTCCTCCAGAGCTGTAAACCACATCAGGTCT
 2341-----+-----+-----+-----+-----+-----+-----+
 CCAATGCCACGATTCCCTACAATCTAGGAACAGGAGGTCTCGACAATTGGTGTAGTCCAGA
 G Y G A K D V R S L S S R A V N H I R S
 GTTTGGGAGGATTTGTTGGAGGACACTGAGACTCCAACTCGACACTACTATCATGGCTAAG
 2401-----+-----+-----+-----+-----+-----+-----+
 CAAACCTCCTAAACAACCTCCTGTGACTCTGAGGTAGCTGTGATGATAGTACCGATTCT
 V W E D L L E D T E T P I D T T I M A K
 AACGAGGTTTTCTGTGTTCAACCAGAGAAGGGTGAAGAAAGCCAGCTAGATTGATCGTT
 2461-----+-----+-----+-----+-----+-----+-----+
 TTGCTCCAAAAGACACAAGTTGGTCTCTTCCCACCTTCTTTTCGGTCGATCTAACTAGCAA
 N E V F C V Q P E K G G R K P A R L I V
 TTTCCAGACTTGGGTGTTAGAGTTTGTGAGAAGATGGCTTTGTACGACGTTGTTTCCACT
 2521-----+-----+-----+-----+-----+-----+-----+
 AAGGGTCTGAACCCACAATCTCAAACACTCTTCTACCGAAACATGCTGCAACAAGGTGA
 F P D L G V R V C E K M A L Y D V V S T
 TTGCCACAGGCTGTTATGGGACCATCTTACGGTTTCCAATACTCCCCAGGACAAAGAGTT
 2581-----+-----+-----+-----+-----+-----+-----+
 AACGGTGTCCGACAATACCCCTGGTAGAATGCCAAAGGTTATGAGGGGTCTCTTTCTCAA
 L P Q A V M G P S Y G F Q Y S P G Q R V
 GAGTTCTTGGTTAACACTTGAAGTCCAAGAAATGTCCAATGGGATTCTCCTACAACACT
 2641-----+-----+-----+-----+-----+-----+-----+
 CTCAAGAACCAATTGTGAACCTTCAGGTTCTTTACAGGTTACCCCTAAGAGGATGTTGTGA
 E F L V N T W K S K K C P M G F S Y N T
 AGATGTTTCGACTCCACTGTTACTGAGAACGACATCAGAACAGAGGAGTCCATCTACCAG
 2701-----+-----+-----+-----+-----+-----+-----+
 TCTACAAAGCTGAGGTGACAATGACTCTTGCTGTAGTCTTGCTCCTCAGGTAGATGGTC
 R C F D S T V T E N D I R T E E S I Y Q
 TGTTGTGACTTGGCTCCAGAAGCTAGACAAGCTATCAAGTCCCTTGACTGAGAGATTGTAC
 2761-----+-----+-----+-----+-----+-----+-----+
 ACAACACTGAACCGAGGTCTTCGATCTGTTTCGATAGTTCAGGAACCTGACTCTCTAACATG
 C C D L A P E A R Q A I K S L T E R L Y
 ATCGGTGGTCTTTTACTAATCCAAGGGACGAAGTGTGGTTACAGAAGATGTAGAGCT
 2821-----+-----+-----+-----+-----+-----+-----+
 TAGCCACCAGGAAACTGATTGAGGTTCCCTGTCTTGACACCAATGTCTTCTACATCTCGA
 I G G P L T N S K G Q N C G Y R R C R A
 TCCGGTGTTTTACTACTTCTGTGGTAACACTTTGACTTGTTACTTGAAGGCTACTGCT
 2881-----+-----+-----+-----+-----+-----+-----+
 AGGCCACAAAACCTGATGAAGGACACCATTTGTGAACTGAACAATGAAGTCCGATGACGA
 S G V L T T S C G N T L T C Y L K A T A
 GCTTGTAGAGCTGCTAAATTGCAGGACTGTACTATGTTGGTTAACGGTAACGACTTGGTT
 2941-----+-----+-----+-----+-----+-----+-----+

CGAACATCTCGACGATTTAACGTCTGACATGATACAACCAATTGCCATTGCTGAACCAA
 A C R A A K L Q D C T M L V N G N D L V
 GTTATCTGTGAGTCCGCTGGTACTCAAGAAGATGCTGCTTCTTTGAGAGTTTTTCACAGAG
 3001-----+-----+-----+-----+-----+
 CAATAGACACTCAGGCGACCATGAGTTCTTCTACGACGAAGAACTCTCAAAAGTGTCTC
 V I C E S A G T Q E D A A S L R V F T E
 GCTATGACTAGATACTCTGCTCCACCTGGTGATCCACCACAACCAGAATACGACTTGGAG
 3061-----+-----+-----+-----+-----+
 CGATACTGATCTATGAGACGAGGTGGACCCTAGGTGGTGGTGGTCTTATGCTGAACCTC
 A M T R Y S A P P G D P P Q P E Y D L E
 TTGATCACTTCTCTGTTCTCCTCAATGTTGGTGGTGGTCTCACGATGCTTCCGGAAAGAGAGTT
 3121-----+-----+-----+-----+-----+
 AACTAGTGAAGGACAAGGAGGTTACAACCACAACGAGTGCTACGAAGGCCCTTTCTCTCAA
 L I T S C S S N V G V A H D A S G K R V
 TACTACTTGACTAGAGACCCAACTACTCCATTGGCTAGAGCTGCTTGGGAAACTGTTAGA
 3181-----+-----+-----+-----+-----+
 ATGATGAAGTCTCTCTGGGTTGATGAGGTAACCGATCTCGACGAACCCCTTTGACAATCT
 Y Y L T R D P T T P L A R A A W E T V R
 CACACTCCAGTTAACTCCTGGTTGGGTAACATCATCATGTACGCTCCAACCTTTGTGGGCT
 3241-----+-----+-----+-----+-----+
 GTGTGAGGTCAATTGAGGACCAACCCATTGTAGTAGTACATGCGAGGTTGAAACACCCGA
 H T P V N S W L G N I I M Y A P T L W A
 AGAATGATCTTGATGACTCACTTCTTCTCCATCTTGTGGCTCAAGAGCAATTGGAAAAG
 3301-----+-----+-----+-----+-----+
 TCTTACTAGAATACTGAGTGAAGAAGAGGTAGAACAACCGAGTTCTCGTTAACCTTTTC
 R M I L M T H F F S I L L A Q E Q L E K
 GCTTTGGACTGTCAGATTTACGGTGCTTGTACTCCATTGAGCCATTGGACTTGGCCACAG
 3361-----+-----+-----+-----+-----+
 CGAAACCTGACAGTCTAAATGCCACGAACAATGAGGTAACCTCGGTAACCTGAACGGTGTG
 A L D C Q I Y G A C Y S I E P L D L P Q
 ATCATTGAGAGATTGCACGGTTTGTCTGCTTTCTCTTGGCACTCTTACTCCCCTGGTGAA
 3421-----+-----+-----+-----+-----+
 TAGTAACCTCTCTAACGTGCCAAACAGACGAAAGAGAAACGTGAGAATGAGGGGACCACTT
 I I E R L H G L S A F S L H S Y S P G E
 ATCAACAGAGTTGCTTCTGTTTGGAGAAAGTTGGGTGTTCCACCATTGAGAGTTTGGAGA
 3481-----+-----+-----+-----+-----+
 TAGTTGTCTCAACGAAGGACAACTCTTTCAACCCACAAGGTGGTAACTCTCAAACCTCT
 I N R V A S C L R K L G V P P L R V W R
 CACAGAGCTAGATCCGTTAGAGCTAAGTTGTTGTCCCAAGGTGGAAGAGCTGCTACTTGT
 3541-----+-----+-----+-----+-----+
 GTGTCTCGATCTAGGCAATCTCGATTCAACAACAGGGTCCACCTTCTCGACGATGAACA
 H R A R S V R A K L L S Q G G R A A T C
 GGTAAGTACTTGTTCAACTGGGCTGTTAGAACAAAGTTGAAGTTGACTCCTATTCTGCT
 3601-----+-----+-----+-----+-----+
 CCATTCATGAACAAGTTGACCCGACAATCTTGTTCAACTTCAACTGAGGATAAGGACGA
 G K Y L F N W A V R T K L K L T P I P A
 GCTTCCCAATTGGATTTGTCCGGTTGGTTTGTGCTGGTTACAACGGTGGTGACATCTAC
 3661-----+-----+-----+-----+-----+
 CGAAGGGTTAACCTAAACAGGCCAACCAACAACGACCAATGTTGCCACCACTGTAGATG
 A S Q L D L S G W F V A G Y N G G D I Y

27/45

CACTCTTTGTCCAGAGCTAGACCAAGATAATAGACTAGTGGATCC
3721-----+-----+-----+-----+-----
GTGAGAAACAGGTCTCGATCTGGTTCTATTATCTGATCACCTAGG
H _ S _ L _ S _ R _ A _ R _ F _ R _ * _ * _

