

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 2004.07.30	(73) Titular(es): NOVARTIS VACCINES AND DIAGNOSTICS INC. 4560 HORTON STREET, EMERYVILLE, CALIFORNIA 94608-2916 US
(30) Prioridade(s): 2003.07.31 US 491822 P 2004.02.03 US 541565 P	
(43) Data de publicação do pedido: 2006.04.26	(72) Inventor(es): GUIDO GRANDI JOHN TELFORD GIULIANO BENSI US US US
(45) Data e BPI da concessão: 2014.07.09 195/2014	(74) Mandatário: LUÍSA MARIA FERREIRA GUERREIRO PRACETA FERNANDO NAMORA, Nº 7, 3º ESQ. 2820-598 CHARNECA DA CAPARICA PT

(54) Epígrafe: **COMPOSIÇÕES IMUNOGÉNICAS PARA ESTREPTOCOCOS PIOGENES**

(57) Resumo:

A INVENÇÃO INCLUI UM ANTIGÉNIO GAS, GAS 40, QUE É PARTICULARMENTE ADEQUADO PARA UTILIZAÇÃO, QUER ISOLADAMENTE OU EM COMBINAÇÃO COM ANTIGÉNIOS GAS ADICIONAIS, TAL COMO GAS 117, GAS 130, GAS 277, 236 GAS, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, E GAS 057.

RESUMO

COMPOSIÇÕES IMUNOGÉNICAS PARA ESTREPTOCOCOS PIOGENES

A invenção inclui um antigénio GAS, GAS 40, que é particularmente adequado para utilização, quer isoladamente ou em combinação com antigénios GAS adicionais, tal como GAS 117, GAS 130, GAS 277, 236 GAS, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, e GAS 057

DESCRIÇÃO

COMPOSIÇÕES IMUNOGÉNICAS PARA ESTREPTOCOCOS PIOGENES

CAMPO DA INVENÇÃO

Esta invenção está no campo da imunologia e vacinologia. Em particular, refere-se a antigénios derivados de *Streptococcus pyogenes* e a sua utilização em imunização.

ANTECEDENTES DA INVENÇÃO

Os estreptococos do Grupo A ("GAS", *S. pyogenes*) é um agente patogénico humano frequente, que se estima estar presente em entre 5-15% dos indivíduos normais, sem sinais de doença. Quando as defesas do hospedeiro estão comprometidas, ou quando o organismo é capaz de exercer a sua virulência, ou quando é introduzido em tecidos ou hospedeiros vulneráveis, no entanto, ocorre uma infecção aguda. Doenças relacionadas incluem febre puerperal, escarlatina, erisipela, faringite, impetigo, fascíte necrosante, miosite e síndrome do choque tóxico estreptocócico.

GAS é uma bactéria Gram positiva em forma de coco, não formadora de esporos, que normalmente ocorre em cadeias ou em pares de células. Apesar da *S. pyogenes* poder ser tratada com antibióticos, é desejada uma vacina profilática para prevenir o aparecimento da doença. Os esforços para desenvolver uma vacina deste tipo estão em curso há muitas décadas. Embora várias abordagens vacinais de GAS tenham sido sugeridas e algumas abordagens estejam atualmente em testes clínicos, até ao momento, não existem vacinas GAS disponíveis para o público.

A SEQ ID N°: 9188 da WO 02/034771 proporciona a sequência de aminoácidos de GAS40. A WO 02/034771 discute o uso de antígenos de GAS em composições imunogénicas.

A WO 2004/078907 (que foi publicada em 16 de Setembro de 2004 e enquadra-se nos termos do artigo 54 (3) EPC) também fornece a sequência de aminoácidos de GAS40 e discute o uso de antígenos de GAS em composições imunogénicas.

É um objecto da invenção proporcionar outras composições melhoradas e proporcionar imunidade contra a doença de GAS e/ou infecção. As composições incluem GAS40, um factor de virulência GAS identificado pelos Requerentes, que é particularmente adequado para uso em vacinas. Além disso, as composições são baseadas numa combinação de dois ou mais (por exemplo, três ou mais) antígenos GAS.

SUMÁRIO DA INVENÇÃO

Os requerentes descobriram um grupo de trinta antígenos de GAS que são particularmente adequados para fins de vacinação, particularmente quando utilizados em combinação. Além disso, os requerentes identificaram um antígeno GAS (GAS 40), que é particularmente imunogénico usado isoladamente ou em combinação com antígenos adicionais de GAS.

A invenção proporciona uma composição imunogénica que compreende uma combinação de antígenos de GAS, consistindo a referida combinação de 2 a 10 antígenos GAS, na qual a referida combinação inclui GAS40, sendo o dito "GAS40" uma proteína que compreende uma sequência de aminoácidos (a) que tem 95% ou mais de identidade à SEQ ID N°: 1 e/ou (b)

que é um fragmento de pelo menos 200 aminoácidos consecutivos da SEQ ID N°: 1 e em que a referida combinação inclui GAS57, sendo o dito "GAS57" uma proteína que compreende uma sequência de aminoácidos (a) que tem 95% ou mais de identidade com a SEQ ID N°: 116 e/ou (b) que é um fragmento de pelo menos 200 aminoácidos consecutivos da SEQ ID N°: 116. A presente invenção proporciona também a composição para utilização em terapia. A invenção também fornece um método de fabrico dessa composição. A invenção também proporciona um kit compreendendo um primeiro componente que compreende esta composição.

Por conseguinte, a invenção proporciona uma composição imunogénica compreendendo GAS 40 (incluindo os seus fragmentos ou um polipeptídeo com uma identidade de sequência com esta). Um fragmento preferido do GAS 40 compreende uma ou mais regiões de espiral enrolada. A invenção proporciona uma composição imunogénica que compreende uma combinação de antigénios de GAS, consistindo a referida combinação de 2 a 10 antigénios GAS, na qual a referida combinação inclui GAS 40 ou um seu fragmento ou um polipeptídeo que possui identidade de sequência com esta, e GAS 57 ou um seu fragmento ou um polipeptídeo com uma identidade de sequência com esta. De preferência, a combinação é constituída por três, quatro, cinco, seis, sete ou antigénios GAS. Ainda mais preferivelmente, a combinação é constituída por três, quatro, ou cinco antigénios GAS.

Também se descreve aqui uma composição imunogénica que compreende uma combinação de antigénios GAS, a referida combinação consistindo de 2 a 31 antigénios GAS de um primeiro grupo de antigénio, o referido primeiro grupo de

antigénio consistindo em: GAS 117, GAS 130, 277 GAS, GAS 236, GAS 40, GAS 389, 504 GAS, GAS 509, GAS 366, GAS 159, 217 GAS, GAS 309, 372 GAS, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, e GAS 057. Estes antigénios são aqui referidos como o "primeiro grupo de antigénio. De preferência, a combinação de antigénios GAS consiste de três, quatro, cinco, seis, sete, oito, nove, dez ou antigénios GAS selecionados a partir do primeiro grupo de antigénio. De preferência, a combinação de antigénios GAS consiste de três, quatro, ou cinco antigénios de GAS selecionados a partir do primeiro grupo de antigénio.

GAS 39, GAS 40, GAS 57, GAS 117, 202 GAS, GAS 294, GAS 527, GAS 533, e GAS 511 são antigénios GAS particularmente preferidos. De preferência, a combinação de antigénios GAS inclui um ou ambos GAS 40 e GAS 117 Preferivelmente, a combinação inclui GAS 40.

Os exemplos representativos de algumas dessas combinações de antigénios são discutidos abaixo.

A combinação dos antigénios GAS pode consistir em três antigénios GAS selecionados a partir do primeiro grupo de antigénio. Por conseguinte, a combinação de antigénios GAS pode consistir de GAS 40, GAS 117 e um terceiro antigénio GAS selecionado a partir do primeiro grupo de antigénio. As combinações preferidas incluem GAS 40, GAS 117 e um terceiro antigénio GAS selecionado a partir do grupo que consiste de GAS 39, GAS 57, GAS 202, GAS 294, GAS 527, GAS 533, e GAS 511.

A combinação dos antígenos GAS pode consistir de GAS 40 e dois antígenos GAS adicionais selecionados a partir do primeiro grupo de antígeno. As combinações preferidas incluem GAS 40 e dois antígenos GAS selecionados a partir do grupo que consiste de GAS 39, GAS 57, GAS 117, GAS 202, GAS 294, GAS 527, GAS 533, e GAS 511. A combinação de antígenos GAS pode consistir de GAS 117 e dois antígenos GAS adicionais selecionados a partir do primeiro grupo de antígeno.

A combinação dos antígenos GAS pode consistir em quatro antígenos GAS selecionados a partir do primeiro grupo de antígeno. A combinação dos antígenos GAS pode consistir de GAS 40, GAS 117 e dois antígenos GAS adicionais selecionados a partir do primeiro grupo de antígeno. As combinações preferidas incluem GAS 40, GAS 117, e dois antígenos GAS selecionados a partir do grupo que consiste de GAS 39, 57 GAS, GAS 202, GAS 294, GAS 527, GAS 533, e GAS 511.

A combinação dos antígenos GAS pode consistir de GAS 40 e três antígenos GAS adicionais selecionados entre o primeiro grupo de antígeno. As combinações preferidas incluem GAS 40 e três antígenos GAS adicionais selecionados a partir do grupo que consiste de GAS 39, GAS 57, GAS 117, 202 GAS, GAS 294, GAS 527, GAS 533, e GAS 511. A combinação de antígenos GAS pode ser constituída por GAS 117 e três antígenos adicionais selecionados a partir do primeiro grupo de antígeno.

A combinação dos antígenos GAS pode consistir de cinco antígenos GAS selecionados a partir do primeiro grupo de

antigénio. A combinação dos antigénios GAS pode consistir de GAS 40, GAS 117 e três antigénios GAS adicionais seleccionados a partir do primeiro grupo de antigénio. As combinações preferidas incluem GAS 40, GAS 117 e três antigénios GAS adicionais seleccionados a partir do grupo que consiste de GAS 39, GAS 57, GAS 202, GAS 294, GAS 527, GAS 533, e 511 GAS.

A combinação dos antigénios de GAS pode consistir de GAS 40 e quatro antigénios GAS adicionais seleccionados a partir do primeiro grupo de antigénio. As combinações preferidas incluem GAS 40 e quatro antigénios GAS adicionais seleccionados a partir do grupo que consiste de GAS 39, GAS 57, GAS 117, GAS 202, GAS 294, 527 GAS, GAS 533, e GAS 511. A combinação de antigénios GAS pode ser constituída por GAS 117 e quatro antigénios GAS adicionais seleccionados entre o primeiro grupo de antigénio.

A combinação dos antigénios GAS pode consistir de oito antigénios GAS seleccionados a partir do primeiro grupo de antigénio. A combinação dos antigénios de GAS pode consistir de GAS 40, GAS 117 e seis antigénios GAS adicionais seleccionados a partir do primeiro grupo de antigénio. A combinação dos antigénios GAS pode consistir de GAS 40 e sete antigénios GAS adicionais seleccionados a partir do primeiro grupo de antigénio. A combinação dos antigénios GAS pode consistir de GAS 117 e sete antigénios GAS adicionais seleccionados a partir do primeiro grupo de antigénio.

A combinação dos antigénios GAS pode ser constituída por dez antigénios GAS seleccionados a partir do primeiro grupo

de antígeno. A combinação dos antígenos GAS pode consistir de GAS 40, GAS 117 e oito antígenos GAS adicionais selecionados a partir do primeiro grupo de antígeno. A combinação dos antígenos GAS pode consistir de GAS 40 e nove antígenos GAS adicionais selecionados a partir do primeiro grupo de antígeno. A combinação de antígeno GAS pode consistir de GAS 117 e nove antígenos GAS adicionais selecionados entre o primeiro grupo de antígeno.

BREVE DESCRIÇÃO DAS FIGURAS

A FIGURA 1 identifica uma sequência do peptídeo líder, duas sequências em espiral enrolada, uma sequência de fecho de leucina e uma sequência transmembranar dentro de uma sequência de aminoácidos GAS 40.

A FIGURA 2 representa um diagrama esquemático de GAS 40 identificando uma sequência peptídica líder, duas sequências em espiral enrolada, uma sequência de fecho de leucina e uma sequência transmembranar, bem como as regiões de espiral enrolada de GAS 40 que têm baixo nível de homologia com outras proteínas de estreptococos de função conhecida ou prevista.

A FIGURA 3 inclui a análise de alinhamento BLAST das regiões em espiral enrolada de GAS 40 identificadas com outras bactérias de estreptococos.

A FIGURA 4 mostra a estrutura secundária prevista para uma sequência de aminoácidos de GAS 40.

A FIGURA 5 ilustra esquematicamente a localização de GAS 40 dentro do genoma GAS. Também inclui a comparação esquemática que representa um mutante de GAS com o GAS 40

excluído. Mais detalhes sobre estes esquemas demonstram a probabilidade de GAS 40 ser adquirido por transferência horizontal através de um fator transposon.

A FIGURA 6 apresenta a comparação de análise FACS retratando a exposição da superfície do GAS 40 numa estirpe de tipo selvagem (e sem exposição da superfície do mutante de deleção de GAS 40).

A FIGURA 7 apresenta dados da opsonofagocitose para GAS 40 (em várias construções de expressão).

A FIGURA 8 apresenta dados de imunização e desafio para vários antígenos GAS da divulgação.

DESCRIÇÃO DETALHADA DA INVENÇÃO

Como discutido acima, são aqui reveladas composições que compreendem uma combinação de antígenos GAS, no qual as combinações podem ser selecionadas a partir de grupos de antígenos, que os requerentes identificaram como sendo particularmente adequados para fins de vacinação, particularmente quando utilizados em combinação. A invenção proporciona composições que compreendem GAS 40.

GAS 40 e os outros antígenos GAS do primeiro grupo de antígeno estão descritas em mais detalhe abaixo. Sequências genômicas de pelo menos três estirpes de GAS estão disponíveis ao público. A sequência genômica de uma estirpe M1 GAS é relatada em Ferretti et al, PNAS (2001) 98 (8):4658-4663. A sequência genômica de uma estirpe M3 GAS é relatada em Beres et al, PNAS (2002) 99 (15):10078-10083. A sequência genômica de uma estirpe M18 GAS é relatada em Smoot et al, PNAS (2002) 99 (7):4668-4673. De preferência, os antígenos GAS da invenção compreendem

uma sequência de polinucleótido ou de aminoácido de estirpes GAS M1, M3 ou M18. Mais preferencialmente, os antígenos GAS da invenção compreendem uma sequência de polinucleótido ou aminoácido de uma estirpe M1.

Como haverá variação entre os antígenos identificados entre as estirpes dos tipos GAS GAS M e GAS isolados, as referências às sequências do aminoácido GAS ou polinucleotídicas aqui descritas incluem, de preferência, sequências de aminoácidos ou de polinucleótidos que têm identidade de sequência com esta. Sequências de aminoácidos ou de polinucleótidos preferidos têm 50% de identidade de sequência ou mais (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95 %, 96%, 97%, 98%, 99%, 99,5% ou mais). Da mesma forma, as referências a sequências de aminoácidos GAS ou polinucleotídicas aqui reveladas incluem, preferencialmente, fragmentos dessas sequências, (isto é, fragmentos que retêm ou codificam para as propriedades imunológicas do antígeno GAS). Fragmentos de aminoácidos preferidos incluem, pelo menos, n aminoácidos consecutivos, em que n é 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 ou mais). Fragmentos de polinucleótidos preferidos incluem, pelo menos, n polinucleótidos consecutivos, em que n é 12 ou mais (por exemplo, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250 ou mais). Os fragmentos de aminoácidos ou polinucleotídicas da invenção podem não ser idênticos às sequências de aminoácidos ou de polinucleótidos de outras (não-GAS) bactérias (por exemplo, os fragmentos não são idênticos às sequências de outras bactérias *Streptococcus*).

(1) GAS 40

GAS 40 corresponde ao número de acesso M1 GenBank GI: 13621545 e GI: 15674449, para o número de acesso M3 GenBank GI: 21909733, para o número de acesso M18 GenBank GI: 19745402, e também é conhecido como 'Spy0269' (M1), 'SpyM3_0197' (M3), 'SpyM18_0256' (M18) e 'prgA'. GAS 40 também tem sido identificado como uma proteína de superfície de exclusão putativa. Sequências de aminoácidos e de polinucleótidos de 40 GAS partir de uma estirpe M1 são apresentados abaixo e na listagem de sequências como SEQ ID N°s: 1 e 2.

SEQ ID NO: 1

MDLEQTKFNOVKQKIALTSTIALLSASVGVSHQVKADDRASGETKASNTHDSDLPRPETIQEAKATIDAVEKTLSSQKAE
LTELATALTKTITAEINHLKEQQDNEQKALTSAGEIYTNNTLASSEETLLAQGAHQHRELTAETELHNAQADQHSKETALS
EQKASISAETTRAQDLVEQVKTSEQNIARLNAMISNPDAITKAAQTANENTKALSSELEKAKADLENQKARVKEQLTEEL
AAQKAALAEKEAELSRLLKSSAPSTQDSIVGNWNTMKAPQGYPLEELKLEASGYIGSASYNVYKEHADQIIAKASPGNQL
NQYQDIAPADRNRFYDFDNLTPEVQNELAQFAHMINSVRRQLGLFPVTVTAGSQEFARLLSTSYKKTHGNTFRSFVYGQP
GVSGHYGVGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGTYDSIKYMLFTDHLHGNTYGHATNFLRVD
KHNPNAFYYLGFSTSNVGSLSNEHFVMPFESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGVSSLENLSAI
HQEADIMAAQAKVSQLOGNLASTLNQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLASLKAAALHOTE
ALAEQAAARVTALVAKKAHLQYLRFKLNPNRLQVIERERIDNTKQDLAKTTSSLENAQEALAAALQAKQSSLEATIATTEH
QLTLLKTLANEKEYRHLDEDIATVPDLQVAPPLTGKPLSYSKIDTTPLVQEMVKETKOLLEASARLAENTSLSVAEALV
GOTSEMVASNAIVSKITSSITQPSSKTSYSGSGSSTTSNLI SDVDESTQRAIKAGVVMIAAVGLTGFRFBKESK

SEQ ID NO: 2

ATGGACTTAGAACAAACGAAGCCAAACCAAGTTAAGCAGAAATTECTTTAACCTCAACAATTGCTTTATTGAGTSCCAG
TGTAGCGGTATCTCACCAGTCAAAGCAGATCATAGAGCCTCAGGAGAAACGAAGGCGAGTAATACTCAGCAGCATAGTT
TACCAAAACGAGAAACAATTCAAGAGGCAAAGCCAACTATTGATGCAGTTGAAAAAATCTCTCAGTCAACAAAAAGCAGAA

CTGACAGAGCTTGCTACCGCTCTGACAAAACTACTGCTGAAATCAACCCCTTAAAAAGAGCAGCAAGATAATGAACAAA
 AGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCTTGGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAAC
 ATCAAAGAGAGTTAAACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAGAGAGCTGCATTGTCA
 GAACAAAAGCTAGCATTTCAGCAGAACTACTCGAGCTCAAGATTASTGGAACAAGTCAAAACGCTGTAACAAAATAT
 TGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCAT
 TAAGCTCAGAAITGGAGAAGGCTAAAGCTGACTTAGAAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAAGAGTTG
 GCAGCTCAGAAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCTGCTTAAATCCCTCAGCTCCGCTACTCAAGATAG
 CATTGTGGGTAATAATACCATGAAGCAGCCGCAAGGCTATCCTCTTGAAGAACCTAAAAAATTAGAAGCTAGTGGTTATA
 TTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAAATCAATTA
 AATCAATACCAAGATATTTCCAGCAGATCGTAATCGCTTTGTTGATCCCGATAATTGACACCAGAGTGCAAAATGAGCT
 AGCGCAGTTTGCAGCTCACATGATTAAATAGTGTAAAGAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCAC
 AAGAAATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACCTCATGGTAATACAAGACCATCATTTGTCTACGGACAGCCA
 GGGGTATCAGGGCATTATGGTGTGGGCTCATGATAAACTATTATTGAAGACTCTGCCGAGCGTCAGGGCTCATTCC
 AAATGATGATAACATGTACGAGATATCGGTGCTTTTAAACGATGTGCATACTGTGAATGGIATTAACCGTGGTATTATG
 ACAGTATCAAGTATATGCTCTTTACAGATCATTACACGGAAATACATACGGCCATGCTATTAACTTTTACGTGTAGAT
 AAACATAACCCCTAATGCGCTGTTTACCTTGGATTTTCAACCCAGCAATGTAGGATCTTTGAATGAACACTTTGTAAATGTT
 TCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCTATAAAAGCGGTGGAGGTACAAAAGATTATGCC
 AAAGAGTAGCCACTGTATCTGATACATTATGCAGCGATCAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTGCGCTATT
 CATCAAGAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTC
 AGACAGCTTAAATCTCCAAGTGAGACAATTAATGATACTAAAGGTTCTTTGAGAACAGAACTACTAGCAGCTAAAGCAA
 AACAGGCACAACCTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAA
 GCCTTAGCAGAGCAAGCCGACCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCTAAGGGACTTTAA
 ATTGAATCCTAACCGCTTCAAGTGATACGTGAGCGCATGATAATTAAGCAAGATTGGCTAAAACTACCTCATCTT
 TGTAAATGCACAGAGAGCTTTAGCAGCCTTACAAGCTAAACAAGCAGTCTAGAAGCTACTATTGCTACCAACAGAACAC
 CAGTTGACTTTGCTTAAACCTTAGCTAACGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCTGATTT
 GCAAGTAGCTCCACCTCTTACGGGGCTAAACCCCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAGGAATGG
 TTAAGGAACGAAACAATATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAAGCGCTTGT
 GGCCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGCTTAAATCACATCTTCGATTACTCAGCCCTCATCTAAGAC
 ATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAAGAGCTCTTAAAGCAG
 GAGTCTGTCATGTTGGCAGCTGTCTGGCTCACAGATTTAGGTTCCGTAAGGAATCTAAGTGA

As proteínas GAS 40 para utilização com a invenção compreendem uma sequência de aminoácidos: (a) possuindo 95% ou mais de identidade (por exemplo, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 1; e/ou (b) que é um fragmento de aminoácidos de, pelo menos, *n* consecutivos da SEQ ID N°: 1, em que *n* é 200 ou mais (por exemplo, 250 ou mais). Estas proteínas GAS 40 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.), da SEQ ID N°: 1. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 1. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 1.

Por exemplo, numa concretização, a sequência de aminoácidos sublinhada na extremidade do terminal N (sequência líder)

da SEQ ID N°: 1 é removida. (As sequências de aminoácidos e de polinucleótidos para esta sequência líder do terminal N estão listados na listagem de sequências como SEQ ID N°s: 3 e 4. As sequências de aminoácidos e de polinucleótidos para o restante fragmento GAS 40 estão listadas na listagem de sequências como SEQ ID N°s: 5 e 6)

Como outro exemplo, numa forma de realização, a sequência de aminoácidos sublinhada na extremidade do terminal C (a região transmembrantar) da SEQ ID N°: 1 é removida. (As sequências de aminoácidos e de polinucleótidos para esta região transmembrantar estão listadas na listagem de sequências como SEQ ID N°s: 7 e 8. As sequências de aminoácidos e de polinucleótidos para o restante fragmento GAS 40 estão listadas na listagem de sequências como SEQ ID N°s: 9 e 10).