Proteína de fusão 4A-3-5B
Otimizada para Escherichia coli

```

1  AGGGCGAATTGGGTACCATGGGGAGCGTTGTTATTGTTGGCCGCATTATTCTGAGCGGTA
   -----+-----+-----+-----+-----+-----+-----+
   TCCCGCTTAACCCATGGTACCCCTCGCAACAATAACAACCGGCGTAATAAGACTCGCCAT
       M _ G _ S _ V _ V _ I _ V _ G _ R _ I _ I _ L _ S _ G _ S

61  GCGGTAGCGCCCCGATTACCGCCTATTCTCAGCAAACCCGTGGTCTGCTGGGTTGTATTA
   -----+-----+-----+-----+-----+-----+-----+
   CGCCATCGCGGGGCTAATGGCGGATAAGAGTCGTTTGGGCACCAGACGACCCAACATAAT
       _ G _ S _ A _ P _ I _ T _ A _ Y _ S _ Q _ Q _ T _ R _ G _ L _ L _ G _ C _ I _ I

121  TTACCAGCCTGACCGGCCGTGATAAAATCAGGTGGATGGCGAAGTTCAGGTTCTGAGCA
   -----+-----+-----+-----+-----+-----+-----+
   AATGGTCGGACTGGCCGGCACTATTTTGTAGTCCACCTACCGCTTCAAGTCCAAGACTCGT
       _ T _ S _ L _ T _ G _ R _ D _ K _ N _ Q _ V _ D _ G _ E _ V _ Q _ V _ L _ S _ T

181  CCGCCACCCAGAGCTTTCTGGCGACCTGTGTGAATGGTGTGTGCTGGACCGTTTATGCCG
   -----+-----+-----+-----+-----+-----+-----+
   GGCGGTGGGTCTCGAAAGACCGCTGGACACACTTACCACACACGACCTGGCAAATACGGC
       _ A _ T _ Q _ S _ F _ L _ A _ T _ C _ V _ N _ G _ V _ C _ W _ T _ V _ Y _ A _ G

241  GTGCCGGTAGCAAAACCCCTGGCGGGTCCGAAAGGTCCGATTACCCAGATGTACACCAACG
   -----+-----+-----+-----+-----+-----+-----+
   CACGGCCATCGTTTTGGGACCGCCAGGCTTTCCAGGCTAATGGGTCTACATGTGGTTGC
       _ A _ G _ S _ K _ T _ L _ A _ G _ P _ K _ G _ P _ I _ T _ Q _ M _ Y _ T _ N _ V

301  TGGATCAGGATCTGGTTGGTTGGCCGGCGCCGCGGGTGCCCGTAGCATGACCCCGTGTA
   -----+-----+-----+-----+-----+-----+-----+
   ACCTAGTCCTAGACCAACCAACCGGCCGCGGGCGGCCACGGGCATCGTACTGGGGCACAT
       _ D _ Q _ D _ L _ V _ G _ W _ P _ A _ P _ P _ G _ A _ R _ S _ M _ T _ P _ C _ T

361  CCTGTGGTAGCAGCGATCTGTATCTGGTTACCCGTCATGCCGATGTTATTCCGGTTTCGTC
   -----+-----+-----+-----+-----+-----+-----+
   GGACACCATCGTCGCTAGACATAGACCAATGGGCAGTACGGCTACAATAAGGCCAAGCAG
       _ C _ G _ S _ S _ D _ L _ Y _ L _ V _ T _ R _ H _ A _ D _ V _ I _ P _ V _ R _ R

421  TCGTGGTGATAGCCGTGGTAGCCTGCTGTCTCCGCGTCCGGTTAGCTATCTGAAAGGTA
   -----+-----+-----+-----+-----+-----+-----+
   CAGCACCCTATCGGCACCATCGGACGACAGAGGCGCAGGCCAATCGATAGACTTTCCAT
       _ R _ G _ D _ S _ R _ G _ S _ L _ L _ S _ P _ R _ P _ V _ S _ Y _ L _ K _ G _ S

481  GCAGCGGTGGTCCGCTGCTGTGTCCGAGCGGTCATGTGGTGGGTATTTTTCGTGCCGCCG
   -----+-----+-----+-----+-----+-----+-----+
   CGTCGCCACCAGGCGACGACACAGGCTCGCCAGTACACCACCCATAAAAAGCACGGCGGC
       _ S _ G _ G _ P _ L _ L _ C _ P _ S _ G _ H _ V _ V _ G _ I _ F _ R _ A _ A _ V

541  TTTGTACCCGTGGTGTGGCGAAAGCGGTGGATTTTATCCCGGTTGAAAGCATGGAACCA
   -----+-----+-----+-----+-----+-----+-----+
   AAAATGGGCACCACACCGCTTTTCGCCACCTAAAATAGGGCCAACCTTTCGTACCTTTGGT
       _ C _ T _ R _ G _ V _ A _ K _ A _ V _ D _ F _ I _ P _ V _ E _ S _ M _ E _ T _ T

601  CCATGCGTAGCCCGGTGTTTACCGATAATAGCAGCCCGCCGCGGTTCCGCAGACCTTTC
   -----+-----+-----+-----+-----+-----+-----+
   GGTACGCATCGGGCCACAAATGGCTATTATCGTCGGGCGCCGCCAAGGCGTCTGGAAAG
       _ M _ R _ S _ P _ V _ F _ T _ D _ N _ S _ S _ P _ P _ A _ V _ P _ Q _ T _ F _ Q

661  AGGTTGCCCATCTGCATGCGCCGACCGGTAGCGGTAAAAGCACCAAAGTTCCGGCGGCGT
   -----+-----+-----+-----+-----+-----+-----+
   TCCAACGGGTAGACGTACGCGGCTGGCCATCGCCATTTTCGTGGTTTCAAGGCCGCCGCA

```

V A H L H A P T G S G K S T K V P A A Y
 ATGCCGCCCAGGGTTATAAAGTGCTGGTGCTGAATCCGAGCGTGGCGGCGACCTGGGTT
 721 -----+-----+-----+-----+-----+-----+-----+
 TACGGCGGGTCCCAATATTTACGACCACGACTTAGGCTCGCACCGCCGCTGGGACCCAA
A A Q G Y K V L V L N P S V A A T L G F
 TTGGTGCCTATATGAGCAAAGCCCATGGCATTGAACCGAACATTTCGTACCGGCGTTTCGTA
 781 -----+-----+-----+-----+-----+-----+-----+
 AACCACGGATATACTCGTTTCGGGTACCGTAACCTGGCTTGTAAAGCATGGCCGCAAGCAT
G A Y M S K A H G I E P N I R T G V R T
 CCATTACCACCGGTGGCCCCGATTACCTATAGCGCCTACGGCAAATTTCTGGCGGATGGTG
 841 -----+-----+-----+-----+-----+-----+-----+
 GGTAATGGTGGCCACCGGGCTAATGGATATCGCGGATGCCGTTTAAAGACCGCCTACCAC
I T T G G P I T Y S A Y G K F L A D G G
 GCTGTAGCGGTGGCGCCTATGATATCATCATCTGTGATGAATGCCATAGCACCGATTGGA
 901 -----+-----+-----+-----+-----+-----+-----+
 CGACATCGCCACCGCGGATACTATAGTAGTAGACACTACTTACGGTATCGTGGCTAACCT
C S G G A Y D I I I C D E C H S T D W T
 CCACCATTCTGGGTATTGGCACCGTTCGGATCAGGCGGAAACCGCCGGTGGCCGTCCTGG
 961 -----+-----+-----+-----+-----+-----+-----+
 GGTGGTAAGACCCATAACCGTGGCAAGACCTAGTCCGCCTTTGGCGGCCACGGGCAGACC
T I L G I G T V L D Q A E T A G A R L V
 TTGTTCTGGCGACCGCAACGCCGCGCGGGTAGCATTACCGTTCCGCATCCGAACATTGAAG
 1021 -----+-----+-----+-----+-----+-----+-----+
 AACAAGACCGCTGGCGTTGCGGCGGCCCATCGTAATGGCAAGGCGTAGGCTTGTAACCTC
V L A T A T P P G S I T V P H P N I E E
 AAGTGGCCCTGAGCAATACCGGCGAAATTCGGTTTTATGGCAAAGCGATTCCGATCGAAG
 1081 -----+-----+-----+-----+-----+-----+-----+
 TTCACCGGGACTCGTTATGGCCGCTTTAAGGCAAAATACCGTTTCGCTAAGGCTAGCTTC
V A L S N T G E I P F Y G K A I P I E A
 CGATTAAAGGCGGCCGTCATCTGATTTTTTTGCCACAGCAAAAAAAATGTGATGAACTGG
 1141 -----+-----+-----+-----+-----+-----+-----+
 GCTAATTTCCGCGCGCAGTAGACTAAAAACGGTGTCGTTTTTTTTTACACTACTTGACC
I K G G R H L I F C H S K K K C D E L A
 CGGCGAAACTGACCGGTCTGGGTCTGAATGCCGTGGCGTATTATCGTGGTCTGGATGTGA
 1201 -----+-----+-----+-----+-----+-----+-----+
 GCCGCTTTGACTGGCCAGACCCAGACTTACGGCACCGCATAATAGCACCGACCTACACT
A K L T G L G L N A V A Y Y R G L D V S
 GCGTTATTCGACCAGCGGTGATGTTGTTCTGGTGGCGACCGATGCCCTGATGACCGGTT
 1261 -----+-----+-----+-----+-----+-----+-----+
 CGCAATAAGGCTGGTCGCCACTACAACAACACCACCGCTGGCTACGGGACTACTGGCCAA
V I P T S G D V V V V A T D A L M T G F
 TTACCGGCGATTTTGATAGCGTGATCGATTGTAACACCTGTGTGACCCAGACCGTTGATT
 1321 -----+-----+-----+-----+-----+-----+-----+
 AATGGCCGCTAAAACTATCGCACTAGCTAACATTGTGGACACACTGGGTCTGGCAACTAA
T G D F D S V I D C N T C V T Q T V D F
 TTAGCCTGGACCCGACCTTTACCATTTGAAACCAACACCGTTCCGCAGGATGCCGTTAGCC
 1381 -----+-----+-----+-----+-----+-----+-----+
 AATCGGACCTGGGCTGGAAATGGTAACCTTTGGTGGTGGCAAGGCGTCCTACGGCAATCGG
S L D P T F T I E T T T V P Q D A V S R
 GTAGCCAGCGTCGTGGTGCCACCGGTCGTGGTCGTAGCGGCATTTATCGTTTTGTGACGC

1441-----+-----+-----+-----+-----+
 CATCGGTGCGAGCACCACGGTGGCCAGCACCAGCATCGCCGTAAATAGCAAAACACTGCG
S Q R R G A T G R G R S G I Y R F V T P
 CGGGTGAACGTCCGAGCGGTATGTTTGATAGCAGCGTGCTGTGTGAATGTTATGATGCCG
 1501-----+-----+-----+-----+-----+
 GCCCACTTGCAGGCTCGCCATACAAACTATCGTCGCACGACACACTTACAATACTACGGC
G E R P S G M F D S S V L C E C Y D A G
 GCTGTGCCTGGTATGAACTGACCCCGGCGGAAACCACCGTTCGTCTGCGCGCGTATCTGA
 1561-----+-----+-----+-----+-----+
 CGACACGGACCATACTTGACTGGGGCCGCCTTTGGTGGCAAGCAGACGCGCGCATAGACT
C A W Y E L T P A E T T V R L R A Y L N
 ATACGCCGGGTCTGCCGGTTTGTCTAGGATCATCTGGAATTCTGGGAAAGCGTTTTTACCG
 1621-----+-----+-----+-----+-----+
 TATGCGGCCCAGACGGCCAAACAGTCTTAGTAGACCTTAAGACCCCTTCGCAAAAATGGC
T P G L P V C Q D H L E F W E S V F T G
 GCCTGACCCATATTGATGCCCATTTTCTGAGCCAGACCAAACAGSCGGGTGATAACTTTC
 1681-----+-----+-----+-----+-----+
 CGGACTGGGTATAACTACGGGTAAAAGACTCGGTCTGGTTTGTCCGCCCACTATTGAAG
L T H I D A H F L S Q T K Q A S D H F P
 CGTATCTGGTGGCGTATCAGGCGACCGTTTGTGCCCCGTGCCAGGCGCGCGCGCGAGCT
 1741-----+-----+-----+-----+-----+
 GCATAGACCACCGCATAGTCCGCTGGCAACACGGGCACGGGTCCGCGGCGGCGGCTCGA
Y L V A Y Q A T V C A R A Q A P P P S W
 GGGATCAGATGTGGAAATGCCTGATTCTGTCTGAAACCGACCCCTGCATGGTCCGACCCCGC
 1801-----+-----+-----+-----+-----+
 CCTAGTCTACACCTTTACGGACTAAGCAGACTTTGGCTGGGACGTACCAGGCTGGCGCG
D Q M W K C L I R L K P T L H G P T P L
 TGCTGTATCGTCTGGGTGCCGTTTCTGAAACGAATTACCCCTGACCCATCCGATCACCAAAT
 1861-----+-----+-----+-----+-----+
 ACGACATAGCAGACCCACGGCAAGTCTTGGCTTTAATGGGACTGGGTAGGCTAGTGSTTTA
L Y R L G A V Q N E I T L T H F I T K F
 TTGTGATGGCGTGTATGAGCGCCGATCTGGAAGTTCTTGTAGCGGTAGCGGCTCTATGA
 1921-----+-----+-----+-----+-----+
 AACACTACCGCACATACTCGCGGCTAGACCTTCAACAACCATCGCCATCGCGGAGATACT
V M A C M S A D L E V V G S G S G S M S
 GCTATACCTGGACCGGTGCCCTGATTACCCCGTGTGCCGCCGAAGAAAGCAAACTGCCGA
 1981-----+-----+-----+-----+-----+
 CGATATGGACCTGGCCACGGGACTAATGGGGCACACGGCGGCTTCTTTCTTTGACGGCT
Y T W T G A L I T P C A A E E S K L P I
 TTAACCCGCTGTCTAATAGCCTGCTGCGTCATCATAGCATCGTGTATAGCACCACCAGCC
 2041-----+-----+-----+-----+-----+
 AATTGGGCGACAGATTATCGGACGACGAGTAGTATCGTACCACATATCGTGGTGGTCCG
N P L S N S L L R H H S M V Y S T T S R
 GTAGCGCCAGCCTGCGTCAGAAAAAAGTGACCTTCGATCGTCTGCAGGTGCTGGATGATC
 2101-----+-----+-----+-----+-----+
 CATCGCGGTGCGACGCACTTTTTTCACTGGAAGCTAGCAGACGTCCACGACCTACTAG
S A S L R Q K K V T F D R L Q V L D D H
 ATTATCGTGATGTGCTGAAAGAAATGAAAGCGAAACCGAGCACCGTTAAAGCCCGTCTGC
 2161-----+-----+-----+-----+-----+
 TAATAGCACTACACGACTTTCTTTACTTTTCGCTTTCGCTCGTGGCAATTTCCGGCAGACG

Y R D V L K E M K A K A S T V K A R L L
 TGTCTATTGAAGAAGCGTGTAAGTACGACCCGCGCATAGCGCCAAAAGCAAATTTGGCT
 2221-----+-----+-----+-----+-----+-----+-----+
 ACAGATAACTTCTTCGCACATTTGACTGGGGCGGCGTATCGCGGTTTTCGTTTTAAACCGA
S I E E A C K L T P P H S A K S K F G Y
 ATGGCGCCAAAGATGTTTCGTAGCCTGAGCAGCCGTGCCGTTAATCATATTCGTAGCGTGT
 2281-----+-----+-----+-----+-----+-----+-----+
 TACCGCGGTTTTCTACAAGCATCGGACTCGTCGGCAGGCAATTAGTATAAGCATCGCACA
G A K D V R S L S S R A V N H I R S V W
 GGAAGATCTGCTGGAAGATACCGAAACCCCGATTGATACCACCATCATGGCGAAAAACG
 2341-----+-----+-----+-----+-----+-----+-----+
 CCCTTCTAGACGACCTTCTATGGCTTTGGGGCTAACTATGGTGGTAGTACCGCTTTTTTGC
E D L L E D T E T P I D T T I M A K N E
 AAGTGTTTTGTGTTTCAGCCGAAAAAGGTGGTTCGTAAACCGGCCCGTCTGATTGTTTTTC
 2401-----+-----+-----+-----+-----+-----+-----+
 TTCACAAAACACAAGTCGGCCTTTTTCCACCAGCATTTGGCCGGGCAGACTAACAAAAAG
V F C V Q P E K G G R K P A R L I V F P
 CGGATCTGGGTGTTTCGTGTGTGTGAAAAATGGCGCTGTACGATGTTGTTAGCACCCCTGC
 2461-----+-----+-----+-----+-----+-----+-----+
 GCCTAGACCCACAAGCACACACACTTTTTTACCGCGACATGCTACAACAATCGTGGGACG
D L G V R V C E K M A L Y D V V S T L P
 CGCAGGCGGTTATGGGTCCGAGCTATCCCTTCCTATTCTCCGGGTCAGCGTGTGAAT
 2521-----+-----+-----+-----+-----+-----+-----+
 GCGTCCGCCAATACCCAGGCTCGATACCGAAAGTCATAAGAGGCCAGTCGCACAACCTTA
Q A V M G P S Y G F Q Y S P Q Q P V E F
 TTCTGGTGAACACCTGGAAAAAGCAAAATGCCCGATGGGCTTCAGCTATAACACCCGCT
 2581-----+-----+-----+-----+-----+-----+-----+
 AAGACCACTTGTGGACCTTTTCGTTTTTTACGGGCTACCCGAAGTCGATATTGTGGGCGA
L V N T W K S K K C P M G F S Y I T R C
 GCTTTGATAGCACCGTGACCGAAAACGATATCCGTACCGAAGAAAGCATTACCAGTGCT
 2641-----+-----+-----+-----+-----+-----+-----+
 CGAAACTATCGTGGCAGTGGCTTTTGCATAGGCATGGCTTCTTTTCGTAAATGGTCACGA
F D S T V T E N D I R T E E S I Y Q C C
 GTGATCTGGCGCCGGAAGCCCGTCAGGCGATTAAAAGCCTGACCGAACGCCCTGTATATTG
 2701-----+-----+-----+-----+-----+-----+-----+
 CACTAGACCGCGGCCCTTCGGGCAGTCCGCTAATTTTCGGACTGGCTTGCGGACATATAAC
D L A P E A R Q A I K S L D E R L Y I G
 GCGGTCCGCTGACCAATAGCAAGGCCAGAACTGTGGTTATCGTCGTTGTCTGCCAGCG
 2761-----+-----+-----+-----+-----+-----+-----+
 CGCCAGGCGACTGGTTATCGTTTTCCGGTCTTGACACCAATAGCGCAACAGCACGGTCCG
G P L T M S K Q C G V R C R A S G
 GTGTTCTGACCACCGCTGTGGTAATACCCCTGACCTGCTACCTGAAAGCGACCGCCGCT
 2821-----+-----+-----+-----+-----+-----+-----+
 CACAAGACTGGTGGTCGACACCATTTATGGGACTGGACGATGGACTTTTCGCTGGCGCGGA
V L T T E C G I T L T C Y L K A T A A C
 GTCGTGCCGCCAAACTGCAGGATTGTACCATGCTGGTTAACGGCAATGATCTGGTGGTGA
 2881-----+-----+-----+-----+-----+-----+-----+
 CAGCACGGCGGTTTGACGTCCTAACATGGTACGACCAATTGCCGTTACTAGACCACCACT
R A A K L Q D C T M L V N G N D L V V I
 TTTGTGAAAGCGCCGCGCCAGCAAGATGCCGCGAGCCTGCGCGTTTTTACCGAAGCGA

2941-----+-----+-----+-----+-----+
 AAACACTTTTCGCGCCCGTGGGTCTTCTACGGCGGTTCGGACGCGCAAAATGGCTTCGCT
C E S A G T Q E D A A S L R V F T E A M
 TGACCCGTTATAGCGCCCCGCGGGTGATCCGCGCAGCCGGAATATGATCTGGAAGTGA
 3001-----+-----+-----+-----+-----+
 ACTGGGCAATATCGCGGGGCGGCCACTAGGCGGCGTCGGCCTTATACTAGACCTTGA
T R Y S A P P G D P P Q P E Y D L E L I
 TCACCAGCTGTAGCAGCAATGTTGGTGTGGCCATGATGCCAGCGGTAAACGTGTGTATT
 3061-----+-----+-----+-----+-----+
 AGTGGTCGACATCGTCGTTACAACCACAACGGGTACTACGGTCGCCATTTGCACACATAA
T S C S S N V G V A H D A S G K R V Y Y
 ACCTGACCCGTGATCCGACCACCCCGCTGGCCCGTGCCGCTGGGAAACCGTTTCGTCATA
 3121-----+-----+-----+-----+-----+
 TGGACTGGGCACTAGGCTGGTGGGGCGACGGGCACGGCGGACCCTTTGGCAAGCAGTAT
L T R D P T T P L A R A A W E T V R H T
 CCCCAGTTAATAGCTGGCTGGGCAACATTATTATGTATGCCCCGACCCTGTGGGCGCGTA
 3181-----+-----+-----+-----+-----+
 GGGGCAATTATCGACCGACCCGTTGTAATAATACATACGGGGCTGGGACACCCGGGCAT
P V N S W L G N I I M Y A P T L W A R M
 TGATTCTGATGACCCACTTCTTTAGCATTCTGCTGGCCAGGAACAGCTGGAAAAAGCGC
 3241-----+-----+-----+-----+-----+
 ACTAAGACTACTGGGTGAAGAAATCGTAAGACGAACGGGTCCCTTGTGACCTTTTTTCGCG
I L M T H F F S I L L A Q E Q L E K A L
 TGGATTGCCAGATTTATGGCGCCTGCTATAGCATTGAACCGCTGGATCTGCCGCAGATTA
 3301-----+-----+-----+-----+-----+
 ACCTAACGGTCTAAATACCGCGGACGATATCGTAACTTGGCGACCTAGACGGCGTCTAAT
D C Q I Y G A C Y S I E P L D L P Q I I
 TTGAACGTCTGCATGGCCTGAGCGCCTTTAGCCTGCATAGCTACTCTCCGGGTGAAATTA
 3361-----+-----+-----+-----+-----+
 AACTTGCAGACGTACCGGACTCGCGGAAATCGGACGTATCGATGAGAGGCCCACTTTAAT
E R L H G L S A F S L H S Y S P G E I N
 ATCGTGTGGCGAGCTGTCTGCGTAAACTGGGTGTTCCGCCGCTGCGTGTCTGGCGTCATC
 3421-----+-----+-----+-----+-----+
 TAGCACACCGCTCGACAGACGCATTTGACCCACAAGGCGGCGACGCACAGACCGCAGTAG
R V A S C L R K L G V P P L R V W R H R
 GTGCCCCGTAGCGTTTCGTGCCAAACTGCTGTCTCAGGGTGGCCGTGCCGCCACCTGTGGTA
 3481-----+-----+-----+-----+-----+
 CACGGGCATCGCAAGCACGGTTTGACGACAGAGTCCACCGGCACGGCGGTGGACACCAT
A R S V R A K L L S Q G G R A A T C G K
 AATACCTGTTTAACTGGGCGGTTTCGTACCAAACTGAAACTGACCCCGATTCCGGCGGCGA
 3541-----+-----+-----+-----+-----+
 TTATGGACAAATTGACCCGCCAAGCATGGTTTGACTTTGACTGGGGCTAAGGCCGCCGCT
Y L F N W A V R T K L K L T P I P A A S
 GCCAGCTGGATCTGAGCGGTTGGTTTGTGGCGGTTATAACGGCGGCGATATCTATCATA
 3601-----+-----+-----+-----+-----+
 CGGTGACCTAGACTCGCCAACCAACAACGGCCAATATTGCCGCCGCTATAGATAGTAT
Q L D L S G W F V A G Y N G G D I Y H S
 GCCTGAGCCGTGCCCGTCCGCGTTAATAAACTAGTGGATCCGAGCTCCAGCTTTGTTCC
 3661-----+-----+-----+-----+-----+
 CGGACTCGGCACGGGCGAGGCGCAATTATTTGATCACCTAGGCCTCGAGGTCGAAACAAGG
L S R A R P R * * *