Outros fragmentos podem omitir um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, dum domínio citoplasmático, dum domínio transmembrantar, ou dum domínio extracelular).

Outra ilustração de domínios dentro de GAS 40 é mostrada nas FIGURAS 1 e 2 Como mostrado nestas figuras, uma sequência de aminoácidos para GAS 40 (SEQ ID N°: 1) contém uma sequência líder de peptídeo nos aminoácidos 1-26 (por exemplo, SEQ ID N°: 3), uma primeira região em espiral enrolada nos aminoácidos 58-261 (SEQ ID N°: 12), uma segunda região em espiral enrolada geralmente nos aminoácidos 556-733 (SEQ ID N°: 13), uma região de fecho de leucina nos aminoácidos 673-701 (SEQ ID N°: 14) e uma região transmembrantar nos aminoácidos 855-866 (SEQ ID N°:

11). A Figura 1 ilustra essas regiões dentro de uma sequência de aminoácidos para GAS 40, enquanto a Figura 2 representa, esquematicamente, estas regiões ao longo do comprimento da proteína GAS 40.

As regiões em espiral enrolada identificadas dentro do GAS 40 são susceptíveis de formar bobinas helicoidais alfa. Estas estruturas são frequentemente envolvidas em interações de oligomerização, por exemplo, entre diferentes regiões da proteína, ou entre regiões de duas proteínas distintas. O motivo de fecho de leucina no interior da segunda região em espiral enrolada contém uma série de resíduos do aminoácido leucina (ou isoleucina), espaçados de tal forma a permitir que a proteína forme uma interação de oligomerização especializada entre duas hélices alfa. Num motivo de fecho de leucina, de preferência, existem seis resíduos de aminoácido intercalados entre os resíduos de leucina repetidos. Numa estrutura oligomérica de fecho de leucina, as hélices alfa são pensadas para serem mantidas juntas por interações hidrófobas entre resíduos de leucina, que estão localizados em cada um dos lados da hélice. Motivos de fecho de leucina estão frequentemente envolvidos em interações de dimerização. A localização do motivo de fecho de leucina no interior da região em espiral enrolada indica ainda a probabilidade de que esta região da proteína GAS 40 esteja envolvida numa interação de oligomerização.

A FIGURA 2 também ilustra que há um baixo nível de homologia entre algumas das regiões identificadas Da proteína GAS 40 e outras proteínas de estreptococos com duas estruturas tridimensionais conhecidas ou previsíveis, ou localização da superfície. Esse baixo nível de

homologia pode indicar estruturas secundárias semelhantes ou mesmo função. Por exemplo, os aminoácidos 33-324 de GAS 40, incluindo a primeira região em espiral enrolada, têm aproximadamente 22% de identidade de sequência com uma região (aminoácidos 112-392) de uma proteína a partir de *Streptococcus gordonii* chamada precursor de proteína de superfície de estreptococos A ("SpA") (referência GenBank GI 25990270, SEQ ID N°: 15). Esta proteína é provavelmente uma proteína de superfície de adesão, envolvida com a adesão de estreptococos com membranas de células hospedeiras de mamíferos. O SpA *S. gordonii* é um membro da família estreptocócica de antigénios I/II de adesões de proteína e reconhece a glicoproteína salivar aglutinina (gp-340) e colagénio tipo I. Os aminoácidos 33-258 de GAS 40 também mostram baixo nível de identidade de sequência (23%) com outra proteína *S. gordonii*, estreptococos na superfície da proteína precursora B (referência GenBank GI 25055226, SEQ ID N°: 16).

Uma região semelhante de GAS 40, que também se sobrepõe com a primeira região em espiral enrolada (aminoácidos 43-238) demonstra cerca de 23% de identidade de sequência com uma região (aminoácidos 43-238) de uma proteína a partir de *Streptococcus pneumoniae* denominada proteína precursora de superfície PspA (referência Genbank GI 282335, SEQ ID N°: 17). O domínio do terminal amino de PspA pensa-se ser essencial para a virulência pneumocócica completa, e os anticorpos monoclonais produzidos contra ele protege os ratos contra infecções por pneumococos. O domínio PspA tem uma forma monomérica com um rácio de forma axial de cerca de 1:12, típico de proteínas fibrosas. As análises das sequências indicam uma estrutura alfa-helicoidal em espiral

enrolada para esta molécula monomérica, com poucas pausas do tipo laço em helicidade.

A segunda região de espiral enrolada de GAS 40 contém cerca de 46% de identidade de sequência com uma região (aminoácidos 509-717) de uma proteína a partir de *Streptococcus equi* chamadas proteínas imunorreativas Se89.9 (referência Genbank GI 2330384, SEQ ID N°: 18) (a sequência de comprimento total para Se89.9 também está disponível em <http://pedant.gsf.de>). Esta proteína de *Streptococcus equi* é predita para ser uma superfície exposta. O alinhamento BLAST de cada uma destas sequências de estreptococos com GAS 40 é apresentado na Figura 3.

Além disso, outra ilustração de duas estruturas bidimensionais de GAS 40 é mostrada na Figura 4. Em primeiro lugar, a Figura 4 (a) apresenta a análise da estrutura secundária prevista alinhada contra a sequência de aminoácidos para GAS 40. As regiões alfa helicoidais previstas da Figura 4 correspondem geralmente às anteriormente observadas regiões em espiral enrolada. Na Figura 4 (b), a previsão PairCoil é utilizada para prever o local das espirais enroladas putativas. Aqui, duas regiões helicoidais são identificados, geralmente correspondentes à primeira e segunda regiões de espiral enrolada. A Figura 4 (c), salienta a região de fecho de leucina e ilustra a leucina regularmente repetida (isoleucina) ou resíduos de aminoácidos que são susceptíveis de participar no fecho de leucina.

Por conseguinte, a primeira região de espiral enrolada de GAS 40 compreende uma sequência de aminoácidos de pelo

menos dez (por exemplo, pelo menos 10, 13, 15, 18, 20, 25, 30, 35, 40, 50, 70, 90, 100 ou mais) resíduos de aminoácidos consecutivos, selecionados a partir da metade do terminal N de um comprimento total da sequência de GAS 40, e prevista para formar um complexo alfa-helicoidal com base nas características funcionais dos resíduos de aminoácidos na sequência. A SEQ ID N°: 12, é uma primeira região em espiral enrolada preferida do GAS 40.

A segunda região em espiral enrolada de GAS 40 compreende uma sequência de aminoácidos de pelo menos dez (por exemplo, pelo menos 10, 13, 15, 18, 20, 25, 30, 35, 40, 50, 70, 90, 100 ou mais) resíduos de aminoácidos consecutivos, selecionados a partir da metade do terminal C de comprimento completo da sequência GAS de 40, e prevista de modo a formar um complexo alfa-helicoidal com base nas características funcionais dos resíduos de aminoácidos na sequência. A segunda região em espiral enrolada, de preferência inclui um motivo de fecho de leucina. A SEQ ID N°: 13, é uma segunda região espiral enrolada preferida de GAS 40.

As regiões em espiral enrolada de GAS 40 são susceptíveis de serem envolvidas na formação de oligómeros, tais como dímeros ou trímeros. Esses oligómeros poderiam ser homómeros (contendo duas ou mais proteínas GAS 40 oligomerizadas juntas) ou heterómeros (contendo uma ou mais proteínas GAS adicionais oligomerizadas com GAS 40). Alternativamente, as primeira e segunda regiões de espiral enrolada podem interagir em conjunto no interior da proteína GAS 40 para formar reações oligoméricas entre a primeira e segunda regiões em espiral enrolada.

Deste modo, numa forma de realização, as composições da invenção incluem um antigénio GAS 40 sob a forma de um oligómero. O oligómero pode compreender mais dois antigénios GAS 40 ou os seus fragmentos, ou pode compreender GAS 40 ou um seu fragmento oligomerizado para um segundo antigénio GAS. Os fragmentos GAS 40 preferidos compreendem uma sequência de aminoácidos selecionada a partir do grupo que consiste na primeira região em espiral enrolada e na segunda região em espiral enrolada. Tais fragmentos GAS 40 preferidos podem ser utilizados sozinhos ou em combinações da invenção.

As sequências de polinucleótidos GAS e de aminoácidos da invenção podem ser manipuladas para facilitar ou otimizar a expressão recombinante. Por exemplo, a sequência líder do terminal N pode ser substituída com uma sequência que codifica para uma proteína de marcação tal como polihistidina ("HIS") ou glutathione-S-transferase ("GST"). Tais proteínas de marcação podem ser utilizadas para facilitar a purificação, detecção e estabilidade da proteína expressa. As variações de tais modificações para GAS 40 são discutidas abaixo. Tais modificações podem ser aplicadas a qualquer uma das proteínas GAS da invenção.

Um exemplo de uma sequência 40 GAS tanto com a tag GST como HIS é aqui denominada como "GST 40 HIS". Esta construção inclui a sequência GAS 40 onde a sequência líder é removida, uma sequência de marcação de codificação de GST é adicionada à extremidade do terminal N, e uma sequência de marcação de codificação HIS é adicionada ao terminal C (usando, por exemplo, um vector pGEXNNH com os locais de restrição NdeI e NotI). As sequências polinucleotídicas e de aminoácidos para a região fundida da marcação GST, a

sequência 40 GAS e o terminal C de marcação HIS de GST 40HIS são mostradas nas SEQ ID N°s: 19 e 20.

Alternativamente, pode ser usada uma única sequência de marcação. Um exemplo de uma sequência GAS 40 com apenas uma marcação HIS é denotado como "40a-HIS". Esta construção inclui a sequência GAS 40, onde a sequência líder do terminal N e o terminal C que contém a sequência transmembranar é removido. Nesta construção, a sequência de marcação HIS é adicionada ao terminal C (utilizando, por exemplo, um vector de clonagem tal como pET21b + (Novagen) para os locais de restrição NdeI e NotI). As sequências polinucleotídicas e de aminoácidos para 40a-HIS são mostradas nas SEQ ID N°s: 21 e 22.

Além da adição de marcações de purificação, a expressão recombinante também pode ser facilitada através da optimização de sequências de codificação para as mais abundantes ou acessíveis ao hospedeiro recombinante. Por exemplo, a sequência polinucleotídica AGA codifica um resíduo de aminoácido arginina. A arginina pode também ser codificada pela sequência polinucleotídica CTG. Este códon CTG é preferido pelas enzimas de translação em *E. coli*. Na sequência polinucleotídica 40a-HIS da SEQ ID N° 21, um terminal C de CTG que codifica para arginina foi substituído com CGT.

Os seguintes códons são geralmente sub-representados no *E.coli*: AGA, AGG e CGA. Quando estes códons ocorrem numa sequência polinucleotídica GAS, podem ser substituídos com um dos outros dois códons opcionais que codificam para o mesmo resíduo de aminoácido.

Um total de três códãos ATG são otimizados para CTG na construção "40a-RR-HIS", a SEQ ID N°s 23 e 24. A SEQ ID N° 23 é também mostrada abaixo, com os códãos otimizados sublinhados. (com exceção do códão de otimização adicional, 40a-RR-HIS é idêntico à 40a-HIS).

SEQ ID N: 23

ATGAGTGTAGCGGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATACTCACGACG
ATAGTTTACCAAAACAGAAACAATTCAAGAGGCAAAAGGCAACTATTGATGCAGTTGAAAAAACTCTCAGTCAACAAAA
AGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAACCTACTGCTGAAATCAACCACTTAAAGAGCAGCAAGATAAT
GAACAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAG
GAGCCGACATCAAAGAGAGTTAACAGCTTCTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAGAGAGC
TGCATTGTTCAGAACAAAAGCTAGCATTTTCAGCAGAACTACTCGAGCTCAAGATTTAGTGAACAAAGTCAAAACGCTCT
GAACAAAATATTGCTAAGCTCAATGCTTATGATTAGCAATCTGTATGCTATCACTAAAGCAGCTCAAACGGCTAATGATA
ATACAAAAGCATTAAGCTCAGAAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAGGAAT

GACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAGAGGCGAGAACTTAGTCTGCTTTAAATCCTCAGCTCCG
TCTACTCAAGATAGCATTTGTGGTAAATAATACCATGAAGACCCGCAAGGCTATCCTCTTGAGAGAACTTAAAAAATTAG
AAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATACAAAGAGCATGCAGATCAATTTATGCCAAAGCTAG
TCCAGGTAAATCAATTAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTTTGTGTGATCCCGATAATTTGACACCA
GAAGTGCAAAATGAGCTAGCGCAGTTTGAGCTCAGATGATTAATAGTGTAACTGCTCAATTAGGTTACCAACAGTTA
CTGTTACAGCAGGATCACAAGAAATTTGCAAGATTACTTAGTACCAGCTATAAGAAACTCATGGTAATACAAGACCATC
ATTTGTCTACGGACAGCCAGGGCTATCAGGGCATTTATGGTGTGGGCTCATGATAAACTATTATTGAAGACTCTGCC
GGAGCGCTCAGGGCTCATTCGAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCACTACTGTGAATG
GTATTAAAGCTGGTATTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTCACGGAATACATACGGCCATGC
TATTAACCTTTTACGCTGATAGATAAACATAACCCCTAATGCGCTGTTTACCTTGGAATTTTCAACAGCAATGTAGGATCT
TTGAATGAACACTTTGTATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCTATPAAAAGCCG
TTGGAAGTACAAAAGATTATGCCAAAGAGTAGGCACTGTATCTGATACTATTGCAGCGATCAAAGGAATAAGTAAGCTC
ATTAGAAAATCGTTTGTGCGCTATTTCATCAAGAAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGST
AAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAAATGAGACAATTAAATGATACTAAAGGTTCTTTGA
GAACAGAAATTAAGCAGCTAAAGCAAAACAAGCAAACTCGAAGCTACTCTGTGATCAATCATTAGCTAAGCTAGCATC
GTTGAAAGCCGCACTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGAGCCAGAGTGCAGCACTGGTGGCTAAAAAA
GCTCATTTGCAATATCTAAGGGACTTTAAATTTGAATCTTAACCGCTTCAAGTGTATCGTGAGCGCATTCGATAATACTA
AGCAAGATTGTGGCTAAAACCTACCTCATCTTTGTTAAATGCCACAAGAAGCTTTAGCAGCCTTCAAGCTAAACAAAGCAG
TCTAGAAGCTACTATTGCTACCAAGAACACCAAGTTGACTTTGCTTAAACCTTAGCTAACGAAAAGGAATATCGCCAC
TTAGACGAAGATATAGCTACTGTGCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAACCGCTATCATATAGTA
AGATAGATACTACTCCGCTTGTTCAGAAATGGTTAAAGAAACGAAACAACCTATTAGAAGCTTCAGCAAGATTAGCTGC
TGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTGGCCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGTCTAAA
ATCAGATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTG
ATGTTGATGAAAGTACTCAAGCTGCGGCCGACTCGAGCACCACCACCACCACC

A otimização do códão também pode ser utilizada sem uma marcação de purificação. A construção "40a-RR-Nat", SEQ ID N°s: 25 e 26, fornece um exemplo. Esta construção compreende GAS 40 sem a sequência líder do terminal N e a sequência da transmembrana do terminal C, com três otimizações de códãos (e não inclui uma sequência de marcação HIS).

Diferentes vectores de clonagem podem ser utilizados para otimizar a expressão em células hospedeiras diferentes ou

sob diferentes condições de cultura. As construções acima discutidas usavam o vector pET21b+ (Novagen) que inclui um promotor indutível por IPTG. Como alternativa, um vector de expressão de transporte de *E. coli*/*B. subtilis* como pSM214gNH pode ser utilizado. Este vector utiliza um promotor constitutivo em vez de um promotor indutível por IPTG. Um exemplo de uma construção de GAS 40 utilizando este vector é denotado como "HIS-40a-NH", SEQ ID N°s 27 e 28. Nesta construção, tanto a sequência líder do terminal N como a sequência transmembranar do terminal C são removidas, e é adicionado um marcador HIS ao terminal N. Aminoácidos adicionais do terminal N são introduzidos com a clonagem. Além disso, são indicadas duas alterações de nucleótidos que provavelmente ocorrem durante a PCR - nenhuma destas modificações resulta em alterações de aminoácidos.

Como outra alternativa, pode ser usado o vector de transporte pSM214gCH. Um exemplo de uma construção de GAS 40 utilizando este vector é denotado como "HIS-40a-CH", SEQ ID N°s: 29 e 30. Nesta construção, a sequência líder do terminal N e a sequência transmembranar do terminal C são removidas e o marcador HIS é colocado no terminal C. Dois aminoácidos adicionais são também introduzidos no terminal amino. Três alterações de nucleótidos introduzidas com a clonagem são mostradas na sequência de ADN, com uma troca de aminoácido resultante indicado na sequência da proteína (a partir dos aminoácidos de F para S).

A optimização do códão também pode ser utilizada com estes vectores de clonagem alternativos. A construção de GAS 40 "HIS-40a-RR-NH" compreende a construção "HIS-40a-NH" com

três otimizações do códon. HIS-40a-RR-NH é apresentada na listagem de sequências como SEQ ID N°s: 31 e 32.

Deste modo, os antígenos GAS utilizados na presente invenção podem ser produzidos de forma recombinante utilizando construções de expressão que facilitam a sua produção recombinante. Modificações de sequência para facilitar a expressão preferidas podem ser selecionadas a partir do grupo que consiste de (1) a adição de uma sequência de marcador de purificação e (2) otimização de códon.

Como discutido acima, os requerentes identificaram GAS 40 como sendo particularmente adequado para utilização em composições imunogénicas, quer isoladamente ou em combinações. O uso de GAS 40 como um antígeno GAS particularmente eficaz é apoiado pela sua associação com a virulência, a sua localização superficial, a sua eficácia em ensaios opsonofagocitose bacterianos e em experiências de desafio de imunização. Além disso, a aquisição horizontal potencial deste factor de virulência indica que este antígeno pode ser específico para GAS (em relação a outras bactérias de estreptococos). Um apoio adicional para as propriedades antigénicas de GAS 40 também inclui a identificação de regiões de espiral enrolada dentro da estrutura bidimensional do GAS 40, e o baixo nível de homologia destas regiões com proteínas de superfície de outras bactérias estreptocócicas, incluindo algumas proteínas de adesão.

A análise da localização de GAS 40 dentro do genoma *Streptococcal pyogenes* feita pelos requerentes

indica que este fator de virulência foi provavelmente adquirido pelo GAS durante a evolução como um resultado de uma transferência horizontal de genes. A Figura 5 descreve GAS 40 dentro do genoma do GAS. É precedido na extremidade 5' por uma sequência designada "repressor do operação purina" ou "purR". É seguido na extremidade 3' por duas sequências que codificam para proteínas ribossomais designadas "proteína ribossômica S12", ou "rpsL" e "proteína ribossomal S7" ou "rpsG". (sequências de aminoácidos e polinucleotídicas para estes genes flanqueadores estão disponíveis ao público no GenBank (sequências PurR podem ser encontradas por exemplo na referência Genbank GI 15674250. As sequências RpsL podem ser encontrada por exemplo em referência Genbank GI 15674250. As sequências RpsG podem ser encontradas por exemplo sob a referência Genbank GI 15674250. Notavelmente, existem duas sequências de promotores putativos designados no início da sequência de rpsL. A Figura 5B retrata um mutante GAS onde uma grande parte do GAS 40 é eliminado. A única porção da sequência de GAS 40 restante corresponde aos polinucleótidos 1-97 da SEQ ID N°: 2. A deleção incluiu um dos promotores rpsL, deixando o segunda, P*, intacto. (As setas horizontais sublinhando o esquema indicar a região excluída).

A Figura 5C fornece detalhes adicionais sobre a sequência GAS de tipo selvagem. Aqui, são mostradas as sequências de repetição direta, designadas "RD", flanqueando as extremidades 5' e 3' de GAS 40 (As sequências correspondentes na deleção do 40 GAS mutante são identificadas na Figura 5D). Estas sequências repetidas diretas são cerca de 8 pares de bases. Um exemplo de um tal par de bases da repetição direta compreende a SEQ ID N°:

136. Tais motivos de sequência dentro de um genoma bacteriano indicam frequentemente uma transferência horizontal de genes. Experiências de infecção *in vivo* mostram que o mutante de deleção de 40 GAS é em vários registos menos virulento do que a estirpe de tipo selvagem. (Os pormenores desta experiência estão indicados no Exemplo 2).

A combinação da presença das sequências de repetição direta de flanqueamento e a virulência associado com GAS 40 sugere fortemente que a sequência GAS 40 foi adquirida horizontalmente por *Streptococcus pyogenes* durante a evolução. Nomeadamente, enquanto *Pur* e *rpsL* afins estão presentes em bactérias estreptocócicas relacionadas *Streptococcus agalactiae* e *Streptococcus mutants*, nenhuma destas bactérias são conhecidas por terem um GAS 40 homólogo. (a Figura 5E fornece esquematicamente a localização dos homólogos *purR*, *rpsL*, e *rpsG* dentro de *S. Agalactiae* (estreptococos do grupo B) e mostra a percentagem de homologia dos GBS homólogos com os homólogos GAS. Notavelmente, os genomas de GBS geralmente só possuem uma das sequências de repetição direta - e não contêm um par de sequências de repetição direta de flanqueamento da sequência GAS 40).

A localização da superfície de GAS 40 está ilustrada pelo diagrama de FACS apresentado na Figura 6. (Discussão dos protocolos relacionados com a análise de FACS é apresentada no Exemplo 1). A Figura 6 inclui diagramas FACS, tanto para o GAS do tipo selvagem (designado DSM 2071, um tipo M23 de GAS) e o mutante de deleção (designado DSM 2071 Δ 40). A mudança de absorvância para a estirpe de tipo selvagem indica que o GAS 40 é reconhecido na superfície da bactéria por anticorpos anti-GAS 40 (e que não é reconhecido na superfície do mutante de supressão).

A exposição da superfície de GAS 40 é adicionalmente demonstrada por um ensaio de opsonofagocitose bacteriana ilustrado na Figura 7 e no Exemplo 3. Neste ensaio, as estirpes de GAS são incubadas com soros pré-ímmes e ímmes, polímmorfonucleatos e complemento. (O soro ímmune é gerado por ímmunização de rato com a proteína GAS indicada). A fagocitose ou crescimento das bactérias são medidos de forma logarítmica. Barras de histograma positivos representam a fagocitose (ou morte bacteriana). Barras de histograma negativos representam o crescimento bacteriano. Como mostrado na Figura 7, os soros ímmes gerados por cada uma das proteínas expressas GAS40 resultou numa redução de bactérias (barras do histograma positivo).

Os estudos de desafio de ímmunização com GAS 40 são discutidos em detalhe no Exemplo 4. Tal como é mostrado no presente exemplo, o GAS 40, tal como produzido usando várias construções, oferece proteção substancial em ratinhos adultos. Notavelmente, a maioria das construções GAS40 fornece quase tanta proteção quanto a proteína GAS M. (a proteína GAS M é utilizada para comparação, uma vez que é conhecida por ser altamente ímmunogénica. No entanto, a proteína M não é geralmente considerada como uma candidata a vacina GAS adequada uma vez que varia amplamente entre as estirpes GAS e possui epítomos com potencial reatividade cruzada com os tecidos humanos). Além disso, um fragmento do terminal N de GAS 40 também proporcionou uma proteção significativa neste modelo. O fragmento do terminal N compreende cerca de 292 aminoácidos a partir do terminal N de GAS 40 que se sobrepõem com a primeira região em espiral enrolada. "40N-HIS" (SEQ ID N°s. 33 e 34) é um exemplo deste fragmento GAS 40, que compreende a região em espiral enrolada de GAS 40 e um marcador de HIS do terminal C.