CTAG
3721-----
GATC

Proteína de fusão 4A-3-4B-5B
otimizada para *Escherichia coli*

```

1  GGGCGAATTGGGTACCATGGGGAGCGTTGTTATTGTTGGCCGCATTATTCTGAGCGGTAG
   -----+-----+-----+-----+-----+
   CCCGCTTAACCCATGGGTACCCCTCGCAACAATAACAACCGGCGTAATAAGACTCGCCATC
       M G S V V I V G R I I L S G S
61  CGGTAGCGCCCCGATTACCGCCTATTCTCAGCAAACCCGTGGTCTGCTGGGTTGTATTAT
   -----+-----+-----+-----+-----+
   GCCATCGCGGGGCTAATGGCGGATAAGAGTCGTTTGGGCACCGAGACGCCAACATAATA
   G S A P I T A Y S Q Q T R G I L L G C I I
121 TACCAGCCTGACCGGCCGTGATAAAATCAGGTGGATGGCGAAGTTCAGGTTCTGAGCAC
   -----+-----+-----+-----+-----+
   ATGGTTCGGACTGGCCGGCACTATTTTTAGTCCACCTACCGCTTCAAGTCCAAGACTCGTG
   T S L T G R D K N Q V D C E V Q V L S T
181 CGCCACCCAGAGCTTTCTGGCGACCTGTGTGAATGGTGTGTGCTGCGACCGTTTATGCCGG
   -----+-----+-----+-----+-----+
   GCGGTGGGTCTCGAAAGACCGCTGGACACACTTACCACACACGACCTGGCAATACCGCC
   A T Q S F L A T C V N G V C W T V Y A G
241 TGCCGGTAGCAAAACCCCTGGCGGGTCCGAAAGGTCCGATTACCCAGATGTACACCAACGT
   -----+-----+-----+-----+-----+
   ACGGCCATCGTTTTTGGGACCGCCCGAGGCTTTCCAGGCTAATGGGTCTACATGTGGTTGCA
   A G S K T L A G P K G P I T Q M Y T N V
301 GGATCAGGATCTGGTTGGTTGGCCGGCGCCCGCGGGTCCCCGTAGCATGACCCCGTGTAC
   -----+-----+-----+-----+-----+
   CCTAGTCCTAGACCAACCAACCGGCGCGCGCGCCACGGGCATCGTACTGGGGCACATG
   D Q D L V G W P A P P G A R S M T P C T
361 CTGTGGTAGCAGCGATCTGTATCTGGTTACCCGTCATGCCGATGTTATTCCGGTTGCTCG
   -----+-----+-----+-----+-----+
   GACACCATCGTCGCTAGACATAGACCAATGGGCAGTACGGCTACAAATAAGGCCAAGCAGC
   C G S S D L Y L V T R H A D V I P V R R
421 TCGTGGTGATAGCCGTGGTAGCCTGCTGTCTCCGCGTCCGGTTAGCTATCTGAAAGGTAG
   -----+-----+-----+-----+-----+
   AGCACCACTATCGGCACCATCGGACGACAGAGGCGCAGGCCAATCGATAGACTTTCCATC
   R G D S R G S L L S P R P V S Y L K G S
481 CAGCGGTGGTCCGCTGCTGTGTCCGAGCGGTCATCTGGTGGGTATTTTTGCTCCCGCGT
   -----+-----+-----+-----+-----+
   GTCGCCACCAAGGCGACGACACAGGCTCGCCAGTACACCACCCATAAAAAGCACGGCGGCA
   S G G P L L C P S G H V V G I F R A A V
541 TTGTACCCGTGGTGTGGCGAAAGCGGTGGATTTTATCCCGGTTGAAAGCATGGAAACCAC
   -----+-----+-----+-----+-----+
   AACATGGGCACCAACCGCTTTCCGCCACCTAAAATAGGGCCAACCTTTCGTACCTTTGGTG
   C T R G V A K A V D F I P V E S M E T T
601 CATGCGTAGCCCGGTGTTTACCGATAATAGCAGCCCGCGCGGTTCGCGCAGACCTTTCA
   -----+-----+-----+-----+-----+
   GTACGCATCGGGCCACAAATGGCTATTATCGTCGGGCGCGGCCAAGGCCTCTGGAAAGT
   M R S P V F T D N S S P P A V P Q T F Q
661 GGTTGCCCATCTGCATGCGCGGACCGGTAGCGGTAAAAGCACCAAAGTTCCGGCGCGGTA
   -----+-----+-----+-----+-----+

```

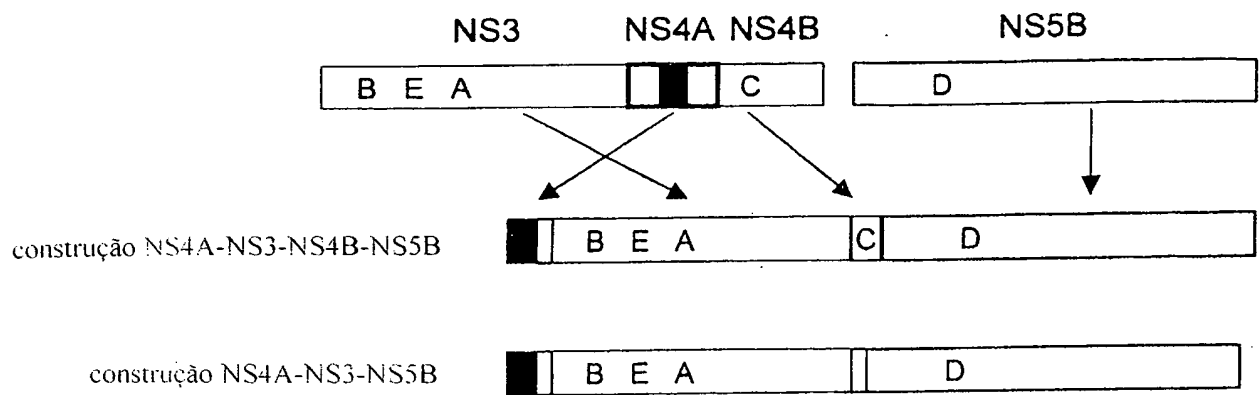
CCAACGGGTAGACGTACGCGGCTGGCCATCGCCATTTTCGTGGTTTCAAGGCCGCCGCAT
V A H L H A P T G S G K S T K V P A A Y
 TGCCGCCCAGGGTTATAAAGTGCTGGTGTGAATCCGAGCGTGGCGGCGACCCCTGGGTTT
 721 -----+-----+-----+-----+-----+-----+-----+
 ACGGCGGGTCCCAATATTTACGACCACGACTTAGGCTCGCACCCGCCGCTGGGACCCAAA
A A Q G Y K V L V L N P S V A A T L G F
 TGGTGCCTATATGAGCAAAGCCCATGGCATTGAACCGAACATTTCGTACCGGCGTTTCGTAC
 781 -----+-----+-----+-----+-----+-----+-----+
 ACCACGGATATACTCGTTTCGGGTACCGTAACCTGGCTTGTAAAGCATGGCCGCAAGCATG
G A Y M S K A H G I E P N I R T G V R T
 CATTACCACCGGTGGCCCGATTACCTATAGCGCCTACGGCAAATTTCTGGCGGATGGTGG
 841 -----+-----+-----+-----+-----+-----+-----+
 GTAATGGTGGCCACCGGGCTAATGGATATCGCGGATGCCGTTTAAAGACCGCCTACCACC
I T T G G P I T Y S A Y G K F L A D G G
 CTGTAGCGGTGGCGCCTATGATATCATCATCTGTGATGAATGCCATAGCACCGATTGGAC
 901 -----+-----+-----+-----+-----+-----+-----+
 GACATCGCCACCGCGGATACTATAGTAGTAGACACTACTTACGGTATCGTGGCTAACCTG
C S G G A Y D I I I C D E C H S T D W T
 CACCATTCTGGGTATTGGCACCGTTCTGGATCAGGCGGAAACCGCCGGTGCCCGTCTGGT
 961 -----+-----+-----+-----+-----+-----+-----+
 GTGGTAAGACCCATAACCGTGGCAAGACCTAGTCCGCTTTGGCGGCCACGGGCAGACCA
T I L G I G T V L D Q A E T A G A R L V
 TGTTCTGGCGACCGCAACGCCGCCGGGTAGCATTACCGTTCCGCATCCGAACATTGAAGA
 1021 -----+-----+-----+-----+-----+-----+-----+
 ACAAGACCGTGGCGTTGCGGCGGCCCATCGTAATGGCAAGGCGTAGGCTTGTAACCTCT
V L A T A T P P G S I T V P H P N I E E
 AGTGGCCCTGAGCAATACCGGCGAAATTCGTTTTATGGCAAAGCGATTCCGATCGAAGC
 1081 -----+-----+-----+-----+-----+-----+-----+
 TCACCGGGACTCGTTATGGCCGCTTTAAGGCAAAATACCGTTTCGCTAAGGCTAGCTTCG
V A L S N T G E I P F Y G K A I P I E A
 GATTAAAGGCGGCCGTCATCTGATTTTTTGCCACAGCAAAAAAATGTGATGAACTGGC
 1141 -----+-----+-----+-----+-----+-----+-----+
 CTAATTTCCGCCGGCAGTAGACTAAAAACGGTGTGCTTTTTTTTTTACACTACTTGACCG
I K G G R H L I F C H S K K K C D E L A
 GGCGAAACTGACCGGTCTGGGTCTGAATGCCGTGGCGTATTATCGTGGTCTGGATGTGAG
 1201 -----+-----+-----+-----+-----+-----+-----+
 CCGCTTTGACTGGCCAGACCCAGACTTACGGCACCGCATAATAGCACCAGACCTACACTC
A K L T G L G L N A V A Y Y R G L D V S
 CGTTATTCCGACCAGCGGTGATGTTGTTGTGGTGGCGACCGATGCCCTGATGACCGGTTT
 1261 -----+-----+-----+-----+-----+-----+-----+
 GCAATAAGGCTGGTCGCCACTACAACAACACCACCGCTGGCTACGGGACTACTGGCCAAA
V I P T S G D V V V V A T D A L M T G F
 TACCGGCGATTTTGATAGCGTGATCGATTGTAACACCTGTGTGACCCAGACCGTTGATTT
 1321 -----+-----+-----+-----+-----+-----+-----+
 ATGGCCGCTAAACTATCGCACTAGCTAACATTGTGGACACACTGGGTCTGGCAACTAAA
T G D F D S V I D C N T C V T Q T V D F
 TAGCCTGGACCCGACCTTTACCATTGAAACCACCGTTCCGCAGGATGCCGTTAGCCG
 1381 -----+-----+-----+-----+-----+-----+-----+
 ATCGGACCTGGGCTGGAAATGGTAACCTTGGTGGTGGCPAGGCGTCTACGGCAATCGGC
S L D P T F T I E T T T V P O D A V S R

TAGCCAGCGTCGTGGTGCCACCGGTCGTGGTCGTAGCGGCATTTATCGTTTTGTGACGCC
 1441-----+-----+-----+-----+-----+-----+-----+
 ATCGGTGCGCAGCACCACGGTGGCCAGCACCAGCATCGCCGTAAATAGCAAAACACTGCGG
 S Q R R G A T G R G R S G I Y R F V T P
 GGGTGAACGTCCGAGCGGTATGTTTGATAGCAGCGTGCTGTGTGAATGTTATGATGCCGG
 1501-----+-----+-----+-----+-----+-----+-----+
 CCCACTTGCAGGCTCGCCATACAACTATCGTCGCACGACACACTTACAATACTACGGCC
 G E R P S G M F D S S V L C E C Y D A G
 CTGTGCCTGGTATGAACTGACCCCGGCGGAAACCACCGTTCTGTCTGCGCGGTATCTGAA
 1561-----+-----+-----+-----+-----+-----+-----+
 GACACGGACCATACTTGACTGGGGCCGCCTTTGGTGGCAAGCAGACGCGCGCATAGACTT
 C A W Y E L T P A E T T V R L R A Y L N
 TACGCCGGGTCTGCCGGTTTGTGTCAGGATCATCTGGAATTCCTGGGAAGCGTTTTTACCGG
 1621-----+-----+-----+-----+-----+-----+-----+
 ATGCGGCCCCAGACGGCCAAACAGTCTTAGTAGACCTTAAGACCCCTTTCGCAAAAATGGCC
 T P G L P V C Q D H L E F W E S V F T G
 CCTGACCCATATTGATGCCCATTTTCTGAGCCAGACCAAACAGCGGGGTGATACTTTCC
 1681-----+-----+-----+-----+-----+-----+-----+
 GGACTGGGTATAACTACGGGTAAAAGACTCGGTCTGGTTTGTCCGCCCACTATTGAAAGG
 L T H I D A H F L S Q T K Q A G D N F P
 GTATCTGGTGGCGTATCAGGCGACCGTTTGTGCCCGTGCCGAGCGCGCGCGCGGAGCTG
 1741-----+-----+-----+-----+-----+-----+-----+
 CATAGACCACCGCATAGTCCGCTGGCAAACACGGGCACGGGTCCGCGCGCGCGGCTCGAC
 Y L V A Y Q A T V C A R A Q A P P P S W
 GGATCAGATGTGGAAATGCCTGATTCTGTGAAACCGACCCCTGCATGGTCCGACCCCGCT
 1801-----+-----+-----+-----+-----+-----+-----+
 CCTAGTCTACACCTTTACGGACTAAGCAGACTTTGGCTGGGACGTACCAGGCTGGGGCGA
 D Q M W K C L I R L K P T L H G P T P L
 GCTGTATCGTCTGGGTGCCGTTTCAGAACGAAATTACCCTGACCCATCCGATCACCAAATT
 1861-----+-----+-----+-----+-----+-----+-----+
 CGACATAGCAGACCCACGGCAAGTCTTGCTTTAATGGGACTGGGTAGGCTACTGGTTTAA
 L Y R L G A V Q N E I T L T H P I T K F
 TGTGATGGCGTGTATGAGCGCCGATCTCGAAGTTGTTAGCCTGATGGCGTTTACCGCCAG
 1921-----+-----+-----+-----+-----+-----+-----+
 AACTACCGCACATACTCGCGCTAGACCTTCAACAATCGGACTACCGCAAATGGCGGTC
 V M A C M S A D L E V V S L M A F T A S
 CATTACCAGCCCGCTGACCACCCAGAATACCCTGCTGTTTAACATCCTGGGCGGTGGGT
 1981-----+-----+-----+-----+-----+-----+-----+
 GTAATGGTGGGCGACTGGTGGGTCTTATGGGACGACAAATTGTAGGACCCGCCAACC
 I T S P L T T Q N T L L F N I L G G W V
 GGCGGCCAGCTGTCTATGAGCTATACCTGGACCGGTGCCCTGATTACCCCGTGTCCCGC
 2041-----+-----+-----+-----+-----+-----+-----+
 CCGCCGGGTCGACAGATACTCGATATGGACCTGSCCAGGGACTAATGGGGCACACGGCG
 A A Q L S M S Y T W T G A L I T P C A A
 CGAAGAAAGCAAACCTGCCGATTAACCCGCTGTCTAATAGCCTGCTGCGTCATCATAGCAT
 2101-----+-----+-----+-----+-----+-----+-----+
 GCTTCTTTTCGTTTACGGCTAATTGGGCGACAGATTATCGGACGACGACAGTAGTATCGTA
 E E S K L P I N P L S N S L L R H H S M
 GGTGTATAGCACCACCAGCCGTAGCGCCAGCCTGCGTCAGAAAAAAGTGACCTTCGATCG
 2161-----+-----+-----+-----+-----+-----+-----+
 CCACATATCGTGGTGGTCCGCATCGCGGTCCGACGACAGTCTTTTTCCTGGAAGCTAGC

V Y S T T S R S A S L R Q K K V T F D R
 TCTGCAGGTGCTGGATGATCATTATCGTGATGTGCTGAAAGAAATGAAAGCGAAAGCGAG
 2221-----+-----+-----+-----+-----+-----+-----+
 AGACGTCCACGACCTACTAGTAATAGCACTACACGACTTTCTTTACTTTTCGCTTTTCGCTC
L Q V L D D H Y R D V L K E M K A K A S
 CACCGTTAAAGCCCGTCTGCTGTCTATTGAAGAAGCGTGTAAGTACGACCCCGCCCATAG
 2281-----+-----+-----+-----+-----+-----+-----+
 GTGGCAATTTCTGGGACAGACAGATAACTTCTTCGCACATTTGACTGGGGCGGCGTATC
T V K A R L L S I E E A C K L T P P H S
 CGCCAAAAGCAAATTTGGCTATGGCGCCAAAGATGTTTCGTAGCCTGAGCAGCCGTGCCGT
 2341-----+-----+-----+-----+-----+-----+-----+
 CGGTTTTTCGTTTAAACCGATACCGCGGTTTCTACAAGCATCGGACTCGTCGGCACGGCA
A K S K F G Y G A K D V R S L S S R A V
 TAATCATATTCGTAGCGTGTGGGAAGATCTGCTGGAAGATACCGAAACCCCGATTGATAC
 2401-----+-----+-----+-----+-----+-----+-----+
 ATTAGTATAAGCATCGCACACCCTTCTAGACGACCTTCTATGGCTTTGGGGCTAACTATG
N H I R S V W E D L L E D T E T P I D T
 CACCATCATGGCGAAAAACGAAGTGTGTTTGTGTTTCAGCCGAAAAAGGTGGTTCGTAAACC
 2461-----+-----+-----+-----+-----+-----+-----+
 GTGGTAGTACCGCTTTTTGCTTCACAAAACACAAGTTCGGCCTTTTTCCACCAGCATTG
T I M A K N E V F C V Q P E K G G R K P
 GGGCGTCTGATTGTTTTTCCGATCTCTGCTGCTGCTGTAAGAAAAATGGCGCTGTA
 2521-----+-----+-----+-----+-----+-----+-----+
 CCGGGCAGACTAACAAAAAGGCCTAGACCCACAAAGCACACACACTTTTTTACCGCGACAT
A R L I V F P D L G V R V C E K M A L Y
 CGATGTTGTTAGCACCCCTGCCGACGGCGGTTATGGGTCCGAGCTATGGCTTTCAGTATTC
 2581-----+-----+-----+-----+-----+-----+-----+
 GCTACAACAATCGTGGGACGGCGTCCGCCAATACCCAGGCTCGATAACGAAAGTCATAAG
D V V S T L P Q A V M G P S Y G F O Y S
 TCCGGGTCAGCGTGTGAATTTCTGGTGAACACCTGGAAAGCAAAAAATGCCCGATGGG
 2641-----+-----+-----+-----+-----+-----+-----+
 AGCCCCAGTCGCACAACTTAAAGACCACTTGTGGACCTTTTCGTTTTTACGGGCTACCC
P G Q R V E F L V N T Y K S K K C P M G
 CTTAGCTATAACACCCGCTGCTTTGATAGCACCGTGACCGAAACGATATCCGTACCGA
 2701-----+-----+-----+-----+-----+-----+-----+
 GAAGTCGATATTGTGGGCGACGAAACTATCGTGGCACTGGCTTTTCTATAGGCATGGCT
F S Y N T R C F D S T V T E N D I R T E
 AGAAAGCATTTACCAAGTGTGTGATCTGGCGCCGGAAGCCCGTCAGGCGATTAAAAGCCT
 2761-----+-----+-----+-----+-----+-----+-----+
 TCTTTTCGTAAATGGTCACGACACTAGACCGCGCCTTCGGGCAGTCCGCTAATTTTCGGA
E S I Y Q C C D L A P E A R Q A I K S L
 GACCGAACGCCTGTATATTGGCGGTCCGCTGACCAATAGCAAAGGCCAGAACTGTGGTTA
 2821-----+-----+-----+-----+-----+-----+-----+
 CTGGCTTGGCGACATATAACCGCCAGCGGACTGGTTATCGTTTCCGGTCTTGACACCAAT
T E R L Y I G G P L T N S K G Q N C G Y
 TCGTCGTTGTGCTGCCAGCGGTGTTCTGACCACCAGCTGTGGTAATACCCTGACCTGCTA
 2881-----+-----+-----+-----+-----+-----+-----+
 AGCAGCAACAGCACGGTCGCCACAAGACTGGTGGTCGACACCATTATGGGACTGGACGAT
R R C R A S G V L T T S C G N T L T C Y
 CCTGAAAGCGACCGCCGCTGTGCTGCCGCCAAACTGCAGGATTGTACCATGCTGGTTAA

2941-----+-----+-----+-----+-----+-----+
 GGACTTTTCGCTGGCGGCGGACAGCACGGCGGTTTGACGTCCTAACATGGTACGACCAATT
L K A T A A C R A A K L Q D C T M L V N
 CGGCAATGATCTGGTGGTGATTTGTGAAAGCGCCGGCAGGAGGATGCCGCCAGCCT
 3001-----+-----+-----+-----+-----+-----+
 GCCGTTACTAGACCACCACTAAACACTTTGCGGCGGTGGGTCTTCTACGGCGGTTCGA
G N D L V V I C E S A G T Q E D A A S L
 GCGCGTTTTTACCAGCGATGACCCGTTATAGCGCCCCGCGGGTGATCCGCCGAGCC
 3061-----+-----+-----+-----+-----+-----+
 CGCGCAAAATGGCTTCGCTACTGGGCAATATCGCGGGCGGCCACTAGGCGGCGTCGG
R V F T E A M T R Y S A P P G D P P Q P
 GGAATATGATCTGGAAGTATCACCAGCTGTAGCAGCAATGTTGGTGTTGCCCATGATGC
 3121-----+-----+-----+-----+-----+-----+
 CCTTATACTAGACCTTGACTAGTGGTTCGACATCGTCGTTACAACCACAAAGGGTACTACG
E Y D L E L I T S C S S N V G V A H D A
 CAGCGGTAAACGTGTGTATTACCTGACCCGTGATCCGACCACCCCGCTGGCCCGTGCCGC
 3181-----+-----+-----+-----+-----+-----+
 GTCGCCATTTGCACACATAATGGACTGGGCACTAGGCTGGTGGGGCGACCGGGCACGGCG
S G K R V Y Y L T R D P T T P L A R A A
 CTGGGAAACCGTTTCGTCATACCCCGGTTAATAGCTGGGCTGGGCAACATTATTATGTATGC
 3241-----+-----+-----+-----+-----+-----+
 GACCCTTTGGCAAGCAGTATGGGGCAATTATCGACCGACCCGTTGTAATAATACATACG
W E T V R H T P V N S W L G N I I M Y A
 CCGGACCCTGTGGGCGCGTATGATTCTGATGACCCACTTCTTTAGCATTCTGCTGGCCCA
 3301-----+-----+-----+-----+-----+-----+
 GGGCTGGGACACCCGGGCATACTAAGACTACTGGGTGAAGAAATCGTAAGACGACCGGGT
P T L W A R M I L M T H F F S I L L A Q
 GGAACAGCTGGAAAAAGCGCTGGATTGCCAGATTTATGGCGCTGCTATAGCATTGAACC
 3361-----+-----+-----+-----+-----+-----+
 CCTTGTCGACCTTTTTTCGCGACCTAACGGTCTAAATACCGCGGACGATATCGTAACTTGG
E Q L E K A L D C Q I Y G A C Y S I E P
 GCTGGATCTGCCGAGATTATTGAACGTCTGCATGGCCTGAGCGCCTTTAGCCTGCATAG
 3421-----+-----+-----+-----+-----+-----+
 CGACCTAGACGGCGTCTAATAACTTGCAGACGTACCGGACTCGCGGAAATCGGACGTATC
L D L P Q I I E R L H G L S A F S L H S
 CTACTCTCCGGGTGAAATTAATCGTGTGGCGAGCTGTCTGCGTAAACTGGGTGTTCCGCC
 3481-----+-----+-----+-----+-----+-----+
 GATGAGAGGCCCACTTTAATTAGCACACCGCTCGACAGACGCATTTGACCCACAAGGCGG
Y S P G E I N R V A S C L R K L G V P P
 GCTGCGTGTCTGGCGTCATCGTGCCCGTAGCGTTCGTGCCAAACTGCTGTCTCAGGGTGG
 3541-----+-----+-----+-----+-----+-----+
 CGACGCACAGACCGCAGTAGCACGGGCATCGCAAGCACGGTTTGACGACAGAGTCCCACC
L R V W R H R A R S V R A K L L S Q G G
 CCGTGCCGCCACCTGTGGTAAATACCTGTTTAACTGGGCGGTTTCGTACCAAAGTAACT
 3601-----+-----+-----+-----+-----+-----+
 GGCACGGCGGTGGACACCATTTATGGACAAATTGACCCGCCAAGCATGGTTTACTTTGA
R A A T C G K Y L F N W A V R T K L K L
 GACCCCGATTCCGGCGGCGAGCCAGCTGGATCTGAGCGGTTGGTTTGTTCGGGTTATAA
 3661-----+-----+-----+-----+-----+-----+
 CTGGGGCTAAGGCCGCGGCTCGGTTCGACCTAGACTCGCCAACCAACAACGGCCAATATT
T P I P A A S Q L D L S G W F V A G Y N