(2) 117 GAS

GAS 117 corresponde ao número de acesso M1 GenBank GI: 13621679 e GI: 15674571, para o número de acesso M3 GenBank GI: 21909852, para número de acesso M18 GenBank GI: 19745578, e também é conhecido como 'Spy0448' (M1), 'SpyM3_0316' (M3), e 'SpyM18_0491' (M18). Exemplos de sequências de aminoácidos e de polinucleótidos de GAS 117 de uma estirpe M1 estão apresentados na listagem de sequências como SEQ ID N°s: 35 e 36.

As proteínas GAS 117 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, de 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91% , 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 35; e/ ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos de SEQ ID N°: 35, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 ou mais). Estas proteínas GAS 117 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) de SEQ ID N°: 35. Fragmentos preferidos de (b) compreender um epítopo a partir da SEQ ID N°: 1. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 35. Por exemplo, a sequência de aminoácidos sublinhada no terminal N da SEQ ID N°: 35 (ver abaixo) pode ser removida. (SEQ ID N°: 37 compreende a sequência de aminoácidos do terminal N removida. A SEQ ID N°: 38 compreende um fragmento de 117 GAS sem a sequência do

terminal N do aminoácido). Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, dum domínio transmembranar, ou de um domínio extracelular).

SEQ ID NO: 35

MTLKKHYLLSLLALVTGAAFNSTQSVSAQVYSNEGYPHQLTDEKSHLQYSKDNAQLQLRNILDGYPNDLGRHYSSYY
YNLRTVMGLSSEQDIEKHYYEELKNKLHDMYNHY

(3) GAS 130

GAS 130 corresponde ao número de acesso M1 GenBank GI: 13621794 e GI: 15674677, para o número de acesso M3 GenBank GI: 21909954, para número de acesso M18 GenBank GI: 19745704, e também é conhecido como 'Spy0591' (M1), 'SpyM3_0418' (M3), e 'SpyM18_0660' (M18). GAS 130 foi identificado como potencialmente uma protease putativa. Exemplos de sequências de aminoácidos e de polinucleótidos de GAS 130 de uma estirpe M1 estão apresentados na listagem de sequências como SEQ ID N°s: 39 e 40.

As proteínas GAS 130 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 39; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 39, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, ou mais). Estas proteínas GAS 130 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) de SEQ ID N°: 39. Fragmentos preferidos de (b) compreendem um epítopo a partir da SEQ ID N°: 39. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4,

5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 39. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(4) GAS 277

GAS 277 corresponde aos números de acesso M1 GenBank GI: 13622962 e GI: 15675742, ao número de acesso M3 GenBank GI: 21911206, ao número de acesso M18 GenBank GI: 19746852, e é também referido como Spy1939 `(M1), 'SpyM3_1670' (M3), e 'SpyM18_2006' (M18). Sequências de aminoácidos e de polinucleótidos de GAS 277 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 41 e 42.

As proteínas GAS 277 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 41; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 41, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, ou mais). Estas proteínas GAS 277 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 41. Fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 41. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2,

3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 41 Por exemplo, a sequência de aminoácidos sublinhada na extremidade do terminal N da SEQ ID N°: 41 (ver abaixo) pode ser removida. (SEQ ID N°: 43 compreende o aminoácido do terminal N sublinhado. A SEQ ID N°: 44 compreende um fragmento de GAS 277 com a sequência de aminoácidos do terminal N removida). Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

SEQ ID NO: 41

MTTQKTIISLLSLALLIGLGTSGKAI SVYAQDQHTDNVIAESTISQVSVEASMRGTEFYIDATVITDQPVROPTQATIT
LKDASDNTINSWVYTMAAQRRFTAWFDLTGQKSGDYHVTVTVHTQEKAVTGQSGTVHFDQNRARKEPTTNMQQKDTSKAM
TNSVDVDTKAQTNQSANQEIDSTSNPFRSATNHRSTSLKRSTKNEKLTPASNSQKNGSNXTKMLVDKKEEVKPTSRRGFP
WVLLGLVLSLAAGLFIAIQKVSRRK

(5) GAS 236

GAS 236 corresponde ao número de acesso M1 GenBank GI: 13622264 e GI: 15675106, ao número de acesso M3 GenBank GI: 21910321, e ao número de acesso M18 GenBank GI: 19746075, e também é conhecido como 'Spy1126' (M1), 'SpyM3_0785' (M3), e 'SpyM18_1087' (M18). As sequências de aminoácidos e de polinucleótidos de GAS 236 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 45 e 46.

As proteínas GAS 236 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 45; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 45, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90,

100, 150 ou mais). Estas proteínas GAS 236 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 45. Fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 45. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 45. Por exemplo, a sequência de aminoácidos sublinhada na extremidade do terminal N da SEQ ID N°: 45 (ver abaixo) pode ser removida. (A SEQ ID N°: 47 compreende o terminal N da sequência de aminoácido. A SEQ ID N°: 48 compreende um fragmento de GAS 236 com a sequência do terminal N removida). Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

SEQ ID NO: 45

MTQMNYTGKVRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGMLLSAFHMVEKELDKVRFVGIHT
GHLGFYTDYRDFEVDKLIIDNLKDKGEQISYPILKVAITLDEGRVVKARALNEATVKRIEKTADVIIINHVKFESFRGD

GISVSTPTGSTAYNKSLGGAVLHPTIEALQLTEISSLNRRVFTLGSSIIIPKKDKIELVPKRLGIYTTISIDNKTYQLKN
 VTKVEYFIDDERIHVFSSPSHTSFWERVKDAFIGEIDS

(6) GAS 389

GAS 389 corresponde ao número de acesso M1 GenBank GI: 13622996 e GI: 15675772, para o número de acesso M3 GenBank GI: 21911237, para o número de acesso M18 GenBank GI: 19746884, e também é conhecido como 'Spy1981' (M1), 'SpyM3_1701' (M3), 'SpyM18_2045' (M18) e 'relA'. GAS 389 também foi identificado como uma sintetase (p) ppGpp. Sequências de aminoácidos e de polinucleótidos de GAS 389 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 49 e 50.

As proteínas GAS 389 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 49; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos de SEQ ID N°: 49, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 ou mais). Estas proteínas GAS 389 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 49. Fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 49. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 49. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(7) GAS 504

GAS 504 corresponde aos números de acesso M1 GenBank GI: 13622806 e GI: 15675600, ao número de acesso M3 GenBank GI: 21911061, ao número de acesso M18 GenBank GI: 19746708, e é também referido como 'Spy1751' (M1), 'SpyM3_1525', 'SpyM18_1823' (M18) e 'FABK'. GAS 504 também foi identificado como uma reductase II de trans-2-enoil-ACP putativo. Sequências de aminoácidos e de polinucleótidos de GAS 504 de uma estirpe M1 são apresentadas abaixo e na listagem de sequências como SEQ ID N°S: 51 e 52.

As proteínas GAS 504 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 51; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 51, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 ou mais). Estas proteínas GAS 504 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 51. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 51. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 51. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(8) GAS 509

GAS 509 corresponde ao número de acesso M1 GenBank GI: 13622692 e GI: 15675496, para o número de acesso M3 GenBank GI: 21910899, para o número de acesso M18 GenBank GI: 19746544, e também é conhecido como 'Spy1618' (M1), 'SpyM3_1363' (M3), 'SpyM18_1627' (M18) e 'cysM'. GAS 509 também tem sido identificado como uma liase O-acetilserina putativa. Sequências de aminoácidos e de polinucleótidos de GAS 509 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 53 e 54.

As proteínas GAS 509 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 53; e/ou (b) que é um fragmento de aminoácidos, pelo menos, *n* consecutivos da SEQ ID N°: 53, em que *n* é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, ou mais). Estas proteínas GAS 509 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 53. Fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 53. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 53. Por exemplo, a sequência de aminoácidos sublinhada na extremidade do terminal C da SEQ ID N°: 53 (ver abaixo) pode ser removida. (A SEQ ID N°: 55 compreende a sequência de aminoácidos do terminal C. A SEQ ID N°: 56 compreende um fragmento de GAS 509 com a sequência do terminal C removida. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

SEQ ID NO: 53

MTKIYKTITELVGQTPIIKLNRLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAEGLISPGDVIIE
PTSGNTGIGLAWVGAAGYRVIIIVMPETMSLERRQIIQAYGAELVLTPGAEGMKGAIAKAETLAIELGAW
MPMQFNPNPANPSIHEKTTAQEILEAFKEISLDAFVSGVGTGGTILSGVSHVLKKANPETVIYAVEAEESAV
LSGQEPGPHKIQGISAGFIPTLDTKAYDQIIIRVKS KDALETARLTGAKEGFLVGISSGAALYAAIEVAK
QLGKGKHAVLTILPDNGERYLSTELYDVPVIKTK

(9) GAS 366

GAS 366 corresponde ao número de acesso M1 GenBank GI: 13622612, GI: 15675424 e GI: 30315979, para o número de acesso M3 GenBank GI: 21910712, para o número de acesso M18 GenBank GI: 19746474, e também é conhecido como 'Spy1525 "(M1), 'SpyM3_1176 "(M3), 'SpyM18_1542" (M18) e 'MuRD'. GAS 366 também foi identificado como uma ligase de UDP-N-acetilemuramoilalanina-D-glutamato ou uma enzima adicionando ácido D-glutâmico. Sequências de aminoácidos e de polinucleótidos de GAS 366 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 57 e 58.

As proteínas GAS 366 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 57; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos da SEQ ID N°: 57, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 ou mais). Estas proteínas GAS 366 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 57. Os fragmentos preferidos de (b) compreendem um epítopo a partir da SEQ ID N°: 57. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 57. Por exemplo, a sequência de aminoácidos sublinhada na extremidade do terminal N da SEQ ID N°: 57 (ver abaixo) pode ser removida. (A SEQ ID N°: 59 compreende a sequência

líder do terminal N. A SEQ ID N°: 60 compreende um fragmento de GAS 366 onde a sequência do terminal N é removida). Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

SEQ ID NO: 57

MRVISNFGNKKILILGLAKSGEAAAKLITLGLALVTVNDSKPFQNPAAQALLEEGIKVICGSHPVVELLDENFEYMKVKNP
GIPYDNPMVKRALAKEIPILTEVELAYFVSEAPIIGITGSNGKTTTTMTIADVLNACGQSALLSGNIGYPASKVVQKAIA

GDTLVMELSSPQLVGVNAFRPHIAVITNLMPTHLDYHGSFEDYVAAKWKMIQAQMTESDYLIILNANQEI SATLAKTTRATV
IPFSTQKVVDGAYLKDGIYFKEQAI IAAATDLGVPGSHNIENALATIAVAKLSGIADDI IAOCLSHFGGVKHLRQVRVQGI
KDITFYNDKSTNIIATQKALSGFDNSRLI IAGGLDRGNEFDLVPDLLGLKQMI ILGESAERMKRAANKAEVSYLEAR
NVAEATELAFKLAQTGDTILLSPANASWDMYPNFVEVRGDEFATFDCLRGDA

(10) GAS 159

GAS 159 corresponde ao número de acesso M1 GenBank GI: 13622244 e GI: 15675088, para o número de acesso M3 GenBank GI: 21910303, para o número de acesso M18 GenBank GI: 19746056, e também é conhecido como 'Spy1105' (M1), 'SpyM3_0767' (M3), 'SpyM18_1067' (M18) e 'POTD'. GAS 159 também tem sido identificado como um transportador ABC de espermidina/putrescina putativo (uma proteína de transporte periplásmica). Sequências de aminoácidos e de polinucleótidos de GAS 159 de uma estirpe M1 são apresentadas abaixo e na listagem de sequências como SEQ ID N°s: 61 e 62.

SEQ ID NO: 61

MRRLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLV IYNWGDYIDFALLKKFTKETGIEVQYETFDSNEAMYTKIKQ
GGTTYDIAVPSDYTIKKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDFQNDYSLPYFWGTVGIVYNDQLVDKAPMHWED
LWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNVKAIVADEMRGYMIQGDAAGITFSGEA
SEMLEDNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFNLNRPENAAQNAAYIGYATPNKKAKALLPDEIKNDPA
FYPTDDIIRKLEVYDNLGSRWLGIYNDLYLQFKMYRK

SEQ ID NO: 62

ATGCGTAAACTTTTATTCCTTTCTAGCAGGAGTTTGGGGTGTATTGTTATTTTAAACAAGTCCTTCTTTCATCTTGCAGAA
AAAATCGGGTTCTGGTAGTCAATCGGATAAAATTAGTTATTTATAACTGGGGAGATTACATGATCCAGCTTTGCTCAAAA
AATTACCCAAAGAACCGGGCATTGAAGTGCAGTATGAAACTTTTCGATTCCAATGAAGCCATGTACACTAAAATCAAGCAG
GGCGGAACCACTTACGACATTGCTCTTCTAGTGAATTACACCATTCATATAAATCATCAAGAAAAACCTACTCAATAAGCT
TGATAAGTCAAAATTAGTTGGCATGGATAATATCGGGAAAGAATTTTATAGGGAAAAGCTTTGACCCACAAAACGACTATT
CTTTGCCCTTATTTCTGGGGAAACCGTTGGGATTGTTTATAATGATCAATTAGTTGATAAGGGCCCTATGCACTGGGAAGAT
CTGTGGCGTCCAGAAATATAAAAATAGTATTATCTGATTTGATGAGACCGCGTGAATGCTAGGGGTTGGTTTAACAACTTT
TGGTTATAGTGTGAATTCTAAAAATCTAGAGCAGTTGCAGGCAGCCGAGAGAAAATGCGAGCAGTTGACGCGGAATGTTA
AAGCCATTGTAGCAGATGAGATGAAAGGCTACATGATTCAAGGTGACGCTGCTATTGGAATTACCTTTTCTGGTGAAGCC
AGTGAGATGTTAGATAGTAACGAACACCTTCACTACATCGTGCCTTCAGAAGGGTCTAACCTTTGGTTTGATAATTTGGT
ACTACCAAAAACCATGAABACACGAAAAAGAAGCTTATGCTTTTTTGAACCTTATCAATCGTCCTGAAAATGCTGGGCAAA
ATGCTGCATATATTGGTTATGCGACACCAAAATAAATAAGCCAAAGGCTTACTTCCAGATGAGATAAATAATGATCTGCT
TTTATCCAAACAGATGACATTATCAAAAAATTGGAAGTTTATGACAATTTAGGGTCAAGATGGGTTGGGGAATTATAATGA
TTTATACCCTCCAAATTAAAAATGATATCGCAATAA

As proteínas GAS 159 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 61; e/ou (b) que é um fragmento de aminoácidos, pelo menos, *n* consecutivos da SEQ ID N°: 61, em que *n* é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 ou mais). Estas proteínas GAS 159 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..), da SEQ ID N°: 61. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 61. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 61. Por exemplo, a sequência de aminoácidos sublinhada na extremidade do terminal N da SEQ ID N°: 61 (ver abaixo) pode ser removida. (A SEQ ID N°: 63 compreende a sequência líder do aminoácido do terminal N. A SEQ ID N°: 64 compreende um fragmento de GAS 159 onde a sequência de aminoácidos líder do terminal N é removida). Noutro exemplo, a sequência de aminoácidos sublinhada na extremidade do terminal C da SEQ ID N°: 61 é removida. (A SEQ ID N°: 65 compreende a região hidrófoba do terminal C.

A SEQ ID N°: 66 compreende um fragmento de GAS 159 onde a região hidrófoba do terminal C foi removida. A SEQ ID N°: 67 compreende um fragmento de GAS 159 onde tanto a sequência líder do terminal N e a região hidrófoba do terminal C são removidas). Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

SEQ ID NO: 61

```
MRKLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLVYINWGDYIDPALLKFTKETGIEVQYETFDSENEAMYTKIKQ
GGTTYDIAVPSDYTIKMIKENLLNKLDKSKLVGMNDNIGKEFLGKSFDPONDYSLPYFWGTGIVYNDQLVDKAPMHWED
LWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLOQLTPNVKAIVADEMKGYMIQGLAAIGITFSGEA
SEMLDSEHLHYIVPSEGSNLWFDNLVLPKTMKEKEAYAFNFINRPENAAQNAAYIGYATPNKKAKALLPDEIKNDPA
FYPTDDI IKKLEVDNLGSRWLGIYNDLYLQFKMYRK
```

(11) GAS 217

GAS 217 corresponde o número de acesso M1 GenBank GI: 13622089 e GI: 15674945, para o número de acesso M3 GenBank GI: 21910174, para o número de acesso M18 GenBank GI: 19745987, e também é conhecido como 'Spy0925' (M1), 'SpyM3_0638' (M3), e 'SpyM18_0982' (M18). GAS 217 também foi identificada como uma oxidoreductase putativa. Sequências de aminoácidos e de polinucleótidos de GAS 217 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 68 e 69.

As proteínas GAS 217 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 68; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos de SEQ ID N°: 68, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, ou mais). Estas proteínas GAS 217 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos,

parálogos, mutantes, etc..) da SEQ ID N°: 68. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 68. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 68. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(12) GAS 309

GAS 309 corresponde ao número de acesso M1 GenBank GI: 13621426 e GI: 15674341, para o número de acesso M3 GenBank GI: 21909633, para o número de acesso M18 GenBank GI: 19745363, e também é conhecido como 'Spy0124' (M1), 'SpyM3_0097' (M3), 'SpyM18_0205' (M18), 'nra' e 'rofA'. GAS 309 também tem sido identificado como uma proteína de regulação e um regulador da transcrição negativo. Sequências de aminoácidos e de polinucleótidos de GAS 309 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 70 e 71.

As proteínas GAS 309 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 70; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 70, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, ou mais). Estas proteínas GAS 309 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos,

parálogos, mutantes, etc..) da SEQ ID N°: 70. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 70. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 70. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(13) GAS 372

GAS 372 corresponde ao número de acesso M1 GenBank GI: 13622698 e GI: 15675501, para o número de acesso M3 GenBank GI: 21910905, para o número de acesso M18 GenBank GI: 19746500 e também é conhecido como 'Spy1625' (M1) 'SpyM3_1369' (M3), e 'SpyM18_1634' (M18). GAS 372 também tem sido identificado como uma proteína quinase putativa ou uma quinase do tipo eucariótico de serina/treonina putativa. Sequências de aminoácidos e de polinucleótidos de GAS 372 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 72 e 73.

As proteínas GAS 372 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 72; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 72, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 ou mais). Estas proteínas GAS 372 incluem variantes (por exemplo, variantes alélicas, homólogos,

ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 72. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 72. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 72. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(14) GAS 039

GAS 039 corresponde ao número de acesso M1 GenBank GI: 13621542 e GI: 15674446, para o número de acesso M3 GenBank GI: 21909730, para o número de acesso M18 GenBank GI: 19745398 e também é conhecido como 'Spy0266' (M1), 'SpyM3_0194' (M3), e 'SpyM18_0250' (M18). Sequências de aminoácidos e de polinucleótidos de GAS 039 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 74 e 75.

As proteínas GAS 039 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 74; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 74, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, ou mais). Estas proteínas GAS 039 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 74. Os fragmentos preferidos de (b) compreendem um epítipo a

partir da SEQ ID N°: 74. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 74. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(15) GAS 042

GAS 042 corresponde ao número de acesso M1 GenBank GI: 13621559 e GI: 15674461, para o número de acesso M3 GenBank GI: 21909745, para o número de acesso M18 GenBank GI: 19745415, e também é conhecido como 'Spy0287' (M1), 'SpyM3_0209' (M3), e 'SpyM18_0275' (M18). Sequências de aminoácidos e de polinucleótidos de GAS 042 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 76 e 77.

As proteínas GAS 042 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 76; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 76, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, ou mais). Estas proteínas GAS 042 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 76. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 76. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4,

5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 76. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(16) GAS 058

GAS 058 corresponde ao número de acesso M1 GenBank GI: 13621663 e GI: 15674556, para o número de acesso M3 GenBank GI: 21909841, para o número de acesso M18 GenBank GI: 19745567 e também é conhecido como 'Spy0430' (M1), 'SpyM3_0305' (M3), e 'SpyM18_0477' (M18). Sequências de aminoácidos e de polinucleótidos de GAS 058 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 78 e 79.

As proteínas GAS 058 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 78; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos da SEQ ID N°: 78, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, ou mais). Estas proteínas GAS 058 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 78. Os fragmentos preferidos de (b) compreendem um epítopo a partir da SEQ ID N°: 78. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7,

8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 78. Por exemplo, a sequência de aminoácidos sublinhada na extremidade do terminal N da SEQ ID N°: 78 (ver abaixo) pode ser removida. (A SEQ ID N°: 80 compreende a sequência líder do terminal N. A SEQ ID N°: 81 compreende um fragmento de GAS 58, onde a sequência líder do terminal N é removida). Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, de omissão um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(17) GAS 290

GAS 290 corresponde aos números de acesso M1 GenBank GI: 13622978 e GI: 15675757, a número de acesso M3 GenBank GI: 21911221, a número de acesso M18 GenBank GI: 19746869 e é também referido como 'Spy1959' (M1), 'SpyM3_1685' (M3), e 'SpyM18_2026' (M18). Sequências de aminoácidos e de polinucleótidos de GAS 290 de uma estirpe M1 são apresentados na listagem de sequências como SEQ ID N°s: 82 e 83.

As proteínas GAS 290 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 82; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos de SEQ ID N°: 82, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 ou mais). Estas proteínas GAS 290 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 82. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou

mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 84. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(18) GAS 511

GAS 511 corresponde aos números de acesso M1 GenBank GI: 13622798 e GI: 15675592, ao número de acesso M3 GenBank GI: 21911053, ao número de acesso M18 GenBank GI: 19746700 e é também referido como 'Spy1743' (M1), 'SpyM3_1517' (M3), 'SpyM18_1815' (M18) e "accA". Sequências de aminoácidos e de polinucleótidos de GAS 511 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 84 e 85.

As proteínas GAS 511 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 84; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos de SEQ ID N°: 84, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 ou mais). Estas proteínas GAS 511 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 84. Os fragmentos preferidos de (b) compreender um epítopo a partir da SEQ ID N°: 84. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°:

84. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(19) GAS 533

GAS 533 corresponde aos números de acesso M1 GenBank GI: 13622912 e GI: 15675696, ao número de acesso M3 GenBank GI: 21911157, ao número de acesso M18 GenBank GI: 19746804 e é também referido como 'Spy1877' (M1), 'SpyM3_1621' (M3), 'SpyM18_1942' (M18) e 'glnA'. GAS 533 também foi identificada como uma sintetase de glutamina putativa. Sequências de aminoácidos e de polinucleótidos de GAS 533 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 86 e 87.

As proteínas GAS 533 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 86; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 86, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 533 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) de SEQ ID N°: 86. fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 86. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 86. Outros fragmentos omitem um ou mais domínios da

proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(20) GAS 527

GAS 527 corresponde ao número de acesso M1 GenBank GI: 13622332, GI: 15675169, e GI: 24211764, para o número de acesso M3 GenBank GI: 21910381, para número de acesso M18 GenBank GI: 19746136, e também é conhecido como 'Spy1204' (M1), 'SpyM3_0845' (M3), 'SpyM18_1155' (M18) e 'guaA'. GAS 527 também foi identificado como uma sintetase GMP putativa (glutamato hidrolisado) (glutamato de amidotransferase). Sequências de aminoácidos e de polinucleótidos de GAS 527 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 88 e 89.

As proteínas GAS 527 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 88; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 88, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 527 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 88. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 88. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°:

88. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(21) GAS 294

GAS 294 corresponde ao número de acesso M1 GenBank GI: 13622306, GI: 15675145, e GI: 26006773, para o número de acesso M3 GenBank GI: 21910357, para o número de acesso M18 GenBank GI: 19746111 e também é conhecido como 'Spy1173' (M1), 'SpyM3_0821' (M3), 'SpyM18_1125' (M18) e 'gid'. GAS 294 também tem sido identificada como uma proteína de divisão inibida por glicose putativa. Sequências de aminoácidos e de polinucleótidos de GAS 294 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 90 e 91.

As proteínas GAS 294 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 90; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 90, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 294 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 90. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID NO: 90. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°:

90. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(22) GAS 253

GAS 253 corresponde o número de acesso M1 GenBank GI: 13622611, GI: 15675423, e GI: 21362716, para o número de acesso M3 GenBank GI: 21910711, para o número de acesso M18 GenBank GI: 19746473 e também é conhecido como 'Spy1524' (M1), 'SpyM3_1175' (M3), 'SpyM18_1541' (M18) e 'murG'. GAS 253 também tem sido identificado como uma transferase undecaprenil-PP-MurNAc-pentaptido-UDPGlcNAc GlcNAc putativa. Sequências de aminoácidos e de polinucleótidos de GAS 253 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 92 e 93.

As proteínas GAS 253 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 92; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 92, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 253 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 92. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 92. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°:

92. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(23) GAS 529

GAS 529 corresponde ao número de acesso M1 GenBank GI: 13622403, GI: 15675233, e GI: 21759132, para o número de acesso M3 GenBank GI: 21910446, para o número de acesso M18 GenBank GI: 19746203 e também é conhecido como 'Spy1280' (M1), 'SpyM3_0910' (M3), 'SpyM18_1228' (M18) e "glmS". GAS 529 também tem sido identificado como uma aminotransferase de L-glutamina-D-frutose-6-fosfato putativa (sintase de glucosamina-6-fosfato). Sequências de aminoácidos e de polinucleótidos de GAS 529 de uma estirpe M1 são apresentadas abaixo e na listagem de sequências como SEQ ID N°s: 94 e 95.

As proteínas GAS 529 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 94; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos da SEQ ID N°: 94, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 529 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 94. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 94. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7,

8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 94. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(24) GAS 045

GAS 045 corresponde ao número de acesso M3 GenBank GI: 21909751, ao número de acesso M18 GenBank GI: 19745421 e é referido como 'SpyM3_0215' (M3), 'SpyM18_oppA' (M18) e 'oppA'. GAS 045 foi identificado como uma permease de oligopeptídeo. Sequências de aminoácidos e de polinucleótidos de GAS 045 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 96 e 97.

As proteínas GAS 045 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 96; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos da SEQ ID N°: 96, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 045 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.), da SEQ ID N°: 96. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 96. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 96. Por exemplo, a sequência de aminoácidos sublinhada na

extremidade do terminal N da SEQ ID N°: 96 (ver abaixo) pode ser removida. (A SEQ ID N°: 98 compreende a sequência líder do terminal N sublinhado. A SEQ ID N°: 99 compreende um fragmento de GAS 45, onde a sequência líder do terminal N é removida). Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

SEQ ID NO: 96

VTFMKKSKNLAASVAILSVSALAACGNKNASGGSEATKTYKYVFVNDPKSLDYILTNGGGTTDVITQMVDGILENDEYG
 NLVPSLAKEDWKVSKOGLTYTYTLRGVSWYTADGEEYAEVTAEDFVTGLKHAVDDKSLALYVVEDSIKNLKAYQNGEVDF
 KEVGVKALEDDKTVQYTLNKPESYWNKSTTYSVLFFVNAKFLKSRGKDFGTTDPSSILVNGAYFLSAFTSKSSMEFHKNE
 YNDAKNVGIESVKLTYSOGSDPGSFYKNFEDKGETSVARLYPNDPTYESAKKNYADNITYGMLTGDI RHLTWNLNRTSERN
 TKKDPAQQDAGKKALNNKDFRQAIQFAFDRASFQAQTAQDAKTALRNMLVPPTFVTIGESDFGSEVEKEMAKLGDEWK
 DVNLADAQDGFYNPEKAKAEFAKAREALTAEGVTFPVQLDYPVDQANAATVQEAQSEKQSVBASLGKENVIVNVLETETS
 THEAQGFYAEETPEQQDYDI LSSWNGFPYQDPRTYLRIMSFPVGGGSVIQKLGIKAGCNKDNVAAAGLDTYQTLLEAAAIT
 DDNDARYKAYAKAQAYLTDNAVDFPVVALGCTFRVTKAVPFSGGFSWAGSEKGLAYKGMKLQDKPVTVKQYEKAREKWMK
 AKAKSNAKYAEKLDHVEK

(25) GAS 095

GAS 095 corresponde aos números de acesso M1 GenBank GI: 13622787 e GI: 15675582, ao número de acesso M3 GenBank GI: 21911042, ao número de acesso M18 GenBank GI: 19746634 e é também referido como 'Spy1733' (M1), 'SpyM3_1506' (M3), 'SpyM18_1741' (M18). GAS 095 também tem sido identificado como um regulador da transcrição putativo. Sequências de aminoácidos e de polinucleótidos de GAS 095 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 100 e 101.

As proteínas GAS 095 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 100; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 100, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35,

40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 095 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.), da SEQ ID N°: 100. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 100. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 100, por exemplo, a sequência de aminoácidos sublinhada na extremidade do terminal N da SEQ ID N°: 100 (mostrado abaixo) pode ser removida. (A SEQ ID N°: 102 compreende a sequência de aminoácidos da sequência líder do terminal N sublinhado. A SEQ ID N°: 103 compreende um fragmento de GAS 95, onde a sequência líder do terminal N é removida.) Outros fragmentos omitem um ou mais domínios de a proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio de membrana ou de um domínio extracelular).

SEQ ID NO: 100

MKIGKIVLMFTAIVLTTLVLAIGVYLTSAYTFSTGELSKTFNDFSTSSNKSDAIKQTRAFSILLMGVDTGSSERASKWEG
 NSDSMILVTVPETKTKTMTSLERDTLTLSGPFNNEMNGVEAKLNAAAYAAGGAQMAIMTVQDILLNITIDNYVQINMQGL
 IDLVNAVGGITVTNEFDFFISIAENEPEYQATVAPGTHKINGEQALVYARMRYDDFEGDYGRQKRQREVIQKVLKKILAL
 DSISYRKILSAVSSNMQTNIEISSRTIPSLLGXYDALRTIKTYQLKGEDATLSGGGSYQIVTSNHLLEIQNRIRTELGL
 HKVNLKTNATVYENLYGSTRSQTVNNYDSSGQAFSYSDSHSSYANYSSGVDTGQSASTDQDSTASSHREATPSSSSDA
 LAADESSSSSGSGSLVFPANINPQT

(26) GAS 193

GAS 193 corresponde ao número de acesso M1 GenBank GI: 13623029 e GI: 15675802, para o número de acesso M3 GenBank GI: 21911267, para o número de acesso M18 GenBank GI: 19746914 e também é conhecido como 'Spy2025' (M1), 'SpyM3_1731' (M3), 'SpyM18_2082' (M18) e 'isp'. GAS 193 também tem sido identificado como um precursor da proteína segregada imunogénica. Sequências de aminoácidos e de

polinucleótidos de GAS 193 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 104 e 105.

As proteínas GAS 193 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 104; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos de SEQ ID N°: 104, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 193 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc..) da SEQ ID N°: 104. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 104. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 104. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(27) GAS 137

GAS 137 corresponde ao número de acesso M1 GenBank GI: 13621842, GI: 15674720 e GI: 30173478, para o número de acesso M3 GenBank GI: 21909998, para o número de acesso M18 GenBank GI: 19745749 e também é conhecido como 'Spy0652' (M1), 'SpyM3_0462', e 'SpyM18_0713' (M18). Sequências de aminoácidos e de polinucleótidos de GAS 137 de uma estirpe

M1 são apresentadas na listagem de sequências como SEQ ID N°s: 106 e 107.

As proteínas GAS 137 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 106; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos de SEQ ID NO: 106, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 137 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.), da SEQ ID N°: 106. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 106. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 106. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(28) GAS 084

GAS 084 corresponde aos números de acesso M1 GenBank GI: 13622398 e GI: 15675229, ao número de acesso M3 GenBank GI: 21910442, ao número de acesso M18 GenBank GI: 19746199 e é também referido como 'Spy1274' (M1), 'SpyM3_0906' e 'SpyM18_1223' (M18). GAS 084 também tem sido identificado como um aminoácido transportador ABC putativo/aminoácido periplasmático da proteína de ligação. Sequências de aminoácidos e de polinucleótidos de GAS 084 de uma estirpe

M1 são apresentadas na listagem de sequências como SEQ ID N°s: 108 e 109.

As proteínas GAS 084 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 108; e/ou (b) que é um fragmento de aminoácidos, pelo menos, *n* consecutivos da SEQ ID N°: 108, em que *n* é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 084 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.), da SEQ ID N°: 108. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 108. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 108, por exemplo, a sequência de aminoácidos sublinhada na extremidade do terminal N da SEQ ID N°: 108 (mostrado abaixo) pode ser removida. (A SEQ ID N°: 110, compreende uma sequência de aminoácidos compreendendo a sequência líder do terminal N sublinhado de GAS 84. A SEQ ID N°: 111 compreende um fragmento de GAS 84, onde a sequência líder do terminal N é removida). Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

SEQ ID NO: 108

MIKKRTVAILAIASSFFLVACQATKSLKSGDAWGVIQKQKSITVGFDNTEFVPMGYKDESGRCKGFDIDLAKVEVHQYGL
KVNFQAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAFTDSYMRNEQIIVVKRSDIKTISDMKHVILGAQSASSGYD
SLLRTPKLLKDFIKNKDANQYETFTQAFIDLKSDRIDGILIDKVYANYLAKEGQLENYRMIPTTFENEAFSVGLRKEDK
TLQAKINRAFRVLYQNGKFQAISEKWFEGDDVATANIKS

(29) GAS 384

GAS 384 corresponde ao número de acesso M1 GenBank GI: 13622908 e GI: 15675693, para o número de acesso M3 GenBank GI: 21911154, para o número de acesso M18 GenBank GI: 19746801 e também é conhecido como 'Spy1874' (M1), 'SpyM3_1618' (M3), e 'SpyM18_1939' (M18). GAS 384 também foi identificado como uma glicoproteína de endopeptidase putativa. Sequências de aminoácidos e de polinucleótidos de GAS 384 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 112 e 113.

As proteínas GAS 384 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 112; e/ou (b) que é um fragmento de aminoácidos, pelo menos, n consecutivos da SEQ ID N°: 112, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 384 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.) da SEQ ID N°: 112. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 112. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 112. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(30) GAS 202

GAS 202 corresponde o número de acesso M1 GenBank GI: 13622431 e GI: 15675258, para o número de acesso M3 GenBank GI: 21910527, para o número de acesso M18 GenBank GI: 19746290 e também é conhecido como 'Spy1309' (M1), 'SpyM3_0991' (M3), 'SpyM18_1321' (M18) e 'dltD'. GAS 202 também tem sido identificado como uma proteína extramembranal putativa. Sequências de aminoácidos e de polinucleótidos de GAS 202 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 114 e 115.

As proteínas GAS 202 preferidas para serem usadas com a divulgação compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 114; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos de SEQ ID N°: 114, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 ou mais). Estas proteínas GAS 202 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.) da SEQ ID N°: 114. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 114. Outros fragmentos preferidos carecem um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 114. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(31) GAS 057

GAS 057 corresponde ao número de acesso M1 GenBank GI: 13621655 e GI: 15674549, para o número de acesso M3 GenBank GI: 21909834, para o número de acesso M18 GenBank GI: 19745560 e também é conhecido como 'Spy0416' (M1), 'SpyM3_0298' (M3), 'SpyM18_0464' (M18) e "prtS". GAS 057 também tem sido identificado como uma proteinase de envelope celular putativa. Sequências de aminoácidos e de polinucleótidos de GAS 057 de uma estirpe M1 são apresentadas na listagem de sequências como SEQ ID N°s: 116 e 117.

As proteínas GAS 057 preferidas para utilização com a invenção compreendem uma sequência de aminoácidos: (a) possuindo 95% ou mais de identidade (por exemplo, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 116; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos de SEQ ID N°: 116, em que n é de 200 ou mais. Estas proteínas GAS 057 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.) da SEQ ID N°: 116. Os fragmentos preferidos de (b) compreendem um epítopo a partir da SEQ ID N°: 116. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 116, por exemplo, numa forma de realização, a sequência de aminoácido sublinhado na extremidade do terminal N da SEQ ID N°: 116 (mostrado abaixo) é removida. (A SEQ ID N°: 118 compreende a sequência líder do terminal N sublinhado. A SEQ ID NO: 119. compreende um fragmento de GAS 57, onde a sequência líder do terminal N foi removida.) Num outro exemplo, a sequência de aminoácidos sublinhada no terminal C da SEQ ID N°: 116 é

removida. (A SEQ ID N°: 120 compreende a região hidrófoba do terminal C sublinhado. A SEQ ID N°: 121 compreende um fragmento de GAS 57 onde a região hidrófoba do terminal C é removida. A SEQ ID N°: 122 compreende um fragmento de GAS 57 onde tanto a sequência líder do terminal N como a região hidrófoba do terminal C são removidas). Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

SEQ ID NO: 116

MEKKQRFSLRKYKSGTFSVLIGSVFLVMTTVADELSTMSEPTITNHAQQQAQHLTNTELSSAESKRSQDTSQITLKTNR
EKEQSQDLVSEPTTTELAETDAASMANTGSSDATQKSASLPPVNTDVRDWVKTGAWDKGYKGGQKVVAVIDTGI DPAHQ3
MRI SDVSTAKVKSKE DMLARQKAAGINYG SWINDKVVF AHNYVENS DNIKENQFEDFEDWENFEFDAEAEFKA I KKKHI
YRPQSTQAPKETVIRKTEETDGS HDI EWTQT DDDTKYESHGMHVTGIVAGNSKEAAATGERFLGIAPFAQVMEMRVFANDI
MGSAESLFIKAIEDAVALGADV INLSLGTANGAQLSGSKPLMEAI EKAKKAGVSVVVAAGNERVYGS SDHDDPLATNPDYG
LVGSPSTGRPTPSVAAINS KRWIQR LMTVKELENRADLNHGKAIYSESVDPKDIKDSLGYDKSHQFAYVKEST DAGYNAQ
DVEGKIALIERDPNKTYDEMIALAKKHGALGVLI FNPKPQGSNRSMRLTANGMGIPSAFISHEFGKAMSQ L NNGNGTGSLE
FDSVYSKAPSQKGNEMNHF SNWGLTSDGYLKPDI TAPGGDIYSTYNDNHYGSQTGTSMAS PQIAGASLLVKQYLEKTQFN
LPKEKIADIVKNLLMENAQIHVNFETR TTTSPRQOQAGLLNIDGAVTSGLYVTGKDNYSISLGNITDTMTFDVTVHNLS
NKKWTLERYDELLTDHVD PQRGFTLTSHSLKTYQGGEVTVFANGKVTVRVTMDVVSQFTKELTRQMPNGYYLEGFVFRD
SQBDQLNRVNIPFVGFKQGFENLAVAEESIYRLKSQGR TGFFYDESGPKDDIYVGKHTFGLVTLGSETNVSTKTI SDNGL
HTLGTFFKNADGKFILEKNAQGNFVLAISPNGDNNQDFAAFKGVFLRKYQGLKASVYHSDREHKNPLWVSPESFRCGRNE
NSDIRFAKSTLLGTAFSGKSLTGAELPDGHYHYVVSYYPDVVGAKRQENTFDMILDRQKPVLSQATFDPETNRFKPEPL
KDRGLAGVRKDSVFYLERKDNKPYTITINDSYKYVSVEDNKT FVERQADGSFILPLDKAKLGDFFYMVEDFAGNVAIAKL
GDHL PQTLGKTFIRKLITDGN YQTRETLKDNLEMTQSDTGLVTNQAQLAVVHFNQPSQLTRNNQDFFISPNEDGKND FV
AFEGKLNNVYNDLTVNVYARDDHQKQYPIWSSQAGASVSAIESTAWYGITARGSKVMFGDYQYVVVYRDERGRENHQKQYT
ISVNEKKFMITQGRFTINGVDHFTPKTKALDSSGIVREEVFYLA KKNGRKFDVTEGKDCITVSDNKVYIEFNPDG SYT
ISKRDGVTLSDIYYLVEDRAGNVSFATLRDLKAVGKDKAVNVFGLDLPVPEDKQIVNETTYLVRDADGKPIENLEYNNNSG
NSLILPYGKYTVLELLTYDTNAAKLES DKIVSFTLSADRNFPQQVTFKITMLATSQITAHFDHLLPEGSRVSLKTAQDQLIP
LEQSLYVPKAYGKTYQEGTYEVVVS LPKGYRIEGNTKVNITLNEVHEL SLRLVVKVGDA SDSTGDHVKVMSKNNSQALTASA
TPTKSTTSATAKALPSTGEKFKGLKLRIVGLVLGLTQVFSRKKSTKD

Exemplos representativos de imunização com antígenos de GAS da divulgação no modelo murino de rato discutidos acima estão resumidos na Figura 8. A primeira coluna identifica o antígeno GAS utilizado na experiência. Em alguns casos, os aspectos de purificação são referenciados nesta lista. Além disso, as modificações da sequência de polinucleótido que foram feitas para facilitar a expressão recombinante do antígeno são indicadas no gráfico com as seguintes anotações: "a" indica que a região hidrofóbica do terminal N ou C foi removida; RR indica a otimização do códon; "NH" e "CH" correspondem a vectores de expressão semelhantes aos indicados nos exemplos da construção de GAS 40. Onde o

valor de p é dado, ele foi calculado com base nos valores de controlo de paragem de HIS no fundo da tabela.

Ratos imunizados com GAS 40 rendem substancialmente melhores taxas de sobrevivência em desafio - numa coleção de mais de 100 imunizações em ratos, a imunização com GAS 40 rendeu mais de 50% de sobrevivência. Os outros antigénios GAS no gráfico oferecem uma quantidade de proteção que, por exemplo, se combinada com GAS 40, poderia oferecer uma melhor proteção.

A imunogenicidade de outros antigénios GAS conhecidos pode ser melhorada por combinação com dois ou mais GAS do primeiro grupo de antigénio. Estes outros antigénios GAS conhecidos incluem um segundo grupo de antigénio consistindo de (1) uma ou mais variantes da proteína de superfície M ou seus fragmentos, (2) proteína de ligação a fibronectina, (3) proteína associada a estreptococos do grupo heme, ou (4) SagA. Estes antigénios são aqui referidos como "o segundo grupo de antigénio".

Descreve-se aqui uma composição imunogénica que compreende uma combinação de antigénios GAS, consistindo a referida combinação de 2-31 antigénios de GAS do primeiro grupo de antigénio e um, dois, três ou quatro antigénios de GAS do segundo grupo de antigénio. De preferência, a combinação é constituída por três, quatro, cinco, seis, sete, oito, nove, dez ou antigénios de GAS a partir do primeiro grupo de antigénio. Ainda mais preferivelmente, a combinação é constituída por três, quatro ou cinco antigénios de GAS a partir do primeiro grupo de antigénio. De preferência, a combinação de antigénios de GAS inclui um ou ambos GAS 40 e GAS 117. Preferivelmente, a combinação de antigénios de GAS inclui uma ou mais variantes da proteína de superfície M.

Cada um dos antígenos de GAS do segundo grupo de antígeno estão descritos em mais detalhe abaixo

(1) Proteína de superfície M

A proteína M é um factor de virulência GAS que tem sido associada tanto com a colonização como com a resistência à fagocitose. Mais de 100 diferentes variantes do tipo da proteína M foram identificadas com base na especificidade antigénica e a proteína M é pensada ser a principal causa da mutação antigénica e da variação antigénica em GAS. A proteína M liga também o fibrinogénio a partir de soro e bloqueia a ligação do complemento para os peptidoglicanos subjacentes. Esta ação é pensada para aumentar a sobrevivência de GAS dentro de um hospedeiro mamífero, inibindo a fagocitose.

Infelizmente, a proteína M GAS contém alguns epítomos que imitam os do músculo e tecido conjuntivo do mamífero. Certas proteínas GAS M podem ser reumatogénicas uma vez que contêm os epítomos relacionados com o músculo cardíaco, e podem levar a cardite reumática autoimune (febre reumática) na sequência de uma infecção aguda.

Os epítomos que aumentam a atividade bactericida e têm menor probabilidade de uma reação cruzada com os tecidos humanos foram identificadas na região do terminal amino e combinados em proteínas de fusão que contêm cerca de seis, sete ou oito fragmentos de proteína M ligados em série. Ver Hu et al, *Infection & Immunity* (2002) 70 (4):2171-2177; Dale, *Vaccine* (1999) 17:193-200; Dale et al, *Vaccine* 14 (10):944-948; WO 02/094851 e WO 94/06465.

Assim, as composições da invenção podem ainda compreender uma proteína de superfície GAS M ou um seu fragmento ou derivado. Um ou mais fragmentos da proteína de superfície M GAS podem ser combinados numa proteína de fusão. Alternativamente, um ou mais fragmentos de proteína de superfície GAS M são combinados com um antigénio GAS ou um seu fragmento do primeiro grupo de antigénio. Um exemplo de uma proteína GAS M é apresentado na listagem de sequências como SEQ ID N°: 123.

Proteínas GAS M preferidas para utilização com a invenção compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para uma proteína M conhecidos, tais como SEQ ID NO: 123; e / ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos de uma proteína M conhecida, tal como a SEQ ID N°: 123, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 ou mais). Estas proteínas GAS M incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.), da SEQ ID N°: 123. Os fragmentos preferidos de (b) compreendem um epítipo de uma proteína M conhecida, tal como a SEQ ID N°: 123. Preferivelmente, o fragmento é um dos descritos nas referências anteriores. Preferivelmente, o fragmento é construído numa proteína de fusão com um ou mais fragmentos da proteína M. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) da extremidade do terminal C e/ou uma ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N de uma proteína M conhecida, tal como a SEQ ID N°: 123. Outros

fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(2) Proteína de ligação de fibronectina

A ligação da proteína GAS de fibronectina ('SfbI') é uma proteína bacteriana multifuncional pensada para mediar a fixação das bactérias às células do hospedeiro, para facilitar a internalização bacteriana em células e de se ligar ao fragmento Fc de IgG humana, interferindo, assim, com a fagocitose mediada pelo receptor Fc e a citotoxicidade celular dependente de anticorpo. A imunização de ratinhos com uma SfbI e 'fragmento H12' (codificado pelas posições 1240-1854 do gene SfbI) é discutida em Schulze et al, Vaccine (2003) 21:1958-1964; Schulze et al, Infection and Immunity (2001) 69 (1):622-625 e Guzman et al, Journal of Infectious Diseases (1999) 179: 901-906. Um exemplo de uma sequência de aminoácidos para GAS SfbI é mostrado na listagem de sequências como SEQ ID N°: 124.