CGGCGGCGATATCTATCATAGCCTGAGCCGTGCCCCGTCCGCGTTAATAAACTAGTGGATC
3721-----+-----+-----+-----+-----+-----+
GCCGCCGCTATAGATAGTATCGGACTCGGCACGGGCAGGCGCAATTATTTGATCACCTAG
_G_G_D_I_Y_H_S_L_S_R_A_R_P_R_*_*_
CGGAGCTCCAGCTTTTGTTCCT
3781-----+-----+-----
GCCTCGAGGTCGAAAACAAGGG



41/45
Fig. 10

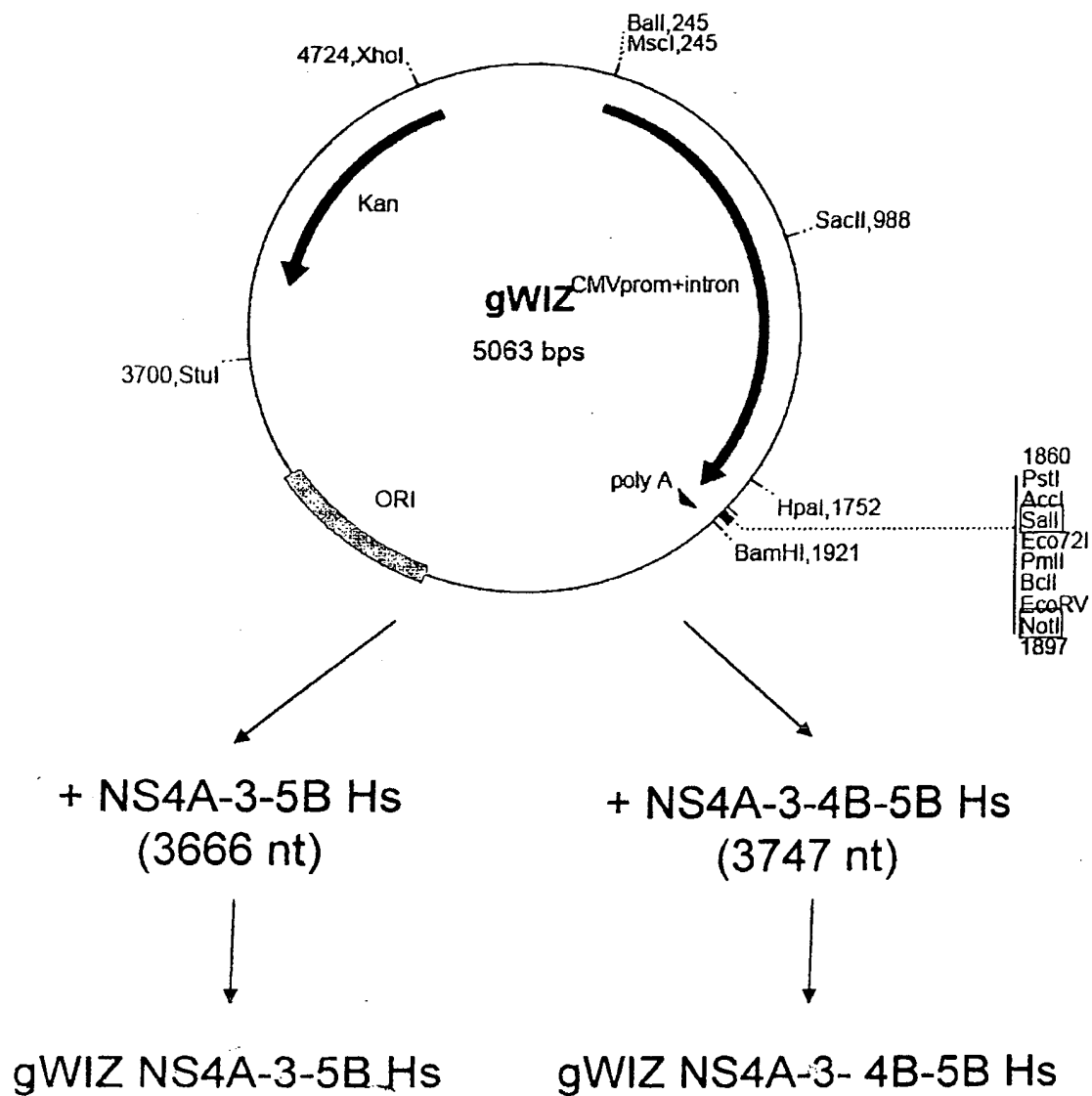
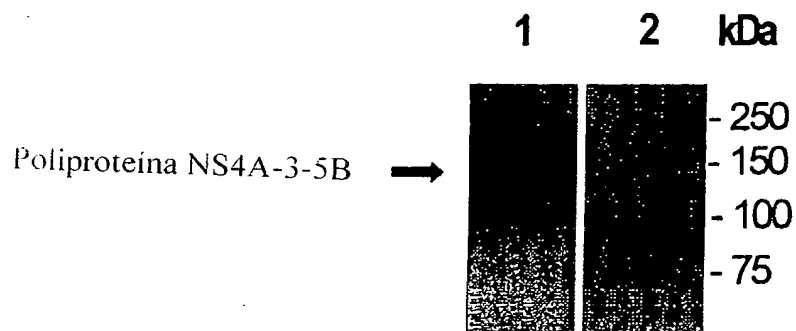
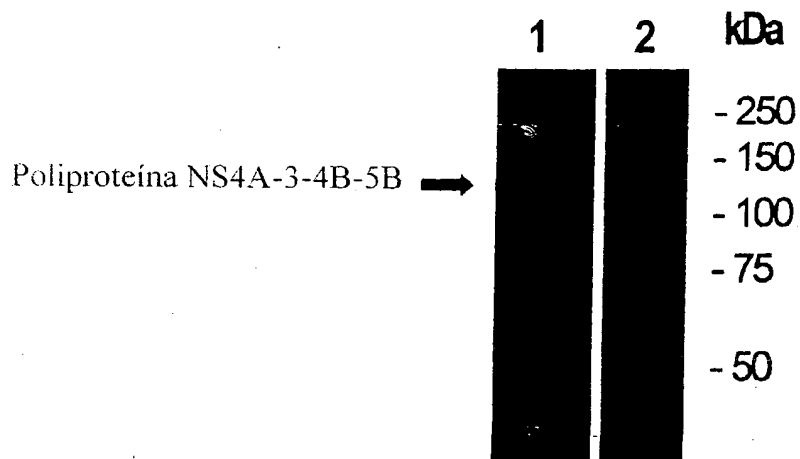


Fig. 11

(A)



(B)



(C)