As proteínas SfbI preferidas para utilização com a invenção compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, de 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 124; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos da SEQ ID N°: 124, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, ou mais). Estas proteínas SfbI incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.) da SEQ ID N°: 124. Os fragmentos

preferidos de (b) compreendem um epítopo a partir da SEQ ID N°: 124. De preferência, o fragmento é um dos descritos nas referências anteriores. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) da extremidade do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 124. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(3) Proteína associada a estreptococos do heme

A proteína GAS associada a estreptococos do heme ('Shp') tem sido identificada como uma proteína da superfície da célula de GAS. Pensa-se ser construída com genes que codificam homólogos de um transportador ABC envolvido na absorção de ferro em bactérias gram-negativas. A proteína Shp é ainda descrita em Lei et al, "Identification and Characterization of a Novel Heme-Associated Cell Surface Protein Made by Streptococcus pyogenes", Infection and Immunity (2002) 70 (8):4494-4500. Um exemplo de uma proteína Shp é mostrado na listagem de sequências como SEQ ID N°: 125.

Proteínas Shp preferidas para utilização com a invenção compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, de 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 125; e/ou (b) que é um fragmento de pelo menos *n* aminoácidos consecutivos da SEQ ID N°: 125, em que *n* é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100

ou mais). Estas proteínas Shp incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.) da SEQ ID N°: 125. Os fragmentos preferidos de (b) compreendem um epítipo a partir da SEQ ID N°: 125. Outros fragmentos preferidos carecem de um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal C e/ou um ou mais aminoácidos (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais) do terminal N da SEQ ID N°: 125. Outros fragmentos omitem um ou mais domínios da proteína (por exemplo, a omissão de um péptido de sinal, de um domínio citoplasmático, um domínio transmembranar, ou de um domínio extracelular).

(4) *SagA*

A estreptolisina S (SLS), também conhecida como "SagA", é pensada ser produzida por quase todas as colónias GAS. Esta toxina citolítica é responsável pela beta-hemólise circundante das colónias de GAS cultivadas em agar do sangue e é pensada ser associada com a virulência. Enquanto o péptido SagA total não tenha sido mostrado ser imunogénico, um fragmento de aminoácidos 10 - 30 (SagA 10 - 30) foi utilizado para produzir anticorpos neutralizantes. Veja Dale et al, "Antibodies against a Synthetic Peptide of SagA Neutralize the Cytolytic Activity of Streptolysin S from Group A Streptococci", *Infection and Immunity* (2002) 70 (4):2166-2170. A sequência de aminoácidos de SagA 10-30 é mostrada na listagem de sequências como SEQ ID N°: 126.

As proteínas SagA 10-30 preferidas para utilização com a invenção compreendem uma sequência de aminoácidos: (a) possuindo 50% ou mais de identidade (por exemplo, de 60%,

65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5% ou mais) para a SEQ ID N°: 126; e/ou (b) que é um fragmento de pelo menos n aminoácidos consecutivos da SEQ ID N°: 126, em que n é de 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, ou 20). Estas proteínas SagA 10-30 incluem variantes (por exemplo, variantes alélicas, homólogos, ortólogos, parálogos, mutantes, etc.) da SEQ ID N°: 126.

Existe um limite superior para o número de antigénios GAS que estará nas composições da invenção. De preferência, o número de antigénios GAS numa composição da presente invenção é menos do que 20, menos do que 19, menos do que 18, menos do que 17, menos do que 16, menos do que 15, menos do que 14, menos do que 13, menos do que 12, menos do que 11, menos do que 10, menos do que 9, inferior a 8, a menos de 7, inferior a 6, menos do que 5, menos do que 4, ou menos do que 3. Ainda mais preferencialmente, o número de antigénios GAS numa composição da invenção é inferior a 6, inferior a 5, ou inferior a 4. Ainda mais de preferência, o número de antigénios GAS numa composição da invenção é 3. Os antigénios GAS utilizados na presente invenção são, de preferência isolados, isto é, separados e discretos, a partir de todo o organismo com o qual a molécula se encontra na natureza ou, quando o polinucleótido ou polipeptídeo não é encontrado na natureza, é suficientemente isento de outras macromoléculas biológicas de modo a que o polinucleótido ou polipeptídeo possa ser utilizado para os fins a que se destinam.

Proteínas de fusão

Os antigénios GAS utilizados na presente invenção podem estar presentes na composição como polipeptídeos

individuais separados, mas prefere-se que pelo menos dois (ou seja, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 ou 20) dos antigénios sejam expressos como uma única cadeia de polipeptídeo (um polipeptídeo 'híbrido'). Polipeptídeos híbridos oferecem duas vantagens principais: em primeiro lugar, um polipeptídeo que pode ser instável ou mal expresso por si só pode ser assistido pela adição de um parceiro híbrido adequado que supera o problema; em segundo lugar, o fabrico comercial é simplificado uma vez que apenas necessita de ser empregue uma expressão e purificação de modo a produzir dois polipeptídeos que são antigenicamente úteis.

O polipeptídeo híbrido pode compreender duas ou mais sequências polipeptídicas do primeiro grupo de antigénio. Deste modo, aqui descrito é uma composição que compreende uma primeira sequência de aminoácidos e uma segunda sequência de aminoácidos, em que a referida primeira e segunda sequência de aminoácidos são selecionadas a partir de um antigénio GAS ou um seu fragmento do primeiro grupo de antigénio. De preferência, as primeiras e segundas sequências de aminoácidos no polipeptídeo híbrido compreendem diferentes epítomos.

O polipeptídeo híbrido pode compreender uma ou mais sequências de polipeptídeos do primeiro grupo de antigénio e uma ou mais sequências de polipeptídeos do segundo grupo de antigénio. Deste modo, aqui descrito é uma composição que compreende uma primeira sequência de aminoácidos e uma segunda sequência de aminoácidos, a referida primeira sequência de aminoácido selecionada a partir de um antigénio GAS ou de um seu fragmento a partir do primeiro grupo de antigénio e a referida segunda sequência de

aminoácidos selecionada a partir de um antígeno GAS ou um seu fragmento, a partir do segundo grupo de antígeno. De preferência, as primeiras e segundas sequências de aminoácidos no polipeptídeo híbrido compreendem diferentes epítomos.

São preferidos os híbridos que consistem em sequências de aminoácidos de dois, três, quatro, cinco, seis, sete, oito, nove, dez ou antígenos GAS. Em particular, são preferidos os híbridos que consistem em sequências de aminoácidos de dois, três, quatro, ou cinco antígenos GAS.

Diferentes polipeptídeos híbridos podem ser misturados em conjunto numa única formulação. Dentro de tais combinações, um antígeno GAS pode estar presente em mais do que um polipeptídeo híbrido e/ou num polipeptídeo que não híbrido. É preferível, no entanto, que um antígeno esteja presente quer como um híbrido ou um não-híbrido, mas não em ambos.

Polipeptídeos híbridos podem ser representados pela fórmula **$\text{NH}_2\text{-A-}\{-\text{X-L-}\}_n\text{-B-COOH}$** , em que: X é uma sequência de aminoácidos de um antígeno GAS ou um seu fragmento a partir do primeiro grupo de antígeno ou do segundo grupo de antígeno; L é uma sequência de aminoácidos ligante opcional; A é uma sequência de aminoácidos do terminal N opcional; B é uma sequência de aminoácidos do terminal C opcional; e *n* é 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ou 15.

Se uma porção -X- tem uma sequência peptídica líder na sua forma de tipo selvagem, tal pode ser incluída ou omitida na

proteína híbrida. Em algumas formas de realização, os peptídeos líder serão eliminados excepto os do radical -X- localizado no terminal N da proteína híbrida, ou seja, o péptido líder de X_1 , será mantido, mas os peptídeos líder de $X_2 \dots X_n$ irão ser omitidos. Isto é equivalente à exclusão de todos os peptídeos líderes e utilizando o péptido líder X_1 como o radical -A-.

Para cada instância n de $\{-X-L-\}$, as sequências de aminoácidos ligantes -L- podem estar presentes ou ausentes. Por exemplo, quando $n = 2$, o híbrido pode ser $NH_2-X_1-L_1-X_2-L_2-COOH$, $NH_2-X_1-X_2-COOH$, $NH_2-X_1-L_1-X_2-COOH$, $NH_2-X_1-X_2-L_2-COOH$, etc. A sequência de aminoácidos ligante -L- será tipicamente curta (por exemplo, 20 ou menos aminoácidos, isto é, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Exemplos incluem sequências peptídicas curtas que facilitam a clonagem, ligantes de poli-glicina (isto é, compreendendo Gly_n onde $n = 2, 3, 4, 5, 6, 7, 8, 9, 10$ ou mais), e marcadores de histidina (ou seja, His_n onde $n = 3, 4, 5, 6, 7, 8, 9, 10$ ou mais). Outras sequências de aminoácido ligante adequadas serão evidentes para os peritos na técnica. Um ligante útil é GSGGGG, com o dipeptídeo Gly-Ser formado a partir de um local de restrição *Bam*HI, facilitando, assim, a clonagem e manipulação, e o tetrapéptido (Gli)₄ sendo um ligante de poli-glicina típico.

-A- é uma sequência de aminoácidos do terminal N opcional. Esta será tipicamente curta (por exemplo, 40 ou menos aminoácidos, ou seja, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Exemplos incluem sequências líder com o tráfico de proteína

direta, ou sequências de peptídeos curtos, que facilitam a clonagem ou purificação (por exemplo, marcadores de histidina, ou seja, His_n onde $n = 3, 4, 5, 6, 7, 8, 9, 10$ ou mais). Outras sequências de aminoácidos do terminal N adequados serão evidentes para os peritos na técnica. Se X_1 carece da sua própria metionina no terminal N, -A- é de preferência um oligopéptido (por exemplo, com 1, 2, 3, 4, 5, 6, 7 ou 8 aminoácidos), que fornece uma metionina no terminal N.

-B- é uma sequência de aminoácidos no terminal C opcional. Esta será tipicamente curta (por exemplo, 40 ou menos aminoácidos, ou seja, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Exemplos incluem sequências de tráfico de proteína direta, sequências peptídicas curtas que facilitam a clonagem ou a purificação (por exemplo, compreendendo n marcadores de histidina, isto é, His_n , onde $n = 3, 4, 5, 6, 7, 8, 9, 10$ ou mais), ou sequências que intensificam a estabilidade da proteína. Outras sequências de aminoácidos do terminal C adequados serão evidentes para os peritos na técnica.

Mais preferivelmente, n é 2 ou 3.

As construções de fusão aqui descritas podem incluir uma combinação de dois ou mais antígenos GAS, em que a referida combinação inclui GAS 40 ou um seu fragmento ou um polipeptídeo com uma identidade de sequência com esta.

As construções de fusão aqui divulgadas podem incluir uma combinação de antígenos GAS, consistindo a dita combinação em 2-31 antígenos GAS do primeiro grupo de antígeno, sendo o dito primeiro grupo de antígeno composto por: GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, 504 GAS, GAS 509, 366 GAS, GAS 159, GAS 217, GAS 309, 372 GAS, GAS 039, 042 GAS, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, e GAS 057. De preferência, a combinação de antígenos GAS é composta por três, quatro, cinco, seis, sete, oito, nove, dez ou antígenos GAS selecionados a partir do primeiro grupo de antígeno. De preferência, a combinação de antígenos GAS consiste de três, quatro, ou cinco antígenos GAS selecionados a partir do primeiro grupo de antígeno.

GAS 39, GAS 40, GAS 57, GAS 117, GAS 202, GAS 294, GAS 527, GAS 533, e GAS 511 são antígenos GAS particularmente preferidos para uso nas construções de fusão aqui divulgadas. De preferência, a combinação de antígenos GAS inclui um ou ambos GAS 40 e GAS 117. Preferivelmente, a combinação inclui GAS 40.

A expressão recombinante das construções de fusão aqui descritas pode ser melhorada ou otimizada por métodos idênticos aos descritos para a expressão de antígenos GAS isolados (discutida acima). Construções de fusão de GAS 40 e GAS 117 são exemplificadas abaixo.

No primeiro exemplo, o GAS 117 está ligada ao GAS 40a-RR. (como discutido acima, GAS 40a-RR é uma sequência de códon GAS 40otimizado, onde a sequência líder no terminal N e a

sequência da transmembrana no terminal C são removidas). Nesta construção um fragmento de GAS 117 (onde a sequência líder no terminal N removida) é colocado no terminal N da sequência de 40 GAS e o marcador HIS é adicionado ao terminal C da sequência de GAS 40. Esta construção é designada "117-40a-RR". Sequências de aminoácidos e de polinucleótidos para esta construção são mostradas na listagem de Sequências como SEQ ID N°s: 127 e 128.

As sequências GAS 117 e GAS 40 são preferencialmente ligadas por uma sequência de ligação compreendendo múltiplos resíduos de glicina. Por exemplo, o ligante utilizado na construção de fusão 117-40a-RR, é usadaa uma sequência de ligação da SEQ ID N°: 129 (YASGGGS).

Num segundo exemplo, as posições relativas das sequências de GAS 40 e GAS 117 podem ser trocadas. Nesta construção, designada "40a-RR-117", a sequência GAS 40a-RR é colocada no terminal N da sequência de GAS 117 e o marcador HIS é adicionado ao terminal C da sequência de GAS 117. Sequências de aminoácidos e de polinucleótidos para esta construção de fusão são apresentadas na listagem de sequências como SEQ ID N°s: 130 e 131.

Alternativamente, as construções de fusão podem ser concebidas sem otimizações de códãos. Por exemplo, as sequências de polinucleótidos e de aminoácidos para a construção de fusão "117-40a" são mostradas na listagem de Sequências como SEQ ID N°s: 132 e 133 (Embora não tenha sido utilizada nenhuma otimização de códão, três mutações pontuais aparentemente ocorreram durante a clonagem, somente uma das quais envolvendo uma mudança de aminoácidos

conservativa (glucina para glicina). No modelo murino de imunização (anteriormente discutido acima), a imunização com "117-40a" produziu até 80% de sobrevivência após o desafio.

A sequência de fusão GAS40 preferida compreende um fragmento de GAS 40 compreendendo uma ou mais das regiões de espiral enrolada. Por exemplo, a construção de fusão pode compreender uma sequência GAS 40 compreendendo a primeira região em espiral enrolada. "117-40N" é um exemplo deste tipo de construção. As sequências de aminoácidos e de polinucleótidos para esta construção são mostradas na listagem de sequências como SEQ ID N°s: 132 e 133.

É também aqui descrito são ácidos nucleicos que codificam polipeptídeos híbridos da divulgação. Além disso, aqui descrito é o ácido nucleico que pode hibridar com este ácido nucleico, de preferência sob condições de "elevado rigor" (por exemplo, 65°C numa solução de 0,1 x SSC, SDS a 0,5%).

Os antigénios GAS aqui revelados também podem ser utilizados para preparar anticorpos específicos para os antigénios GAS. Os anticorpos são de preferência específicos para as primeiras ou segundas regiões de espiral enrolada GAS 40 Também é aqui descrito o uso da combinação de dois ou mais tipos de anticorpos selecionados a partir do grupo que consiste de anticorpos específicos de GBS 80, GAS 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095,

GAS 193, GAS 137, GAS 084, GAS 384, GAS 202 e GAS 057. De preferência, a combinação inclui um anticorpo específico para GAS 40, ou um seu fragmento.

Os anticorpos específicos GAS aqui divulgados incluem uma ou mais unidades biológicas que, através de meios químicos ou físicos, se podem ligar ou associar com um epítipo de um polipeptídeo GAS. Os anticorpos aqui descritos incluem anticorpos que se ligam especificamente a um antigénio GAS, de preferência GAS 80. Aqui descritos são anticorpos obtidos a partir de preparações de anticorpos policlonais e monoclonais, bem como as seguintes: moléculas de anticorpos híbridos (quiméricos) (ver, por exemplo, Winter et al. (1991) Nature 349: 293-299; e Patente US 4816567; fragmentos F (ab')₂ e F(ab); moléculas F_v (heterodímeros não covalentes, ver, por exemplo, Inbar et al. (1972) Proc Natl Acad Sci USA 69:2659-2662; e Ehrlich et al. (1980) Biochem 19: 4091-4096); moléculas Fv de cadeia simples (sFv) (ver, por exemplo, Huston et al. (1988) Proc Natl Acad Sci USA 85:5897-5883); construções de fragmentos de anticorpos diméricos e triméricos; minicorpos (ver, por exemplo, Pacxk et al. (1992) Biochem 31:1579-1584; Cumber et al. (1992) J Immunology 149B:120-126); moléculas de anticorpo humanizado (ver, por exemplo, Riechmann et al. (1988) Nature 332: 323-327; Verhoeyan et al. (1988) Science 239:1534-1536; e Publicação de Patente GB 2276169, publicada 21 de Setembro de 1994); e, quaisquer fragmentos funcionais obtidos dessas moléculas, em que esses fragmentos retêm propriedades de ligação imunológicas da molécula do anticorpo original. Além disso são aqui descritos os anticorpos obtidos através de processos não-convencionais, tais como a apresentação de fagos.

De preferência, os anticorpos específicos GAS aqui divulgados são anticorpos monoclonais. Os anticorpos monoclonais aqui descritos incluem uma composição de anticorpos com uma população homogênea de anticorpos. Os anticorpos monoclonais aqui descritos podem ser obtidos a partir de hibridomas de murino, assim como, anticorpos monoclonais humanos obtidos utilizando hibridomas humanos em vez de murino. Ver, *por exemplo*, Cote, et al. *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, p 77.

Os polipeptídeos da invenção podem ser preparados através de vários meios (por exemplo, expressão recombinante, purificação a partir de cultura celular, síntese química, etc.) e em diversas formas (por exemplo, nativa, fusões, não glicosilada, lipidadas, etc.). Eles são preferivelmente preparados na forma substancialmente pura (isto é, substancialmente livre de outras proteínas GAS ou da célula hospedeira).

Ácidos nucleicos aqui divulgados podem ser preparados de muitas maneiras (por exemplo, por síntese química, a partir de bibliotecas genômicas ou de cADN, a partir do próprio organismo, etc.) e podem assumir diversas formas (por exemplo, cadeia simples, cadeia dupla, vectores, sondas, etc.). Eles são preferivelmente preparados na forma substancialmente pura (isto é, substancialmente livre de GAS ou ácidos nucleicos da célula hospedeira).

O termo "ácido nucleico" inclui ADN e ARN, e também os seus análogos, tais como os que contêm estruturas modificadas (por exemplo, fosforotioatos, etc.), e também os ácidos nucleicos peptídicos (PNA), etc. Aqui divulgado é ácido

nucleico compreendendo sequências complementares àqueles descritos acima (por exemplo, antessentido ou para fins de sondagem).

A invenção também fornece um processo para a produção de um polipeptídeo da invenção, compreendendo o passo de cultivar uma célula hospedeira transformada com o ácido nucleico aqui divulgado sob condições que induzem a expressão do polipeptídeo.

A invenção proporciona um processo para a produção de um polipeptídeo da invenção, compreendendo o passo de síntese, pelo menos de parte do polipeptídeo, por meios químicos.

Descreve-se aqui um processo para a produção de ácido nucleico aqui divulgado, compreendendo o passo de amplificação de ácido nucleico utilizando um método de amplificação à base de iniciadores (por exemplo, PCR).

Descreve-se aqui um processo para a produção de ácido nucleico aqui divulgado, compreendendo o passo de síntese, de pelo menos, parte do ácido nucleico, por meios químicos.

Estirpes

Os polipeptídeos preferidos da presente invenção compreendem uma sequência de aminoácidos encontrada numa estirpe M1, M3 ou M18 de GAS. A sequência genômica de uma estirpe GAS M1 é relatada em Ferretti et al, PNAS (2001) 98 (8):4658-4663. A sequência genômica de uma estirpe M3 GAS relatada em Beres et al, PNAS (2002) 99 (15):10078-10083. A

sequência genómica de uma estirpe M18 GAS é relatada em Smooet et al, PNAS (2002) 99 (7):4668-4673.

Quando são utilizados polipeptídeos híbridos, os antigénios individuais dentro do híbrido (isto é, porções de -X-individuais) podem ser de uma ou mais estirpes. Quando $n = 2$, por exemplo, X_2 pode ser da mesma estirpe que X_1 ou a partir de uma estirpe diferente. Onde $n = 3$, as estirpes podem ser (i) $X_1 = X_2 = X_3$ (ii) $X_1 = X_2 \neq X_3$ (iii) $X_1 \neq X_2 = X_3$ (iv) $X_1 \neq X_2 \neq X_3$ ou (v) $X_1 = X_3 \neq X_2$, etc.

Purificação e expressão recombinante

Os antigénios GAS da invenção podem ser isolados a partir de *Streptococcus pyogenes*, ou podem ser produzidos de forma recombinante, por exemplo, num hospedeiro heterólogo. De preferência, os antigénios GAS são preparados usando um hospedeiro heterólogo. O hospedeiro heterólogo pode ser procariótico (por exemplo, uma bactéria) ou eucariótico. É de preferência *E.coli*, mas outros hospedeiros apropriados incluem *Bacillus subtilis*, *Vibrio cholerae*, *Salmonella typhi*, *Salmonella typhimurium*, *Neisseria lactamica*, *Neisseria cinerea*, *Mycobacteria* (por exemplo, *M. tuberculosis*), leveduras, etc.

A produção recombinante de polipeptídeos é facilitada pela adição de uma proteína marcada para o antigénio GAS a ser expresso como uma proteína de fusão que compreende a proteína de marcação e o antigénio GAS. Tais proteínas de marcação podem facilitar a purificação, detecção e a estabilidade da proteína expressa. As proteínas de marcação adequadas para utilização na presente invenção incluem uma poliarginina marcada (Arg-tag), uma poli-histidina marcada

(His-marcada), FLAG-marcada, Strep-marcada, c-myc-marcada, S-marcada, o péptido de ligação à calmodulina, o domínio de ligação à celulose, domínio de ligação a quitina, "SBP-marcada", a glutational S-transferase-marcada (GST), proteína de ligação a maltose, o factor de terminação da transcrição anti-terminação (NusA), tioredoxina de *E. coli* (TrxA) e e proteína de isomerase I de dissulfeto (DsbA). As proteínas marcadas preferidas incluem His-marcada e GST. Uma discussão completa sobre o uso de proteínas de marcação pode ser encontrada em Terpe et al, Appl Microbiol Biotechnol (2003) 60:523-533.

Após a purificação, as proteínas de marcação podem opcionalmente ser removidas da proteína de fusão expressa, isto é, por tratamento enzimático especificamente adaptado conhecido na técnica. Proteases comumente usadas incluem enteroquinase, vírus do tabaco (TEV), trombina e fator X_a.

Composições imunogénicas e medicamentos

As composições da invenção são composições imunogénicas, e são com maior preferência as composições de vacina. O pH da composição é, de preferência, entre 6 e 8, de preferência cerca de 7. O pH pode ser mantido pelo uso de um tampão. A composição pode ser esterilizada e/ou livre de pirogénio. A composição pode ser isotónica no que diz respeito aos seres humanos.

As vacinas de acordo com a invenção tanto podem ser profiláticas (isto é, para evitar a infecção) ou terapêuticas (ou seja, para tratar a infecção), mas normalmente serão profiláticas. Assim, é aqui revelado um método para o tratamento terapêutico ou profilático de uma

infecção por *Streptococcus pyogenes* num animal susceptível à infecção por estreptococos, compreendendo a administração ao referido animal de uma quantidade terapêutica ou profilática de composições imunogénicas da invenção. De preferência, a composição imunogénica compreende uma combinação de antigénios GAS, consistindo a referida combinação de 2-31 antigénios GAS do primeiro grupo de antigénio. De preferência, a combinação de antigénios GAS consiste de três, quatro, cinco, seis, sete, oito, nove, dez ou antigénios GAS selecionados a partir do primeiro grupo de antigénio. De preferência, a combinação de antigénios GAS consiste de três, quatro, ou cinco antigénios de GAS selecionados a partir do primeiro grupo de antigénio. De preferência, a combinação de antigénios de GAS inclui um ou ambos GAS 40 e GAS 117.

Alternativamente, aqui divulgada é uma composição imunogénica que compreende uma combinação de antigénios GAS, consistindo a referida combinação de 2-31 antigénios de GAS do primeiro grupo de antigénio e um, dois, três ou quatro antigénios de GAS do segundo grupo de antigénio. De preferência, a combinação é constituída por três, quatro, cinco, seis, sete, oito, nove, dez ou antigénios GAS a partir do primeiro grupo de antigénio. Ainda mais preferivelmente, a combinação é constituída por três, quatro ou cinco antigénios GAS a partir do primeiro grupo de antigénio. De preferência, a combinação de antigénios GAS inclui um ou ambos GAS 40 e GAS 117. Preferivelmente, a combinação de antigénios GAS inclui uma ou mais variantes da proteína de superfície M.

A invenção proporciona também uma composição da invenção para uso como um medicamento. O medicamento é de

preferência capaz de suscitar uma resposta imunitária num mamífero (ou seja, é uma composição imunogénica) e é mais preferencialmente uma vacina.

A invenção também proporciona a utilização das composições da invenção no fabrico de um medicamento para aumentar uma resposta imune num mamífero. O medicamento é de preferência uma vacina. A invenção também proporciona um kit compreendendo um primeiro componente que compreende uma combinação de antigénios GAS. A combinação dos antigénios de GAS pode consistir numa mistura de 2-31 antigénios GAS selecionados a partir do primeiro grupo de antigénio. De preferência, a combinação é constituída por três, quatro, cinco, seis, sete, oito, nove, dez ou antigénios GAS a partir do primeiro grupo de antigénio. De preferência, a combinação é constituída por três, quatro, ou cinco antigénios GAS a partir do primeiro grupo de antigénio. De preferência, a combinação inclui um ou ambos de GAS 117 e GAS 040.

O kit pode compreender um primeiro componente que compreende uma combinação de antigénios GAS constituída por uma mistura de 2-31 antigénios GAS do primeiro grupo de antigénio e um, dois, três ou quatro antigénios GAS do segundo grupo de antigénio. De preferência, a combinação é constituída por três, quatro, cinco, seis, sete, oito, nove, dez ou antigénios GAS a partir do primeiro grupo de antigénio. Ainda mais preferivelmente, a combinação é constituída por três, quatro ou cinco antigénios GAS a partir do primeiro grupo de antigénio. De preferência, a combinação de antigénios GAS inclui um ou ambos GAS 40 e GAS 117. Preferivelmente, a combinação de antigénios GAS inclui uma ou mais variantes da proteína de superfície M.

A invenção também fornece um dispositivo de distribuição pré-cheio com as composições imunogénicas da invenção.

A invenção também fornece um método para aumentar uma resposta imune num mamífero, compreendendo o passo de administrar uma quantidade eficaz de uma composição da invenção. A resposta imunitária é preferencialmente de proteção e, de preferência envolve imunidade mediada por anticorpos e/ou por células. O método pode aumentar a resposta ao reforço.

O mamífero é preferencialmente um ser humano. Quando a vacina é para uso profilático, o ser humano é, de preferência uma criança (por exemplo, uma criança ou bebé) ou um adolescente; em que a vacina é para uso terapêutico, o ser humano é de preferência um adolescente ou um adulto. Uma vacina destinada a crianças pode também ser administrada a adultos, por exemplo, para avaliar a segurança, dosagem, imunogenicidade, etc.

Estas utilizações e métodos são, de preferência, para a prevenção e/ou tratamento de uma doença causada por *Streptococcus pyogenes* (por exemplo, faringite (tais como dor de garganta provocada por estreptococos), escarlatina, impetigo, erisipela, celulite, septicemia, síndrome do choque tóxico, fascite necrotizante (doença que come a carne) e sequelas (como a febre reumática e a glomerulonefrite aguda)). As composições podem também ser eficazes contra outras bactérias estreptocócicas.

Uma maneira de verificar a eficácia do tratamento terapêutico envolve o acompanhamento da infecção por GAS após a administração da composição da invenção. Uma maneira de verificar a eficácia do tratamento profilático envolve a monitorização de respostas imunitárias contra os antígenos GAS em composições da invenção após administração da composição.

As composições da invenção irão geralmente ser administradas diretamente a um paciente. A entrega direta pode ser conseguida por injeção parentérica (por exemplo, por via subcutânea, por via intraperitoneal, por via intravenosa, por via intramuscular, ou para o espaço intersticial de um tecido), ou por administração rectal, bucal (por exemplo, comprimido, pulverização), vaginal, tópica, transdérmica (por exemplo, ver WO99/27961) ou transcutânea (por exemplo, ver WO02/074244 e WO02/064162), intranasal (por exemplo, ver WO03/028760), ocular, oral, através da mucosa pulmonar ou outra. A invenção pode ser utilizada para induzir a imunidade sistémica e/ou das mucosas.

O tratamento por dosagem pode ser um programa de dose única ou um programa de múltiplas doses. Doses múltiplas podem ser usadas num esquema de imunização primária e/ou num esquema de imunização de reforço. Num plano de dose múltipla as várias doses podem ser administradas pela mesma ou por diferentes vias, por exemplo, uma primeira parenteral e o reforço mucosal, uma primeira mucosal e o reforço parenteral, etc.

As composições da invenção podem ser preparadas de várias formas. Por exemplo, as composições podem ser preparadas como injetáveis, quer como soluções líquidas ou suspensões. Formas sólidas adequadas para solução em, ou suspensão em veículos líquidos antes da injeção podem também ser preparadas (por exemplo, uma composição liofilizada). A composição pode ser preparada para administração tópica, por exemplo, como uma pomada, creme ou pó. A composição pode ser preparada para administração por via oral, por exemplo, como um comprimido ou uma cápsula, como uma pulverização, ou como um xarope (opcionalmente aromatizado). A composição pode ser preparada para administração pulmonar, por exemplo, como um inalador, utilizando-se um pó fino ou um spray. A composição pode ser preparada como um supositório ou pessário. A composição pode ser preparada para administração nasal, oral ou ocular, por exemplo, na forma de gotas. A composição pode estar na forma de kit, concebido de tal modo que uma composição combinada é reconstituída imediatamente antes da administração a um paciente. Tais kits podem compreender um ou mais antigénios em forma líquida e um ou mais antigénios liofilizados. As composições imunogénicas utilizadas como vacinas compreendem uma quantidade imunologicamente eficaz do antigénio (s), bem como quaisquer outros componentes, conforme necessário. Por "quantidade imunologicamente eficaz" pretende-se dizer que a administração dessa quantidade a um indivíduo, quer numa dose única ou como parte de uma série, é eficaz para tratamento ou prevenção. Esta quantidade varia dependendo da saúde e condição física do indivíduo a ser tratado, idade, grupo taxonómico do indivíduo a ser tratado (por exemplo, primata não-humano, primata, etc.), a capacidade do sistema imunitário do indivíduo para sintetizar anticorpos, o grau de proteção desejado, a formulação da vacina, a avaliação do médico

assistente sobre a situação médica e outros fatores relevantes. Espera-se que a quantidade irá cair numa gama relativamente ampla que pode ser determinada através de ensaios de rotina.

Outros componentes da composição

A composição da invenção compreenderá tipicamente, em adição aos componentes mencionados acima, um ou mais dos transportadores farmaceuticamente aceitáveis", que incluem qualquer transportador que não induza ele próprio a produção de anticorpos prejudiciais ao indivíduo que recebe a composição. Os veículos adequados são tipicamente grandes macromoléculas lentamente metabolizadas, tais como proteínas, polissacarídeos, ácidos polilácticos, ácidos poliglicólicos, aminoácidos poliméricos, copolímeros de aminoácidos, e agregados lipídicos (tais como gotículas de óleo ou lipossomas). Tais transportadores são bem conhecidos dos peritos na técnica. As vacinas também podem conter diluentes, tais como água, solução salina, glicerol, etc. Além disso, substâncias auxiliares, tais como agentes molhantes ou emulsionantes, substâncias de tamponamento do pH e semelhantes, podem estar presentes. Uma discussão exaustiva de excipientes farmaceuticamente aceitáveis está disponível em Gennaro (2000) Remington: The Science and Practice of Pharmacy. 20a ed, ISBN: 0683306472.

As vacinas da invenção podem ser administradas em conjunto com outros agentes imunorreguladores. Em particular, as composições geralmente incluem um adjuvante.

Outros adjuvantes preferidos incluem, mas não estão limitados a, um ou mais dos seguintes estabelecidos abaixo:

A. Compostos contendo minerais

Composições contendo minerais adequadas para uso como adjuvantes na invenção incluem sais minerais, tais como sais de alumínio e sais de cálcio. A invenção pode ser usada com sais minerais, tais como os hidróxidos (por exemplo, oxihidróxidos), fosfatos (por exemplo, hidroxifosfatos, ortofosfatos), sulfatos, etc. {por exemplo, ver capítulos 8 e 9 de Vaccine Design: the subunit and adjuvant approach (1995) Powell & Newman. ISBN 0-306-44867-X}}, ou misturas de compostos minerais diferentes, com os compostos tendo qualquer forma adequada (por exemplo, gel, cristalina, amorfa, etc.), e com a adsorção sendo preferida. As composições contendo minerais podem também ser formuladas como uma partícula de sal metálico. Ver WO00/23105.

B. Emulsões em óleo

As composições em emulsão em óleo adequadas para uso como adjuvantes na presente invenção incluem emulsões de esqualeno em água, tal como o MF59 (5% de esqualeno, 0,5% de Tween 80, e 0,5% de Span 85, formulado em partículas submicrônicas usando um microfluidificador). Ver WO90/14837. Ver também, Podda, "The adjuvanted influenza vaccines with novel adjuvants: experience with the MF59-adjuvanted vaccine", Vaccine (2001) 19:2673-2680; Frey et al, "Comparison of the safety, tolerability, and immunogenicity of a MF59-adjuvanted influenza vaccine and a non-adjuvanted influenza vaccine in non-elderly adults", Vaccine (2003) 21:4234-4237. MF59 é utilizado como adjuvante na subunidade da vacina trivalente do vírus da gripe FLUAD™.

Os adjuvantes particularmente preferidos para utilização nas composições são emulsões submicrônicas óleo-em-água. Emulsões óleo-em-água submicrônicas preferidas para uso na invenção são emulsões de esqualeno/água contendo opcionalmente quantidades variáveis de MTP-PE, tal como uma emulsão submicrônica óleo-em-água, contendo 4-5% p/v de esqualeno, 0,25-1,0% p/v de Tween 80™ (monooleato de polioxiétilenosorbitano), e/ou 0,25-1,0% de Span 85™ (trioleato de sorbitano), e, opcionalmente, N-acetilmuramil-L-alanil-D-isoglutaminil-L-alanina-2-(1'-2'-dipalmitoil-sn-glicero-3-hidroxifosforiloxi)-etilamina (MTP-PE), por exemplo, a emulsão de óleo-em-água submicrônica conhecida como "MF59" (Publicação Internacional WO90/14837; Patentes US 6299884 e 6451325; e Ott et al, "MF59 - Design and Evaluation of a safe and potente adjuvante for human vaccines" em Vaccine Design: The Subunit and Adjuvant Approach (Powell, MF e Newman, MJ eds Plenum Press, New York, 1995, pp. 277-296). MF59 contém 4-5% p/v de esqualeno (por exemplo, 4,3%), 0,25-0,5% p/v de Tween 80™, e 0,5% p/v de Span 85™ e contém opcionalmente várias quantidades de MTP-PE, formulados em partículas submicrônicas usando um microfluidificador tal como o microfluidificador Modelo 110Y (Microfluidics, Newton, MA). Por exemplo, a MTP-PE poderá estar presente numa quantidade de cerca de 0-500 µg/dose, mais preferencialmente 0-250 µg/dose e, mais preferivelmente, 0-100 µg/dose. Tal como aqui utilizado, o termo "MF59-0" refere-se à emulsão de óleo-em-água submicrônica acima sem MTP-PE, enquanto o termo MF59-MTP indica uma formulação que contém MTP-PE. Por exemplo, "MF59-100" contém 100 µg de MTP-PE por dose, e assim por diante. MF69, outra emulsão submicrônica óleo-em-água para utilização na presente invenção, contém 4,3% p/v de esqualeno, 0,25% p/v de Tween 80™ e 0,75% p/v de Span 85™ e opcionalmente MTP-PE. Ainda outra emulsão

submicrónica óleo-em-água é MF75, também conhecida como SAF, contendo 10% de esqualeno, 0,4% de Tween 80™, 5% de polímero de plurónico bloqueado L121, e thr-MDP, também microfluidificado numa emulsão submicrónica. MF75-MTP indica uma formulação MF75 que inclui MTP, tal como 100-400 µg de MTP-PE por dose.

Emulsões submicrónicas óleo-em-água, os métodos de fazer as mesmas e agentes imunoestimulantes, tais como, péptidos de muramil, para utilização nas composições, estão descritos em detalhe na Publicação Internacional WO90/14837 e nas Patentes US 6299884 e 6451325.

O adjuvante completo de Freund (CFA) e o adjuvante incompleto de Freund (IFA) podem igualmente ser utilizados como adjuvantes na presente invenção.

C. Formulações de saponina

As formulações de saponina também podem ser utilizadas como adjuvantes na invenção. As saponinas são um grupo heterólogo de glicósidos de esterol e glicósidos triterpenóides que são encontrados na casca, folhas, caules, raízes e até flores de uma ampla gama de espécies de plantas. A saponina da casca da árvore *Quillaia saponaria* Molina tem sido amplamente estudada como adjuvantes. A saponina também pode ser obtida comercialmente a partir de *Smilax ornata* (sarsapilla), *Gypsophilla paniculata* (véu de noiva) e *Saponaria officianalis* (raiz de sabão). Formulações de adjuvantes de saponina incluem formulações purificadas, tais como QS21, bem como formulações de lípidos, tais como ISCOM. As composições de saponinas foram purificadas usando

cromatografia em camada fina de alta eficiência (HP-LC) e Cromatografia Líquida de Fase Reversa de Alta Eficiência (HPLC). Fracções purificadas específicas usando estas técnicas têm sido identificadas, incluindo QS7, QS17, QS18, QS21, QH-A, QH-B e QH-C. De preferência, a saponina é QS21. Um método de produção de QS21 é divulgado na Patente US 5057540. As formulações de saponina podem também compreender um esterol, tal como colesterol (ver WO 96/33739).

Combinações de saponinas e colesteróis podem ser usados para formar partículas únicas chamadas complexos imunoestimulantes (ISCOM). ISCOM tipicamente também incluem um fosfolípido, tal como, fosfatidiletanolamina ou fosfatidilcolina. Qualquer saponina conhecida pode ser utilizada em ISCOMs. Preferencialmente, o ISCOM inclui um ou mais de Quil A, QHA e QHC. ISCOM são ainda descritos em EP 0 109 942, WO 96/11711 e WO 96/33739. Opcionalmente, os ISCOMS podem ser desprovidos de detergente adicional. Ver a WO00/07621.

A avaliação do desenvolvimento de adjuvantes com base em saponina pode ser encontrada em Barr, et al, Advanced Drug Delivery Reviews (1998) 32:247-271. Ver também Sjolander, et al, Advanced Drug Delivery Reviews (1998) 32:321-338.

C. Virossomas e partículas do tipo viral (VLPs)

Os virossomas e as partículas do tipo viral (VLPs) podem também ser utilizados como adjuvantes na presente invenção. Estas estruturas geralmente contêm uma ou mais proteínas do vírus, opcionalmente combinado ou formulado com um fosfolípido. Elas são geralmente não-patogénicas, não

replicantes e, geralmente, não contêm qualquer do genoma viral nativo. As proteínas virais podem ser recombinantemente produzidas ou isoladas de vírus inteiros. Estas proteínas virais adequadas para o uso em virossomas ou VLPs incluem proteínas derivadas do vírus da gripe (tais como HA ou NA), vírus da Hepatite B (tal como proteínas da cápside ou do núcleo), vírus da hepatite E, vírus do sarampo, vírus Sindbis, rotavírus, vírus da febre aftosa, retrovírus, vírus Norwalk, vírus do papiloma humano, VIH, fagos de ARN-, fago Q β - (tais como proteínas de revestimento), fago GA, fago fr-, fago AP205, e Ty (tal como a proteína p1 de retrotransposições Ty). VLPs são discutidos em WO 03/024480, WO 03/024481, e Niikura et al, Virology (2002) 293:273-280, Lenz et al, Journal of Immunology (2001) 5246-5355; Pinto, et al, Journal of Infectious Diseases (2003) 188:327-338 e Gerber et al, Journal of Virology (2001) 75 (10):4752-4760. Os virossomas são discutidos, também, por exemplo, em Gluck et al, Vaccine (2002) 20:B10 -B16.

D. Derivados bacterianos ou microbianos

Os adjuvantes adequados para uso na invenção incluem derivados bacterianos ou microbianos tais como:

(1) os derivados não-tóxicos de lipopolissacarídeos enterobacterianos (LPS)

Tais derivados incluem monofosforil-lípido A (MPL) e MPL 3-O-desacilado (3dMPL). 3dMPL é uma mistura de 3-De-O-acilado monofosforil lípido A com 4, 5 ou 6 cadeias aciladas. Uma forma de "partícula pequena" preferida de 3-De-O-acilado monofosforil lípido A é descrita em EP 0 689 454. Tais "partículas pequenas" de 3dMPL são suficientemente pequenas para serem esterilizadas por filtração através de uma membrana de 0,22 micron (ver EP 0 689 454). Outros

derivados não-tóxicos de LPS incluem imitações de monofosforil lípido A, tais como derivados de fosfato de aminoalquil glucosaminida, por exemplo, RC-529. Ver Johnson et al. (1999) Bioorg Med Chem Lett 9:2273-2278.

(2) Derivados de lípidos A

Derivados de lípido A incluem derivados de lípido A de *Escherichia coli* tais como OM-174. OM-174 é descrito, por exemplo, em Meraldi et al, Vaccine (2003) 21:2485-2491 e Pajak et al, Vaccine (2003) 21:836-842.

(3) Oligonucleótidos imunoestimulantes

Oligonucleótidos imunoestimulantes adequados para uso como adjuvantes na presente invenção incluem sequências de nucleótidos que contêm um motivo CpG (uma sequência contendo uma citosina não metilada seguida de guanossina e ligadas por uma ligação de fosfato). ARN bacteriano de cadeia dupla, ou oligonucleótidos contendo sequências palindrômicas ou poli (dG) também tem demonstrado ser imunoestimulante.

CpG pode incluir modificações de nucleotídeos/análogos, tais como modificações de fosforotioato e pode ser de cadeia dupla ou de cadeia simples. Opcionalmente, a guanossina pode ser substituída por um análogo, tais como 2'-desoxi-7-deazaguanossina. Ver, Kandimalla et al, Nucleic Acids Research (2003) 31 (9):2393-2400; WO 02/26757 e WO 99/62923 para exemplos de possíveis substituições de análogos. O efeito adjuvante de oligonucleotídeos CpG é também discutido em Krieg, Nature Medicine (2003), 9 (7): 831-835; McCluskie et al, FEMS Immunology and Medical Microbiology (2002) 32:179-185; WO 98/40100, Patente US 6207646, Patente US 6239116 e Patente US 6429199.

A sequência CpG pode ser direcionada para TLR9, como o motivo GTCGTT ou TTCGTT. Ver Kandimalla, et al, Biochemical Society Transactions (2003) 31 (parte 3):654-658. A sequência CpG pode ser específica para a indução de uma resposta imune Th1, tal como CpG-A ODN, ou pode ser mais específico para a indução de uma resposta de células B, tal CpG-B ODN. CpG-A e CpG-B ODN são discutidos em Blackwell et al., J. Immunol. (2003) 170 (8):4061-4068; Krieg, Trends in Immunology (2002) 23 (2):64-65 e WO 01/95935. De preferência, o CpG é um CpG-A ODN.

De preferência, o oligonucleótido CpG é construído de modo que a extremidade 5' seja acessível para o reconhecimento do receptor. Opcionalmente, duas sequências de oligonucleótido CpG podem estar ligadas às suas extremidades 3' para formar "imunómeros". Ver, por exemplo, Kandimalla et al, BBRC (2003) 306:948-953; Kandimalla, et al, Biochemical Society Transactions (2003) 31 (parte 3):664-658; Bhagat et al, BBRC (2003) 300:853-861 e WO 03/035836.

(4) ADP-ribosilação de toxinas e seus derivados destoxificados.

Toxinas bacterianas ribosilantes de ADP e derivados destoxificados podem ser usados como adjuvantes na invenção. De preferência, a proteína é derivada de *E. coli* (isto é, enterotoxina lábil ao calor de *E. coli*, "LT), da cólera ("TC"), ou da tosse convulsa ("PT"). A utilização de toxinas ribosilantes desintoxicadas de ADP como adjuvantes de mucosa é descrita em WO 95/17211 e como adjuvantes parenterais em WO 98/42375. De preferência, o adjuvante é um mutante LT destoxificado tal como LT-K63.

E. Imunomoduladores humanos

Imunomoduladores humanos adequados para uso como adjuvantes na invenção incluem citocinas, tais como interleucinas (por exemplo, IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferões (por exemplo, o interferão- γ), factor de estimulação de colónias de macrófagos, e factor de necrose tumoral.

F. Bioadesivos e mucoadesivos

Bioadesivos e mucoadesivos podem também ser utilizados como adjuvantes na presente invenção. Bioadesivos adequados incluem microesferas de ácido hialurónico esterificado (Singh et al. (2001) J. Cont. Rele. 70: 267-276) ou mucoadesivos tais como derivados reticulados de poli (ácido acrílico), álcool polivinílico, polivinilpirrolidona, carboximetilcelulose e polissacarídeos. O quitosano e seus derivados podem também ser utilizados como adjuvantes na presente invenção. Por exemplo, W099/27960.

G. Micropartículas

As micropartículas podem também ser utilizadas como adjuvantes na invenção. As micropartículas (isto é, uma partícula de ~100nm para ~150 μ m de diâmetro, mais de preferência de ~200nm a ~30 μ m de diâmetro, e mais preferencialmente de ~500nm a ~10 μ m de diâmetro) formadas a partir de materiais que são biodegradáveis e não tóxicos (por exemplo, um poli (ácido hidróxi- α), um ácido polihidroxibutírico, um poliortoéster, um polianidrido, uma policaprolactona, etc.), com poli (lactido-co-ácido glicólico) são preferidos, opcionalmente tratados para terem uma superfície carregada negativamente (por exemplo, SDS) ou uma superfície carregada positivamente (por exemplo, com um detergente catiónico, tal como CTAB).