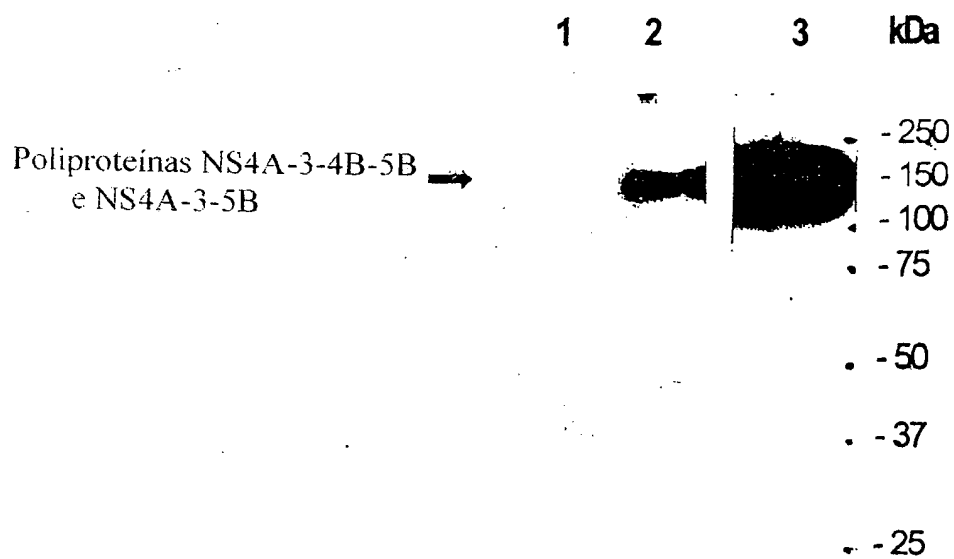
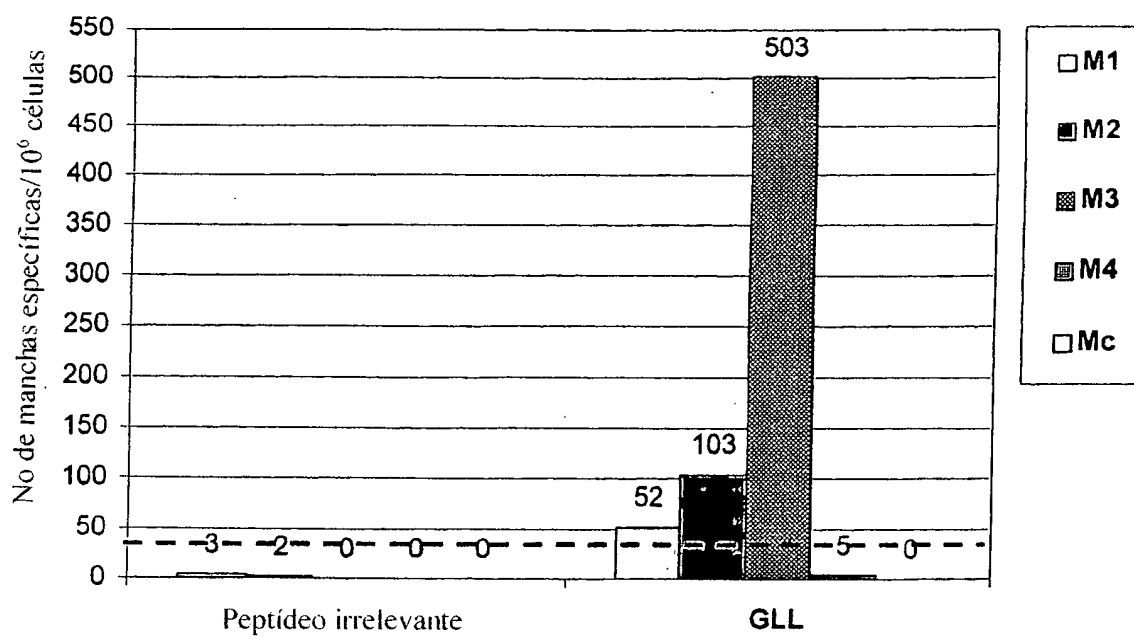
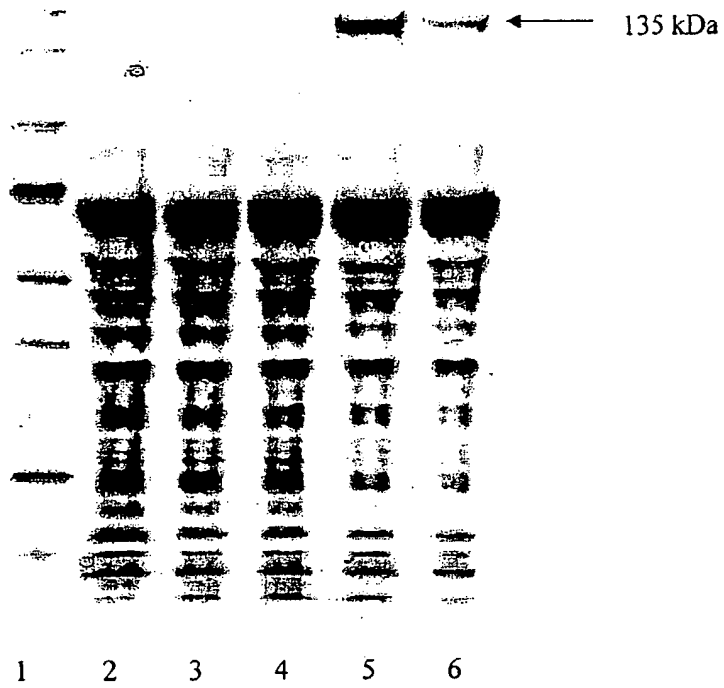


Fig. 12



44/45

Figure 13



45/45
Figure 14

Fig. 14 A

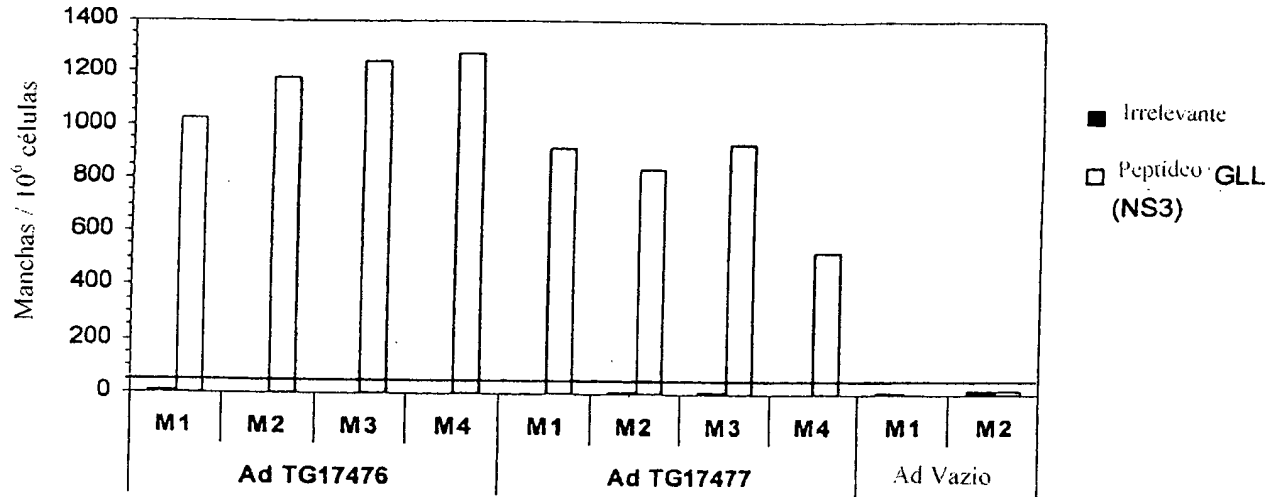


Fig. 14 B

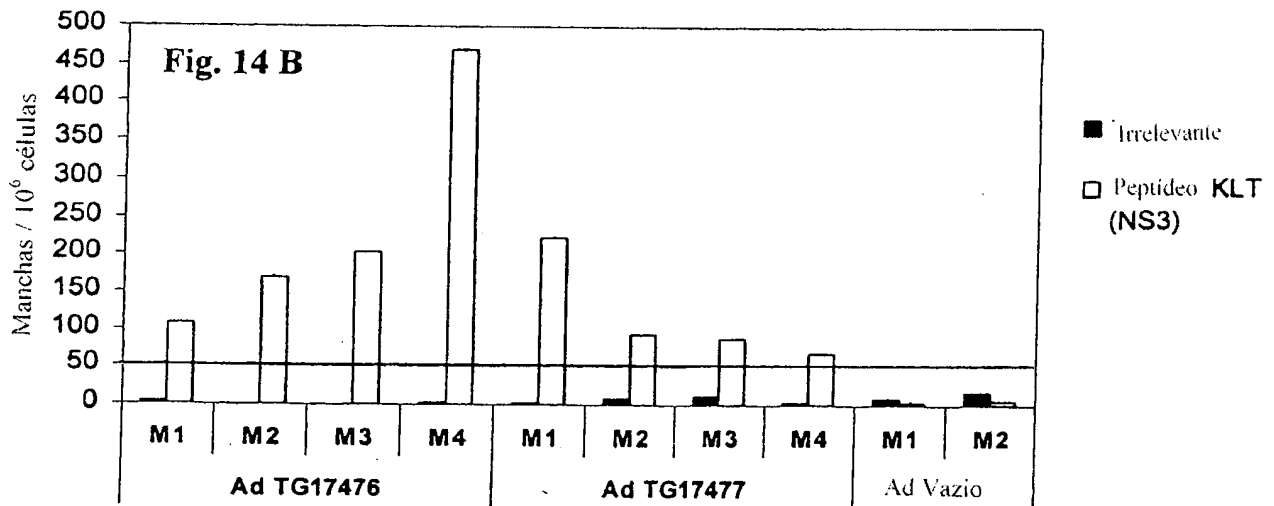
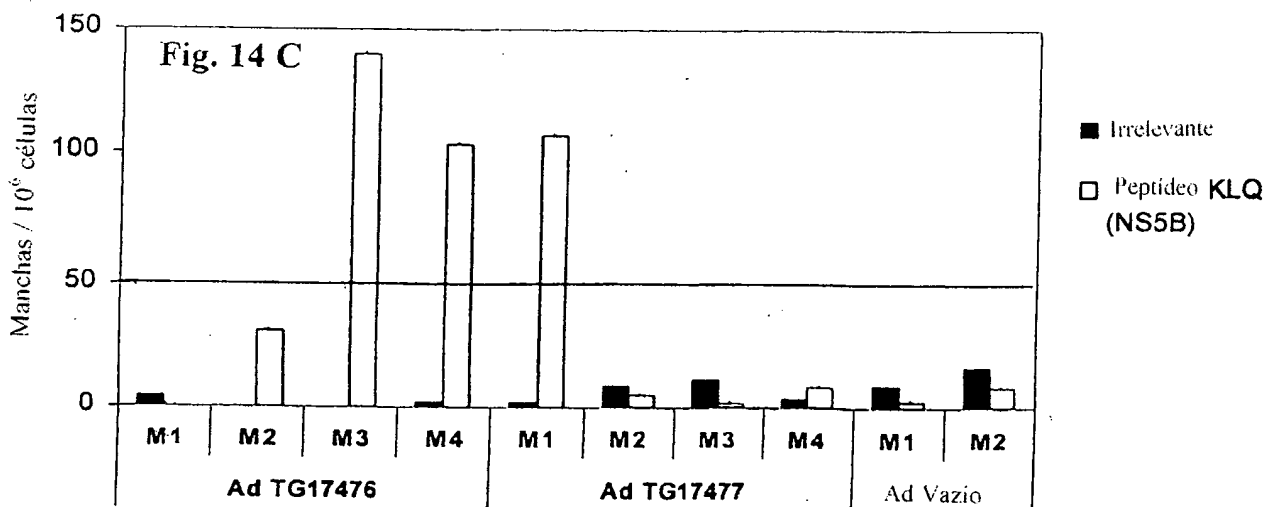


Fig. 14 C



REIVINDICAÇÕES

1. – Proteína de fusão isolada compreendendo pelo menos três polipeptídeos NS, dentre qual polipeptídeo NS4A, que se originam de um vírus da hepatite C (VHC),
5 caracterizada pelo fato de que os referidos polipeptídeos NS são configurados na referida proteína de fusão em uma ordem diferente da ordem na qual eles aparecem na configuração nativa, onde o polipeptídeo NS4A é localizado no N-terminal.

2. – Proteína de fusão isolada, de acordo com a
10 reivindicação 1, caracterizada pelo fato de que a fusão entre cada um dos polipeptídeos NS é direta ou através de um peptídeo de ligação.

3. – Proteína de fusão isolada, de acordo com a reivindicação 1 ou 2, caracterizada pelo fato de que todos os
15 polipeptídeos NS se originam da mesma linhagem ou isolado do VHC.

4. – Proteína de fusão isolada, de acordo com a reivindicação 1 ou 2, caracterizada pelo fato de que pelo menos dois dos polipeptídeos NS se originam de linhagens ou isolados
20 diferentes do VHC.

5. – Proteína de fusão isolada, de acordo com qualquer uma das reivindicações 1 a 4, caracterizada por compreender um polipeptídeo NS4A, um polipeptídeo NS3 e um polipeptídeo NS5B, com NS5B no C-terminal da referida proteína
25 de fusão.

6. – Proteína de fusão isolada, de acordo com qualquer uma das reivindicações 1 a 4, caracterizada por compreender um polipeptídeo NS4A, um polipeptídeo NS3, um polipeptídeo NS4B e um polipeptídeo NS5B com NS5B no C-terminal da referida proteína de fusão.