H. Lipossomas

Exemplos de formulações de lipossomas adequados para utilização como adjuvantes encontram-se descritos na Patente US 6090406, Patente US 5916588 e EP 0 626 169.

I. Formulações de éter de polioxietileno e de éster de polioxietileno

Os adjuvantes adequados para uso na invenção incluem os éteres de polioxietileno e ésteres de polioxietileno. W099/52549. Tais formulações incluem ainda tensioativos de ésteres polioxietileno de de sorbitano em combinação com um octoxinol (W001/21207), bem como éteres de polioxietileno de alquilo ou de éster de tensioativos em combinação com pelo menos um tensioativo não iónico adicional, tal como um octoxinol (W001/21152).

Éteres de polioxietileno preferidos são selecionados a partir do seguinte grupo: éter de polioxietileno-9-lauril (laureth 9), éter de polioxietileno-9-estearilo, éter de polioxietileno-8-estearilo, éter de polioxietileno-4-laurilo, éter de polioxietileno-35-laurilo, e éter de polioxietileno-23-laurilo.

J. Polifosfazeno (PCPP)

As formulações de PCPP são descritas, por exemplo, em Andrianov et al, "Preparation of hydrogel microspheres by coacervation of aqueous polyphosphazene solutions", Biomaterials (1998) 19 (1 a 3):109-115 e Payne et al., "Protein Release from Polyphosphazene Matrices", Adv. Drug. Delivery Review (1998) 31 (3):185-196.

K. Péptidos de muramilo

Exemplos de péptidos de muramilo adequados para uso como adjuvantes na invenção incluem N-acetil-muramil-L-treonil-D-isoglutamina (thr-MDP), N-acetil-normuramil-L-alanil-D-isoglutamina (nor-MDP), e MTP-PE).

L. Compostos de imidazoquinolona

Exemplos de compostos de imidazoquinolona adequados para uso como adjuvantes na invenção incluem Imiquamod e os seus homólogos, descrito mais adiante em Stanley, "Imiquimod and the imidazoquinolones: mechanism of action and therapeutic potential" Clin Exp Dermatol (2002) 27 (7):571- 577 e Jones, "Resiquimod 3M", Curr Opin Investig Drugs (2003) 4 (2):214-218.

A invenção pode também compreender combinações de aspectos de um ou mais dos adjuvantes identificados acima. Por exemplo, as seguintes composições de adjuvantes podem ser utilizadas na presente invenção:

(1) uma saponina e uma emulsão de óleo-em-água (W099/11241);

(2) uma saponina (por exemplo, QS21) + um derivado não-tóxico de LPS (por exemplo, 3dMPL) (ver WO 94/00153);

(3) uma saponina + um derivado não tóxico de LPS (por exemplo, 3dMPL) + um colesterol (por exemplo, QS21);

(4) uma saponina (por exemplo, QS21) + 3dMPL + IL-12 (opcionalmente + um esteroide) (W098/57659);

(5) combinações de 3dMPL com, por exemplo, QS21 e/ou emulsões de óleo-em-água (pedidos de patente europeia 0835318, 0735898 e 0761231);

(6) SAF, contendo 10% de esqualeno, 0,4% de Tween 80, 5% de polímero em bloco Pluronic L121-, e thr-MDP, quer microfluidizado numa emulsão submicrónica ou vortexado para gerar uma emulsão de maior tamanho de partícula.

(7) sistema adjuvante Ribi[™] (RAS), (Ribi Immunochem) contendo 2% de esqualeno, 0,2% de Tween 80, e um ou mais componentes da parede celular bacteriana a partir do grupo que consiste em monofosforilípido A (MPL), dimicolato de trealose (TDM), e estrutura da parede celular (CWS), preferencialmente MPL + CWS (Detox[™]); e

(8) um ou mais sais minerais (tal como um sal de alumínio) + um derivado não tóxico de LPS (tal como 3dPML).

(9) um ou mais sais minerais (tal como um sal de alumínio) + um oligonucleotídeo imunoestimulador (por exemplo, uma sequência de nucleótidos, incluindo um motivo CpG).

Os sais de alumínio e MF59 são adjuvantes preferidos para imunização parentérica. Toxinas bacterianas mutantes são adjuvantes preferidos de mucosa.

A composição pode incluir um antibiótico.

Outros antigénios

As composições da invenção podem ainda compreender um ou mais antigénios não GAS adicionais, incluindo antigénios bacterianos, virais ou parasitários adicionais.

Numa forma de realização, as combinações de antigénios GAS da invenção são combinadas com um ou mais antigénios não

GAS adicionais, adequados para utilização numa vacina pediátrica. Por exemplo, as combinações de antigénios GAS podem ser combinadas com um ou mais antigénios derivados de uma bactéria ou vírus selecionados entre o grupo que consiste em *N. meningitidis* (incluindo o sorogrupos A, B, C, W135 e/ou Y), *Streptococcus pneumoniae*, *Bordetella pertussis*, *Moraxella catarrhalis*, *Tetanus*, *Diphtheria*, vírus Sincicial Respiratório (RSV), poliomielite, sarampo, papeira, rubéola e rotavírus.

Numa outra forma de realização, as combinações de antigénios GAS da invenção são combinadas com um ou mais antigénios não-GAS adicionais, adequados para uso numa vacina concebida para proteger indivíduos idosos ou imunocomprometidos. Por exemplo, as combinações de antigénios GAS podem ser combinadas com um antigénio derivado de um grupo consistindo de *Enterococcus faecalis*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Pseudomonas aeruginosa*, *Legionella pneumophila*, *Listeria monocytogenes*, gripe, e vírus Parainfluenza ('PIV').

No caso de um antigénio de hidrato de carbono ou sacarídeo ser utilizado, ele é preferencialmente conjugado com uma proteína transportadora, a fim de aumentar a imunogenicidade {por exemplo, Ramsay et al. (2001) Lancet 357 (9251):195-196; Lindberg (1999) Vaccine 17 Suppl 2: S28-36; Buttery & Moxon (2000) JR Coll Médico Lond 34:163-168; Ahmad & Chapnick (1999) Infect Dis Clin North Am 13: 113-133, vii. Goldblatt (1998) J. Med. Microbiol. 47:563-567; Patente europeia 0 477 508; Patente US 5306492; WO98/42721; Conjugate Vaccines (eds. Cruse et al.) ISBN 3805549326, particularmente vol. 10:48-114; Hermanson (1996) Bioconjugate Techniques ISBN: 0123423368 ou

012342335X}. Proteínas de transporte preferidas são toxinas ou toxóides bacterianos, tais como de difteria ou do tétano. O toxóide diftérico CRM₁₉₇ é particularmente preferido {Research Disclosure, 453077 (Jan 2002)}. Outros polipeptídeos transportadores incluem a proteína da membrana externa de *N. meningitidis* {EP-A-0372501}, peptídeos sintéticos {EP-A-0378881 e EP-A-0427347}, proteínas de choque térmico {WO93/17712 e WO94/03208}, proteínas da tosse convulsa {WO98/58668 e EP-A-0471177}, Proteína D de *H. influenzae* {WO00/56360}, citocinas, {WO91/01146}, linfocinas, hormonas, fatores de crescimento, toxina A ou B de *C. difficile* {WO00/61761}, proteínas de captação de ferro {WO01/72337}, etc. Quando uma mistura compreende sacarídeos capsulares de ambos os serogrupos A e C, pode ser preferível que a proporção (p/p) de sacarídeo Mena: sacarídeo MenC seja maior do que 1 (por exemplo, 2:1, 3:1, 4:1, 5:1, 10:1 ou superior). Diferentes sacarídeos podem ser conjugados com o mesmo ou diferente tipo de proteína transportadora. Qualquer reação de conjugação adequada pode ser usada, com qualquer ligante adequado, quando necessário.

Antigénios de proteína tóxica podem ser desintoxicados, quando necessário, por exemplo, desintoxicação da toxina da tosse convulsa por meios químicos e/ou genéticos.

No caso de um antigénio da difteria ser incluído na composição, é preferível incluir também o antigénio do tétano e antigénios da tosse convulsa. Similarmente, onde um antigénio do tétano é incluído, é preferível incluir também os antigénios da difteria e da tosse convulsa. Similarmente, onde um antigénio da tosse convulsa é

incluído, é preferível incluir também os antígenos da difteria e do tétano.

Os antígenos na composição estarão tipicamente presentes numa concentração de pelo menos 1 µg/ml cada. Em geral, a concentração de um determinado antígeno irá ser suficiente para provocar uma resposta imune contra esse antígeno.

Como alternativa à utilização de antígenos de proteínas na composição da invenção, pode ser utilizado o ácido nucleico que codifica o antígeno {por exemplo, Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly et al. (1997) *Annu Rev Immunol* 15:617-648; Scott-Taylor & Dalglish (2000) *Expert Opin Investig Drugs* 9:471-480; Apostolopoulos & Plebanski (2000) *Curr Opin Mol Ther* 2: 441-447; Ilan (1999) *Curr Opin Mol Ther* 1: 116-120 Dubensky et al. (2000) *Mol Med* 6:723-732; Robinson & Pertmer (2000) *Adv Virus Res* 55:1-74; Donnelly et al. (2000) *Am J Respir Crit Care Med* 162 (4 Pt 2): S190-193; Davis (1999) *Mt. Sinai J. Med.* 66:84-90}. Componentes proteicos das composições da invenção podem, assim, ser substituídos por ácido nucleico (preferencialmente ADN, por exemplo, sob a forma de um plasmídeo) que codifica a proteína.

Definições

O termo "compreendendo" significa "incluindo", bem como "consistindo", por exemplo, uma composição "compreendendo" X pode consistir exclusivamente de X ou pode incluir algo adicional, por exemplo, X + Y.

O termo "acerca" em relação a um valor numérico x significa, por exemplo, $x \pm 10\%$.

As referências a uma percentagem de identidade de sequência entre duas sequências de aminoácidos significam que, quando alinhadas, essa percentagem de aminoácidos é a mesma, na comparação das duas sequências. Este alinhamento e a homologia percentual ou identidade de sequência podem ser determinados utilizando programas de software conhecidos na técnica, por exemplo, os descritos na secção de 7.7.18 de Current Protocols in Molecular Biology (FM Ausubel et al., Eds., 1987) suplemento 30. Um alinhamento preferido é determinado pelo algoritmo de pesquisa de homologia Smith-Waterman, utilizando uma pesquisa de lacuna afim com uma lacuna de penalidade aberta de 12 e uma penalidade de extensão de lacuna de 2, matriz BLOSUM de 62. O algoritmo de pesquisa de homologia de Smith-Waterman é divulgado em Smith & Waterman (1981) Adv. Appl. Math. 2:482-489. Métodos de identidade de sequência semelhantes podem ser utilizados para determinar a homologia de sequências entre duas sequências de polinucleótidos.

O exemplo seguinte demonstra uma forma de preparação de antígenos recombinantes GAS da invenção testando a sua eficácia num modelo de murino.

EXEMPLO 1: Preparação de antígenos recombinantes GAS e demonstração de eficácia em modelo de murino

Proteínas GAS recombinantes correspondentes a dois ou mais dos antígenos GAS do primeiro grupo de antígeno são expressas como se segue.

1. Clonagem de antígenos de GAS para expressão em E. coli

Os antígenos GAS seleccionados foram clonados de modo a obter dois tipos diferentes de proteínas recombinantes: (1) as proteínas com uma marca de hexa-histidina no terminal

carboxilo (GAS-His) e (2) as proteínas tendo o marcador de hexa-histidina no terminal carboxi e de GST no terminal amino (Gst-GAS-His). As proteínas do Tipo (1) foram obtidas por clonagem num vector pET21b+ (disponível a partir de Novagen). As proteínas do Tipo (2) foram obtidas por clonagem de um vector pGEX-NND. Esta estratégia de clonagem permitiu ao ADN genómico de GAS ser utilizado para amplificar os genes seleccionados por PCR, para realizar uma digestão de enzima de restrição única dos produtos de PCR e cloná-los, em seguida, simultaneamente em ambos os vectores.

(a) Construção de vectores de expressão pGEX-NNH

Dois pares de oligodesoxirribonucleótidos complementares são sintetizados utilizando o sintetizador de ADN ABI394 (Perkin Elmer) e os reagentes de Cruachem (Glasgow, Escócia). Quantidades equimolares dos pares de oligo (50 ng de cada oligo) são recozidos em tampão de ligase do ADN T4 (New England Biolabs) durante 10 min num volume final de 50 µl e, em seguida, deixados arrefecer lentamente à temperatura ambiente. Com o procedimento descrito obtêm-se os seguintes ligantes de ADN:

Ligante gexNN

```

      NdeI  NheI  KmaI  EcoRI  NcoI      SalI      XhoI      SacI
GATCCCATATGGCTAGCCCGGGGAATTCGTCATGGAGTGAGTCGACTGACTCGAGTGATCGAGCTC
GGTATACCGATCGGGCCCCCTTAAGCAGGTACCTCACTCAGCTGACTGAGCTCACTAGCTCGAG

```

```

      NotI
CTGAGCGGCCCGCATGAA
GACTCGCCGGCGTACTTTCCA

```

Ligante gexNNH

HindIII	NotI	XhoI	Hexa-Histidine
TCGACAAGCTT	GCGGCCG	CACTCGAG	CATCACCATCACCATCACTGAT
GTT	CGAACG	CCGCGT	GAGCACGTAGAGGTAGTGGTACTGACTATCGA

O plasmídeo pGEX-KG [KL Guan e JE Dixon, Canal Biochem. 192, 262 (1991)] foi digerido com BamHI e HindIII e 100 ng foi ligado durante a noite a 16°C para o ligante gexNN com uma razão molar de 3:1 de ligante/plasmídeo usando 200 unidades de ligase de ADN T4 (New England Biolabs). Após a transformação do produto de ligação em *E. coli* DH5, um clone contendo o plasmídeo pGEX-NN, tendo o ligante correto, é selecionado por meio da análise de enzima de restrição e sequenciação de ADN.

O novo plasmídeo pGEX-NN é digerido com Sall e HindIII e ligado ao ligante gexNNH. Após a transformação do produto de ligação em *E. coli* DH5, um clone contendo o plasmídeo pGEX-NNH, tendo o ligante correto, é selecionado por meio da análise de enzima de restrição e sequenciação de ADN.

(b) Reparação do ADN cromossómico

A estirpe GAS SF370 é cultivada em meio THY até OD₆₀₀ ser 0,6-0,8. As bactérias são então centrifugadas, suspensas em tampão de TES com lisozima (10 mg/ml) e mutanolisina (10 U/mL) e incubadas 1 h a 37°C. Após o tratamento da suspensão bacteriana com ARNase, Proteinase K e 10% de sarcosil/EDTA, é levada a cabo a extração de proteínas com fenol saturado e fenol/clorofórmio. O sobrenadante resultante é precipitado com acetato de sódio/etanol e o ADN extraído é sedimentado por centrifugação, suspenso em tampão Tris e mantido a -20°C.

(C) Concepção de *oligonucleotídeo*

Iniciadores sintéticos de oligonucleotídeos são projetados com base da sequência de codificação de cada antigénio GAS usando a sequência da estirpe SF370 M1de *Streptococcus pyogenes*. Qualquer péptido de sinal previsto é omitido, por deduzir a sequência de iniciador de amplificação da extremidade 5' imediatamente a jusante da sequência líder prevista. Para a maioria dos antigénios GAS, a cauda 5' dos iniciadores (ver Tabela 1, abaixo) inclui apenas um sítio de reconhecimento de enzima de restrição (NdeI, ou NheI ou SpeI, dependendo do próprio padrão de restrição do gene); as caudas 3' do iniciador (ver Tabela 1) incluem um local de restrição XhoI ou NotI ou HindIII.

5' tails		3' tails	
NdeI	5' GTGCGT CATATG 3'	XhoI	5' GCGT CTCGAG 3'
NheI	5' GTGCGT GCTAGC 3'	NotI	5' ACTCGCTAG CGGCCGC 3'
SpeI	5' GTGCGT ACTAGT 3'	HindIII	5' GCGT AAGCTT 3'

Tabela 1. Caudas oligonucleotídicas dos iniciadores utilizados para amplificar os genes que codificam os antigénios GAS selecionados.

Assim como contendo as sequências de reconhecimento de enzimas de restrição, os iniciadores incluem os nucleótidos que hibridam com a sequência a ser amplificada. O número de nucleótidos que hibridam depende da temperatura de fusão dos iniciadores que pode ser determinada como descrito [(Breslauer et al. Proc. Nat. Acad. Sci. 83, 3746-50 (1986))]. A temperatura média de fusão dos oligos selecionados é de 50-55°C, durante a região de hibridação isolada e 65-75°C para oligos inteiros. Os oligos podem ser comprados em MWG-Biotech SpA (Firenze, Itália).

(d) Amplificação por PCR

O protocolo de PCR padrão é como se segue: 50 ng de ADN genómico são utilizadas como modelo na presença de 0,2 µM de cada iniciador, 200 µM de cada dNTP, 1,5 mM de MgCl₂, 1x tampão de PCR menos Mg (Gibco-BRL), e 2 unidades de polimerase Taq ADN (Platinum Taq, Gibco-BRL) num volume final de 100 µl. Cada amostra passa por um passo de amplificação duplo: os primeiros 5 ciclos são realizados utilizando-se a temperatura de hibridação de um dos oligos excluindo a enzimas de restrição da cauda, seguido de 25 ciclos realizados de acordo com a temperatura de hibridação de todo o comprimento dos iniciadores. Os ciclos de padrão são os seguintes: um ciclo:

desnaturação: 94°C, 2 min,

5 cycles:

denaturation: 94 °C, 30 seconds,
hybridization: 51 °C, 50 seconds,
elongation: 72 °C, 1 min or 2 min and 40 sec, }

25 cycles:

denaturation: 94 °C, 30 seconds,
hybridization: 70 °C, 50 seconds,
elongation: 72 °C, 1 min or 2 min and 40 sec, }

72 °C, 7 min,
4 °C

O tempo de alongamento é de 1 min para antígenos GAS codificados por ORF mais curtos de 2000 pb e 2 min e 40 segundos para ORF maiores do que 2000 pb. As amplificações são realizadas usando um sistema Gene Amp PCR 9600 (Perkin Elmer).

Para verificar os resultados das amplificações, 4 µL de cada produto de PCR é carregado em gel de agarose a 1-1,5 e o tamanho dos fragmentos amplificados comparado com padrões de peso molecular (marcadores de ADN III ou IX, Roche). Os produtos de PCR são carregados em gel de agarose após electroforese e as bandas de tamanho adequado são excisadas do gel. O ADN é purificado a partir da agarose, utilizando o kit de extração de Gel (Qiagen) seguindo as instruções do fabricante. O volume de eluição final do ADN é de 50 µl de TE (10 mM Tris-HCl, EDTA 1 mM, pH 8). Um µl de cada um de ADN purificado é carregado em gel de agarose para avaliar a produtividade.

(e) Digestão de fragmentos de PCR

Um ou dois µg de produtos de PCR purificados são duplamente digeridos durante a noite a 37°C com as enzimas de restrição apropriadas (60 unidades de cada enzima) utilizando o tampão de restrição apropriado no volume final de 100 µl. As enzimas de restrição e os tampões de digestão são da New England Biolabs. Após a purificação do ADN digerido (kit de purificação de PCR, Qiagen) e eluição com 30 µl TE, 1 µl é submetido a electroforese em gel de agarose para avaliar o rendimento em comparação com os padrões de peso molecular (marcadores de ADN titulados III ou IX, Roche).

(f) Digestão dos vectores de fecho (pET21b + e pGEX-NND)

10 µg de plasmídeo foi duplamente digerido com 100 unidades de cada enzima de restrição em 400 µl de volume de reação, na presença de tampão adequado por incubação durante a noite a 37°C. Após electroforese num gel de agarose a 1%, a banda correspondente ao vector digerido foi purificado a partir do gel utilizando o Kit Qiagen Qiaex II Gel

Extraction e o ADN foi fluido com 50 µl TE. A concentração de ADN é avaliada medindo a OD₂₆₀ da amostra.

(g) Clonagem dos produtos de PCR

Setenta e cinco ng dos vectores apropriadamente digeridos e purificados e os fragmentos digeridos e purificados, correspondentes a cada um antigénio GAS seleccionado estão ligados em volumes finais de 10-20 mL com uma razão molar de 1:1 de fragmento/vector, com 400 unidades de ligase de ADN T4 (New England Biolabs) na presença do tampão fornecido pelo fabricante. As reacções são incubadas durante a noite a 16°C.

A transformação de células de *E. coli* BL21 (Novagen) e *E. coli* BL21-DE3 (Novagen) electrocompetentes é realizada utilizando ligações pGEX-NND e ligações pET21b+, respectivamente. O procedimento de transformação é como se segue: 1-2 µL da reacção de ligação é misturada com 50 µl de células competentes de gelo frio, em seguida, as células são despejadas num pulsador de genes de cuvette de eléctrodo de 0,1 cm (Biorad). Após pulsar as células num electroporador MicroPulser (Biorad), seguindo as instruções do fabricante, as células são suspensas em 0,95 ml de meio SOC e incubadas durante 45 min a 37°C, sob agitação. 100 e 900 µl de suspensões de células foram semeadas em placas de agar LB separadas de 100 µg/ml de ampicilina e as placas são incubadas durante a noite a 37°C. O rastreio dos transformantes é realizada por PCR: transformantes escolhidos aleatoriamente, foram repicadas e suspensas em 30 µl de mistura de reacção PCR contendo tampão de PCR, as quatro dNTP, 1,5 mM de MgCl₂, a polimerase Taq e iniciadores oligonucleotídicos para a frente e para trás apropriados, que são capazes de hibridar a montante e a

jusante do ligante poli dos vectores pET21b+ ou pGEX-NNH. Depois de 30 ciclos de PCR, 5 µl dos produtos resultantes são executados em electroforese em gel de agarose, de modo a seleccionar os clones positivos a partir dos quais a banda de PCR esperada é obtida. Os clones positivos de PCR são seleccionados com base no tamanho correto do produto de PCR, tal como avaliado por comparação com marcadores de peso molecular (marcadores de ADN III ou IX adequados, Roche).

2. Expressão proteica

Colónias PCR positivas são inoculados em 3 ml de LB de 100 µg/ml de ampicilina e cultivadas a 37°C durante a noite. 70 mL da cultura durante a noite foi inoculada em 2 ml de LB/Amp e cultivada a 37°C até a OD₆₀₀ dos clones pET atingir o valor de 0,4-0,8 ou até OD₆₀₀ dos clones pGEX atingir o valor de 0,8-1. A expressão da proteína é então induzida por adição de 1 mM de IPTG (isopropil β-D-tio-galacto piranoside) para as mini-culturas. Após 3 horas de incubação a 37°C, a OD₆₀₀ final é verificada e as culturas são arrefecidas em gelo. Após a centrifugação de 0,5 ml de cultura, o sedimento de células é suspenso em 50 µl de tampão de amostra de proteína (60 mM Tris-HCl pH 6,8, 5% p/v de SDS, 10% v/v de glicerina, 0,1% p/v de azul de bromofenol, 100 mM de DTT) e incubados a 100°C durante 5 min. Um volume de amostra fervida correspondente a 0,1 OD₆₀₀ da cultura foi analisado por SDS-PAGE e coloração com Azul de Coomassie para verificar a presença da banda de proteína induzida.