7. – Proteína de fusão isolada, de acordo com qualquer uma das reivindicações 1 a 6, caracterizada pelo fato de que o polipeptídeo NS3 é modificado se comparado a um polipeptídeo NS3 nativo, de modo a apresentar uma atividade da protease significativamente reduzida.

8. – Proteína de fusão isolada, de acordo com qualquer uma das reivindicações 1 a 7, caracterizada pelo fato de que o polipeptídeo NS3 é modificado se comparado a um polipeptídeo NS3 nativo, de modo a apresentar uma atividade da helicase significativamente reduzida.

9. – Proteína de fusão isolada, de acordo com qualquer uma das reivindicações 1 a 8, caracterizada pelo fato de que o polipeptídeo NS5B é modificado se comparado a um polipeptídeo NS5B nativo, de modo a apresentar uma atividade da RNA polimerase dependente do RNA significativamente reduzida.

10. – Proteína de fusão isolada de acordo com qualquer uma das reivindicações 1 a 9, caracterizada pelo fato de que a referida proteína de fusão não compreende um ou mais dos sítios de clivagem reconhecidos por NS3 normalmente presentes

em um precursor de poliproteína do VHC nas junções NS3/NS4A, NS4A/NS4B, NS4B/NS5A e NS5A/NS5B.

11. – Proteína de fusão isolada, de acordo com qualquer uma das reivindicações 1 a 10, caracterizada pelo fato de
5 que os polipeptídeos NS4A, NS4B e/ou NS5B são modificados de modo a deletar um ou mais domínios hidrófobos normalmente envolvidos na ancoragem membranar dos polipeptídeos NS4A, NS4B e/ou NS5B nativos.

12. – Proteína de fusão isolada, de acordo com
10 qualquer uma das reivindicações 1 a 11, caracterizada pelo fato de que o polipeptídeo NS3 se origina da proteína NS3 apresentada no SEQ ID NO: 1 que é modificada pelo menos de tal maneira que:

o ela retenha a parte que se estende da posição
15 12 até a posição 56, da posição 70 até a posição 155 e da posição 218 até a posição 248;

o ela compreenda a substituição do resíduo His na posição 57 por um resíduo de aminoácido diferente de His, tal como um resíduo Ala;

20 o ela compreenda a substituição do resíduo Thr na posição 269 por um resíduo de aminoácido diferente de Thr, tal como um resíduo Ala; e

o ela compreenda a substituição do resíduo Arg na posição 464 por um resíduo de aminoácido diferente de Arg,
25 tal como um resíduo Ala; e

o ela não compreenda o resíduo Thr na posição 631.

13. – Proteína de fusão isolada, de acordo com a reivindicação 12, caracterizada pelo fato de que o polipeptídeo
5 NS3 compreende a seqüência de aminoácidos apresentada no SEQ ID NO: 5.

14. – Proteína de fusão isolada, de acordo com qualquer uma das reivindicações 1 a 13, caracterizada pelo fato de que o polipeptídeo NS4A se origina da proteína NS4A
10 apresentada no SEQ ID NO: 2 que é modificada pelo menos de tal maneira que:

o ela contenha a parte do SEQ ID NO: 2 da posição 21 até a posição 33;

o ela não contenha a parte do SEQ ID NO: 2 da
15 posição 1 até a posição 20.

15. – Proteína de fusão isolada, de acordo com a reivindicação 14, caracterizada pelo fato de que o polipeptídeo NS4A consiste essencialmente da seqüência de aminoácidos apresentada no SEQ ID NO: 6 precedida por um resíduo Met
20 iniciador.

16. – Proteína de fusão isolada, de acordo com qualquer uma das reivindicações 1 a 15, caracterizada pelo fato de que o polipeptídeo NS4B se origina da proteína NS4B apresentada no SEQ ID NO: 3 que é modificada pelo menos de tal
25 maneira que: o ela compreenda a parte que se estende do resíduo

Ser na posição 78 até o resíduo Leu na posição 109; o ela não compreenda o resíduo Cys na posição 261.

17. – Proteína de fusão isolada, de acordo com a reivindicação 16, caracterizada pelo fato de que o polipeptídeo
5 NS4B opcional consiste essencialmente da seqüência de aminoácidos apresentada no SEQ ID NO: 7.

18. – Proteína de fusão isolada, de acordo com qualquer uma das reivindicações 1 a 17, caracterizada pelo fato de que o polipeptídeo NS5B se origina da proteína NS5B nativa
10 apresentada no SEQ ID NO: 4 que é modificada pelo menos de tal maneira que:

o ela compreenda a parte que se estende do resíduo Arg na posição 155 até o resíduo Leu na posição 182;

o ela compreenda a substituição do resíduo Asp
15 na posição 220 por um resíduo de aminoácido diferente de um Asp, tal como um resíduo Asn;

o ela compreenda a substituição do resíduo Asp na posição 318 por um resíduo de aminoácido diferente de um Asp, tal como um resíduo Asn; e

20 o ela não compreende a parte que se estende do resíduo Trp na posição 571 até o resíduo Arg na posição 591.

19. – Proteína de fusão isolada, de acordo com a reivindicação 18, caracterizada pelo fato de que o polipeptídeo NS5B compreende a seqüência de aminoácidos conforme
25 apresentada no SEQ ID NO: 8.

20. – Proteína de fusão isolada, de acordo com qualquer uma das reivindicações 12 a 19, caracterizada pelo fato de que a referida proteína de fusão compreende uma seqüência de aminoácidos que é homóloga ou idêntica à seqüência de aminoácidos apresentada no SEQ ID NO: 9 ou 10.

21. – Molécula de ácido nucléico isolada, caracterizada por codificar a proteína de fusão de acordo com qualquer uma das reivindicações 1 a 20.

22. – Molécula de ácido nucléico isolada, de acordo com a reivindicação 21, caracterizada por compreender uma seqüência de nucleotídeos otimizada para expressão em uma célula hospedeira de mamífero.

23. – Molécula de ácido nucléico isolada, de acordo com a reivindicação 22, caracterizada por compreender uma seqüência de nucleotídeos que é homóloga ou idêntica às seqüências de nucleotídeos apresentadas no SEQ ID NO: 11 ou SEQ ID NO: 12.

24. – Molécula de ácido nucléico isolada, de acordo com a reivindicação 21, caracterizada por compreender uma seqüência de nucleotídeos otimizada para expressão em uma célula hospedeira de levedura.

25. – Molécula de ácido nucléico isolada, de acordo com a reivindicação 24, caracterizada pelo fato de que a célula hospedeira de levedura é selecionada dentre o grupo que consiste de *Saccharomyces cerevisiae*, *Saccharomyces pombe* e *Pichia pastoris*.

26. – Molécula de ácido nucléico isolada, de acordo com a reivindicação 25, caracterizada por compreender uma seqüência de nucleotídeos que é homóloga ou idêntica às seqüências de nucleotídeos apresentadas no SEQ ID NO: 13 ou
5 SEQ ID NO: 14.

27. – Molécula de ácido nucléico isolada, de acordo com a reivindicação 21, caracterizada por compreender uma seqüência de nucleotídeos otimizada para expressão em uma célula hospedeira procariótica.

10 28. – Molécula de ácido nucléico isolada, de acordo com a reivindicação 27, caracterizada por compreender uma seqüência de nucleotídeos que é homóloga ou idêntica às seqüências de nucleotídeos apresentadas no SEQ ID NO: 15 ou SEQ ID NO: 16.

29. – Vetor, caracterizado por compreender uma
15 ou mais cópias da molécula de ácido nucléico de acordo com qualquer uma das reivindicações 21 a 28.

30. – Partícula viral infecciosa, caracterizada por compreender a molécula de ácido nucléico de acordo com qualquer uma das reivindicações 5 a 28 ou o vetor de acordo com
20 a reivindicação 29.

31. – Célula hospedeira, caracterizada por compreender a molécula de ácido nucléico de acordo com qualquer uma das reivindicações 21 a 28 ou o vetor de acordo com a reivindicação 29 ou a partícula viral infecciosa de acordo
25 com a reivindicação 30.

32. – Método para produzir de forma recombinante a proteína de fusão de acordo com qualquer uma das reivindicações 1 a 20, caracterizado por compreender:

(a) introduzir o vetor de acordo com a
5 reivindicação 29 ou a partícula viral infecciosa de acordo com a reivindicação 30 em uma célula hospedeira para produzir uma célula hospedeira transfectada ou infectada de acordo com a reivindicação 31,

(b) cultivar *in vitro* a referida célula hospedeira
10 transfectada ou infectada sob condições adequadas para o crescimento da célula hospedeira,

(c) recuperar a proteína de fusão da cultura celular, e

(d) como opção, purificar a proteína de fusão
15 recombinante.

33. – Composição, caracterizada por compreender a proteína de fusão de acordo com qualquer uma das reivindicações 1 a 20, a molécula de ácido nucléico de acordo com qualquer uma das reivindicações 21 a 28, o vetor de acordo
20 com a reivindicação 29, a partícula viral infecciosa de acordo com a reivindicação 30, a célula hospedeira de acordo com a reivindicação 31 ou qualquer combinação destes.

34. – Uso da proteína de fusão de acordo com qualquer uma das reivindicações 1 a 20, da molécula de ácido
25 nucléico de acordo com qualquer uma das reivindicações 21 a 28, do vetor de acordo com a reivindicação 29, da partícula viral

infecciosa de acordo com a reivindicação 30, da célula hospedeira de acordo com a reivindicação 31 ou da composição de acordo com a reivindicação 33, caracterizado por ser para a preparação de um fármaco destinado ao tratamento ou prevenção de
5 infecções pelo VHC, doenças associadas ao VHC e condições patológicas.

35. – Uso, de acordo com a reivindicação 34, caracterizado pelo fato de que o referido uso é realizado de acordo com uma modalidade terapêutica de indução-reforço, em que a
10 composição de acordo com a reivindicação 33 é utilizada tanto para ou reforçar, ou para induzir e reforçar uma resposta imunológica anti-VHC.

36. – Uso, de acordo com a reivindicação 34 ou 35, caracterizado pelo fato de que o referido uso é realizado em
15 combinação com quimioterapia com um ou mais fármacos para o VHC.

RESUMO

Patente de Invenção para "PROTEÍNA DE FUSÃO NÃO ESTRUTURAL DO VÍRUS DA HEPATITE C".

A presente invenção revela uma proteína de fusão isolada compreendendo pelo menos três polipeptídeos NS originados de um vírus da hepatite C (VHC) configurados na referida proteína de fusão em uma ordem diferente da ordem na qual eles aparecem na configuração nativa. A presente invenção também se refere a uma molécula de ácido nucléico codificando tal proteína de fusão e a um vetor compreendendo tal molécula de ácido nucléico. A presente invenção também prevê partículas virais infecciosas e células hospedeiras compreendendo tal molécula de ácido nucléico ou tal vetor. A presente invenção também diz respeito a um método para produzir tal proteína de fusão de forma recombinante. Por fim, a presente invenção também prevê uma composição farmacêutica compreendendo tal proteína de fusão, molécula de ácido nucléico, vetor, partículas virais infecciosas e célula hospedeira, bem como o uso terapêutico desta composição para o tratamento ou prevenção de infecções pelo VHC, doenças associadas ao VHC e condições patológicas, assim como um método para induzir ou estimular uma resposta imunológica contra o VHC em um organismo hospedeiro.