3. Purificação das proteínas recombinantes

As colónias simples são inoculadas em 25 ml de LB de 100 µg/ml de ampicilina e cultivadas a 37°C durante a noite. A cultura durante a noite foi inoculada em 500 ml de LB/Amp e

cultivada sob agitação a 25°C até OD₆₀₀ 0,4-0,7. A expressão da proteína é então induzida por adição de 1 mM de IPTG para as culturas. Após 3,5 horas de incubação a 25°C, a OD₆₀₀ final é verificada e as culturas são arrefecidos em gelo. Após centrifugação a 6000 rpm (rotor JA10 Beckman), o sedimento de células foi processado para purificação ou congelado a -20°C.

(a) Processo para a purificação de proteínas solúveis marcadas com His de *E. coli*

(1) Transferir o sedimento a partir de -20°C para um banho de gelo e reconstituí-los com 10 ml de tampão de 50 mM de NaHPO₄, 300 mM de NaCl, pH 8,0, passar em tubos de centrifugação de 40-50 ml e romper as células de acordo com o seguinte contorno.

(2) Quebrar o sedimento na prensa francesa realizando três passagens com lavagem em linha.

(3) Centrifugar cerca de 30-40000 xg por 15-20 min. Se possível, usar o rotor JA 25,50 (21000 rpm, 15 min.) ou JA-20 (18000 rpm, 15 min.)

(4) Equilibrar as colunas Poly-Prep com 1 ml de resina Fast Flow Chelating Sepharose com 50 mM de tampão fosfato, 300 mM de NaCl, pH 8,0.

(5) Conservar o sedimento de centrifugação a -20°C, e carregar o sobrenadante em colunas.

(6) Recolher o fluxo.

(7) Lavar as colunas com 10 ml (2 ml + 2 ml + 4 ml) de 50 mM de tampão fosfato, 300 mM de NaCl, pH 8,0.

(8) Lavar novamente com 10 ml de tampão de 20 mM de imidazole, 50 mM de fosfato, 300 mM de NaCl, pH 8,0.

(9) eluir as proteínas ligadas às colunas com 4,5 ml (1,5 ml + 1,5 ml + 1,5 ml) a 250 mM de tampão de imidazol, 50 mM de fosfato, 300 mM de NaCl, pH 8,0 e recolher as 3 fracções correspondentes 1,5 ml cada. Adicionar a cada tubo 15 µL de 200 mM de DTT (concentração final 2 mM)

(10) Medir a concentração de proteína das duas primeiras fracções com o método de Bradford, recolher uma alíquota de 10 µg de proteínas de cada uma das amostras e analisar por SDS-PAGE. (NB: a amostra deve estar demasiado diluída, carga de 21 µl + 7 µl tampão de carga).

(11) Armazenar as fracções recolhidas a +4°C, enquanto se espera pelos resultados da análise de SDS-PAGE.

(12) Para a imunização preparar 4-5 alíquotas de 100 µg cada, em 0,5 ml de glicerol a 40%. O tampão de diluição é o tampão de eluição anterior, mais 2 mM de DTT. Armazenar as alíquotas a -20°C até à imunização.

(b) *Purificação de proteínas marcadas com His a partir de corpos de inclusão*

As purificações são efectuadas essencialmente de acordo com o protocolo seguinte:

(1) As bactérias são recolhidas a partir de culturas de 500 ml por centrifugação. Se necessário, armazenar os sedimentos bacterianos a -20°C. Para a extração, ressuspender cada sedimento bacteriano em 10 ml de tampão de 50 mM de TRIS-HCl, pH 8,5 num banho de gelo.

(2) Dissociar as bactérias ressuspensas com uma prensa

francesa, realizando duas passagens.

(3) Centrifugar a 35000 xg durante 15 minutos e recolher os sedimentos. Usar um rotor de Beckman JA 25,50 (21000 rpm, 15 min.) ou JA-20 (18000 rpm, 15 min.).

(4) Dissolver os sedimentos de centrifugação com 50 mM de Tris-HCl, 1 mM de TCEP {Tris (2-carboxietil) fosfina cloridrato, Pierce}, 6 M de cloreto de guanidínio, pH 8,5. Agita-se durante ~10 min. com uma barra magnética.

(5) Centrifugar como descrito acima, e recolher o sobrenadante.

(6) Preparar um número adequado de colunas Poly-Prep (Bio-Rad) contendo 1 ml de Fast Flow Chelating Sepharose (Pharmacia) saturado com Nichel de acordo com as recomendações do fabricante. Lavar as colunas duas vezes com 5 ml de H₂O e equilibrar com 50 mM de TRIS-HCl, 1 mM de TCEP, 6 M de cloreto de guanidínio, pH 8,5.

(7) Carregar os sobrenadantes a partir do passo 5 para as colunas, e lavar com 5 ml de tampão de 50 mM Tris-HCl, 1 mM de TCEP, 6 M de ureia, pH 8,5

(8) Lavar as colunas com 10 ml de 20 mM de imidazol, 50 mM de TRIS-HCl, 6 M de ureia, 1 mM de TCEP, pH 8,5. Recolher e reservar os primeiros 5 ml para eventuais controles.

(9) Eluir as proteínas ligadas às colunas com 4,5 ml de um tampão contendo 250 mM de imidazol, 50 mM de TRIS-HCl, 6 M de ureia, 1 mM de TCEP, pH 8,5. Adicionar o tampão de eluição em três aliquotas de 1,5 ml, e recolher as 3 frações correspondentes. Adicionar a cada fração 15 µl de DTT (concentração final 2 mM).

(10) Medir a concentração de proteína eluída com o método

de Bradford, e analisar as aliquotas de 10 µg de proteína por SDS-PAGE.

(11) Armazenar as proteínas a -20°C em 40% (v/v) de glicerol, 50 mM de Tris-HCl, 2 M de ureia, 0,5 M de arginina, 2 mM de DTT, 0,3 mM de TCEP, 83,3 mM de imidazol, pH 8,5.

(c) Processo para a purificação de proteínas de fusão GST a partir de *E. coli*

(1) Transferir o sedimento bacteriano a -20°C para um banho de gelo e suspender com 7,5 ml de PBS, pH 7,4, ao qual foi adicionado uma mistura de inibidores de protease (COMPLETE™ - Boehringer Mannheim, um comprimido a cada 25 ml de tampão).

(2) Transferir para tubos de de centrifugação de 40-50 ml e sonicar de acordo com o seguinte procedimento:

a. Posicionar a sonda a cerca de 0,5 cm a partir do fundo do tubo

b. Bloquear o tubo com o grampo

c. Mergulhar o tubo num banho de gelo

d. Definir o sonicador da seguinte forma: Temporizador → Hold, Ciclo de Trabalho → 55, out. Controle → 6.

e. Realizar 5 ciclos de 10 impulsos num lapso de tempo de 1 minuto (ou seja, um ciclo = 10 impulsos + ~45" Hold; b. 10 impulsos + ~45" Hold; c. 10 impulsos + ~45" Hold; d. 10 impulsos + ~45" hold; e. 10 impulsos + ~45" Hold).

(3) Centrifugar a cerca de 30-40000 xg durante 15-20 min.

Por exemplo: usar um rotor Beckman JA 25,50 a 21000 rpm, por 15 min.

(4) Armazenar os sedimentos de centrifugação à temperatura de -20°C , e carregar os sobrenadantes sobre as colunas de cromatografia, como a seguir

(5) Equilibrar a colunas Poly-Prep (Bio-Rad) com 0,5 ml (suspensão $\cong 1$ ml) de resina de Sepharose 4B-glutationala, lavar com 2 ml (1 + 1) de H_2O , e, em seguida, com 10 ml (2 + 4 + 4) de PBS, pH 7,4.

(6) Colocar o sobrenadante sobre as colunas e descartar o fluxo.

(7) Lavar as colunas com 10 ml (2 + 4 + 4) de PBS, pH 7,4.

(8) Eluir as proteínas ligadas às colunas com 4,5 ml de tampão de 50 mM de TRIS, 10 mM de glutationala reduzida, pH 8,0, adicionar 1,5 ml + 1,5 ml + 1,5 ml e recolher as respectivos três fracções de $\sim 1,5$ ml cada.

(9) Medir a concentração de proteína das duas primeiras fracções com o método de Bradford, analisar uma alíquota de 10 μg de proteínas de cada uma das amostras por SDS-PAGE. (NB: se a amostra é muito diluída carregar 21 μL (+ 7 μL de tampão de carregamento)).

(10) Armazenar as fracções recolhidas a $+4^{\circ}\text{C}$, enquanto se espera pelos resultados da análise de SDS-PAGE.

(11) Para cada uma das proteínas destinadas para a imunização preparar 4-5 alíquotas de 100 μg cada, em 0,5 ml de glicerol a 40%. O tampão de diluição é de 50 mM de Tris.HCl, 2 mM de DTT, pH 8,0. Armazenar as alíquotas a -20°C até à imunização.

4. Modelo murino de proteção contra infecções GAS

(a) Protocolo de imunização

Grupos de 10 ratinhos fêmea CD1 com idades compreendidas entre 6 e 7 semanas são imunizados com dois ou mais antigénios GAS, (20 µg de cada antigénio GAS recombinante), em suspensão em 100 µl de uma solução adequada. Cada grupo recebe três doses nos dias 0, 21 e 45. A imunização é realizada por meio de injeção intra-peritoneal da proteína com um volume igual de adjuvante completo de Freund (CFA) para a primeira dose e adjuvante Incompleto de Freund, (IFA) para as duas seguintes doses. Em cada esquema de imunização são utilizados grupos de controlo positivo e negativo.

Para o grupo de controlo negativo, os ratos são imunizados com proteínas de *E. coli* eluídas das colunas de purificação seguintes do processamento de extracto total de bactérias de uma estirpe de *E. coli* contendo o vector pET21b ou pGEX-NNH (expressando assim apenas GST) sem qualquer GAS ORF clonado (grupos podem ser indicados como HisStop ou GSTStop respectivamente).

Para os grupos de controlo positivo, os ratos são imunizados com GAS M purificado clonado a partir de qualquer estirpe GAS SF370 ou GAS DSM 2071 (grupos indicados como 192SF e 192DSM respectivamente).

Os soros obtidos a partir de cada grupo são recolhidos antes da primeira imunização e duas semanas após a última. Os ratos são infectadas com o GAS cerca de uma semana depois.

Os ratinhos imunizados são infectados com uma estirpe GAS diferente da que é utilizada para a clonagem de proteínas selecionadas. Por exemplo, a estirpe de GAS pode ser do tipo DSM 2071 M23, disponível na Coleção Alemã de Microrganismos e Culturas Celulares (DSMZ).

Para as experiências de infecção, DSM 2071 é cultivado a 37°C em caldo THY até OD₆₀₀ 0,4. As bactérias são sedimentadas por centrifugação, lavadas uma vez com PBS, suspensas e diluídas com PBS para obter a concentração adequada de bactérias/ml e administradas a ratos por injeção intraperitoneal. Entre 50 e 100 bactérias são dadas para cada rato, como determinado através do plaqueamento de alíquotas da suspensão bacteriana em 5 placas de THY. Os animais são observados diariamente e verificados para a pela sobrevivência.

5. Análise de soros Imunes

(a) Preparação de extractos de proteína total GAS

Extractos de proteínas totais, são preparados por incubação de uma cultura bacteriana cultivada a OD₆₀₀ de 0,4-0,5 em 50 mM de Tris, pH 6,8/mutanolisina (20 unidades/ml) durante 2 horas a 37°C, seguido por incubação durante 10 minutos sobre gelo em 0,24 N de NaOH e 0,96% de β-mercaptoetanol. As proteínas extraídas são precipitados por adição de ácido tricloroacético, lavadas com acetona gelada e suspensas em tampão de carga de proteína.

(b) Análise de Western blot

Alíquotas de extrato de proteína total misturado com tampão de carga SDS (1x: 60 mM TRIS-HCl pH 6,8, 5% p/v de SDS, 10% v/v de glicerina, 0,1% de azul de bromofenol, 100 mM de DTT) e fervido 5 minutos a 95°C, foram carregados num gel

de 12,5% de SDS-PAGE pré-moldado (Biorad). O gel é administrado utilizando um tampão de corrida de SDS-PAGE contendo 250 mM de Tris, 2,5 mM de glicina e 0,1% de SDS. O gel é electrotransferido para uma membrana de nitrocelulose a 200 mA durante 60 minutos. A membrana é bloqueada durante 60 minutos com PBS/0,05% de Tween-20 (Sigma), 10% de leite em pó desnatado e incubada O/N a 4°C com PBS/0,05% Tween 20, 1% de leite desnatado em pó, com a apropriada diluição dos soros. Depois de ser lavada duas vezes com PBS/0,05% de Tween, a membrana foi incubada durante 2 horas com anticorpo anti-ratinho secundário conjugado com peroxidase (Amersham) diluído a 1:4000. A nitrocelulose foi lavada três vezes durante 10 minutos com PBS/Tween a 0,05% e uma vez com PBS e, subsequentemente desenvolvida por Opti-4CN Substrate Kit (Biorad).

(c) *Elaboração de culturas GAS tratadas com paraformaldeído*

Uma cultura bacteriana cultivada a OD₆₀₀ 0,4-0,5 é lavada uma vez com PBS e concentrada quatro vezes em PBS/0,05% de paraformaldeído. Seguindo uma hora de incubação a 37°C com agitação, a cultura tratada é mantida durante a noite a 4°C e a inativação completa da bactéria é, então, controlada por plaqueamento de alíquotas em placas de agar de sangue THY.

(d) *Análise FACS de culturas GAS criadas em paraformaldeído com os soros imunes de rato*

Cerca de 10⁵ bactérias inativadas de paraformaldeído são lavadas com 200 µL de PBS numa placa de fundo de 96 poços em U e centrifugadas por 10 min. a 3000 g, a 4°C. O sobrenadante é rejeitado e as bactérias são suspensas em 20 µl de PBS-BSA a 0,1%. Oitenta µl de soros de rato pré-imunes ou imunes diluídos em PBS-BSA a 0,1% são adicionados

à suspensão bacteriana para uma diluição final a 1:100, 1:250 ou 1:500, e foram incubados em gelo durante 30 min. As bactérias são lavadas uma vez por adição de 100 µl de PBS-BSA a 0,1%, centrifugadas durante 10 min. a 3000 g, 4°C, em suspensão em 200 µl de PBS-BSA a 0,1%, novamente centrifugadas e suspensas em 10 µl de anticorpo de cabra anti-IgG de rato F (ab')₂ específico do fragmento de R-Ficoeritrina conjugada (Jackson ImmunoResearch Laboratories Inc., cat.N ° 115-116-072) em PBS-BSA a 0,1% a uma diluição final de 1:100, e incubadas em gelo durante 30 min. no escuro. As bactérias são lavadas uma vez por adição de 180 µl de PBS-BSA a 0,1% e centrifugadas durante 10 min. a 3000 g, 4°C. O sobrenadante é rejeitado e as bactérias foram suspensas em 200 µl de PBS. A suspensão bacteriana é passada através de uma câmara de citometria dum FACS Calibur (Becton Dickinson, Mountain View, CA EUA) e são adquiridos 10000 eventos. Os dados são analisados usando o software Cell Quest (Becton Dickinson, Mountain View, CA EUA) pelo desenho de um gráfico de pontos morfológicos (utilizando parâmetros de dispersão frontal e lateral) em sinais bacterianos. Um histograma é, então, criado em intensidade FL2 da escala logarítmica de fluorescência recordando a região morfológica das bactérias.

EXEMPLO 2: Comparação da virulência das estirpe de tipo selvagem de GAS (incluindo GAS 40) e exclusão do mutante de GAS 40

O exemplo a seguir apresenta uma comparação entre a virulência de uma estirpe de tipo selvagem GAS e mutante de deleção de GAS 40. As estirpes mutantes GAS onde a maioria da sequência de GAS 40 foi removida foram preparadas por métodos padrão. Os grupos de imunização de dez ratinhos por grupo foram injetados quer com o tipo selvagem ou as estirpes mutantes GAS. Como mostrado abaixo, a injeção de

uma gama de concentrações do tipo selvagem isolado de rato resultou em mortes, enquanto a injeção com o mutante GAS $\Delta 40$ não.

GAS strain	concentration	number of fatalities
wild type	2×10^5	10
wild type	2×10^6	9
wild type	2×10^7	5
GAS $\Delta 40$	2×10^2	0
GAS $\Delta 40$	2×10^3	0
GAS $\Delta 40$	2×10^4	0
GAS $\Delta 40$	2×10^5	0
GAS $\Delta 40$	2×10^6	0
GAS $\Delta 40$	2×10^7	0

EXEMPLO 3: Ensaio de opsonofagocitose bacteriana em construções GAS 40

O exemplo que se segue demonstra a exposição da superfície de GAS 40 por utilização de um ensaio de opsonofagocitose bacteriana. As seguintes construções de GAS, cada uma das quais está descrita em pormenor acima, foram usadas no ensaio: 40a-CH, 40a- RR-NH, 40a-RR, GST-40, 40a, 40a e 40a-NH. (As duas referências a "40a" na Figura 7, referem-se a soros preparados em dias diferentes.

O ensaio foi realizado como se segue.

1. Preparação do inoculo bacteriano. As bactérias GAS são cultivadas em meio de THY até que atinjam a fase exponencial média (OD_{600} de 0,4) a 37°C . As bactérias são lavadas duas vezes em solução salina gelada e são suspensas em meio HBSS com o volume ajustado para cada estirpe, dependendo da quantidade de bactérias que irá ser utilizada. As células bacterianas são mantidos em gelo até

à utilização.

2. Preparação de PMN. PMN são preparados a partir de crostas inflamatórias de sangue heparinizado de voluntários sadios. O creme leucocitário foi incubado durante 30 minutos numa solução contendo dextrano, NaCl e Heparina (taxa de 1:1). Após a incubação o sobrenadante, rico dos leucócitos, é removido, transferido num tubo limpo e centrifugado a 700xg durante 20 minutos. A uma curta lavagem em água é realizada para quebrar as células vermelhas do sangue e, em seguida, é adicionada uma solução de cloreto de sódio para restaurar a concentração de sal apropriado. Após este passo as células são centrifugadas, lavadas e suspensas em meio MEM a uma concentração apropriada.

3. Ensaio de opsonofagocitose. Estirpes GAS (preparados como descrito) são incubadas com soro imune inativado pelo calor de ratinhos derivados a partir da imunização com o antigénio GAS indicado (ou pré-imune para o controlo), PMN humanos e complemento de coelho bebé. Uma hora de incubação a 37°C. As amostras recolhidas imediatamente antes e após a incubação foram semeadas em placas de ágar de sangue THY. A fagocitose é avaliada comparando a diferença no número de colónias nos dois tempos para o soro pré-imune e imune. Os dados são apresentados como número de logaritmo de colónias crescidas em $t = 0$ - número de logaritmo de colónias crescidas em $t = 60$

Os resultados do ensaio são mostrados na Figura 7. O eixo Y indica uma diferença entre o logaritmo do número de colónias no tempo 0 e o logaritmo do número de colónias ao fim de 60 segundos: $\log (\text{CFU @ } T_0) - \log (\text{CFU @ } T_{60'})$. Se

houve um crescimento (isto é, as bactérias não são ativamente mortas), resulta em números negativos (barras negativos). Se as bactérias são mortas, resulta em números positivos (barras do histograma positivos). Como mostrado na Figura 7, são relatados as barras do histograma positivo, para cada uma das construções GAS. As últimas quatro barras amarelas na Figura 7 representam controles: **B** = bactérias isoladas, **B PMN** = Bactérias + polimorfonucleatos, **BC** = Bactérias + complemento, **P PMN C** = bactérias + + polimorfonucleatos + complemento (sem soro).

EXEMPLO 4: Experiências de desafio de imunização GAS 40 em modelo de proteção murino de rato

Estão listados abaixo exemplos da percentagem de sobrevivência resulta de numerosas experiências em modelo murino do rato, usando o antigénio GAS 40. As anotações indicam onde uma construção utilizada para expressar o antigénio recombinante GAS 40 foi modificada para facilitar a expressão.

GAS antigen	% Survival in Mouse Challenge Model
40a	55
40a-RR	70
40a-RR-NH	60

Listagem de Sequências

SEQ ID NO: 1 amino acid sequence comprising GAS 40

MDLEQTKPNQVKQKIALTSTIALLSASVGVSHQVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVE
KTLSQQKAEITELATALTKTTAEINHLKEQQDNEQKALTSAQEITYNTLASSEETLLAQGAHQRELTATE
TELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTKA
LSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRKSSAPSTQDSTVGNNTMKAPQGYPLEE
LKKLEASGYIGSASYNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFDPPDNLTPEVQNELAQFAAHMI
NSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGPHDKTTIEDSAGASGLI
RNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAINFLRVDKHNPNAPVYLGFSTSNV
GSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGVSSLENRLSAIHQEADIMA
AQAKVSQLOQKLASTLKQSDSLMLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLASLKAALHOT
EALAEQAAARVTALVAKKAHLQYLKDFKLNPNRLQVIRERIDNTKQDLAKTTSSLLNAQEALALQAKQSS
LEATIAATTEHQLTLLKTLANEKEYRHLDEDIATVPDLQVAPPLTGVKELSYSKIDTTELVQEMVKETKQLL
EASARLAAENTSLVAEALVGQTESEMVASNAIVSKITSSITQPSSKTSYSGSGSSTTSNLSIDVDESTQRAK
AGVVMMLAAVGLTGFRFRKESK

SEQ ID NO: 2 polynucleotide sequence encoding for GAS 40

ATGGACTTAGAACAAACGAAGCCAAACCAAGTTAAGCAGAAAATTGCTTTAACCTCAACAATTGCTTTATT
 GAGTGGCAGTGTAGGGGTATCTCACCAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTA
 ATACTCACGACGATAGTTTACCAAAACAGAAAACAATTCAAAGAGGCAAGGCAACTATTGGATGCAATTGAA
 AAAACTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAAT
 CAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATA
 CTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCCAGGAGCCGAACATCAAAGCAGAGTTAACAGCTACTGAA
 ACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCAT
 TTCAGCAGAACTACTCGAGCTCAAGATTTAGTGGAAACAAGTCAAAACGCTGTAACAAAATATTGCTAAGC
 TCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCA
 TTAAGCTCAGAAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGAC
 TGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGCTTTAAATCCTCAG
 CTCCGCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAA
 CTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGA
 TCAAATTAATGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATC
 GCTTTGTGTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGACGCTCATATGATT
 AATAGTGTAAAGAGACAATTAGGTCTACCACCACTACTGTTACAGCAGGATCACAAGAATTTGCAAGATT
 ACTTAGTACCAGCTATAGAAAACCATGGTAAATCAAGACCATCATTTGTCTACGGACAGCCAGGGGTAT
 CAGGGCATTATGGTGTGGGCCCTCATGATAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCAT
 CGAAATGATGATAACATGTACGAGAAATATCGGTGCTTTTAAACGATGTGCATACTGTGAATGGTATTAAACG
 TGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAATAACATACGCCCATGCTA
 TTAATCTTTTACGTTAGATAAACTAACCCCTAATGCCCTGTATTACCTTGGATTTTCAACCAGCAATGTA
 GGATCTTTGAATGAACACTTTGTAAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGAC
 CCCATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCATGTATCTGATACTATTGCAAG
 CGATCAAAGGAAAAGTAAGCTCATTAAGAAAATCGTTTGTGCGCTATTCAATCAAGAAAGCTGATATTATGGCA
 GCGCAAGCTTAAAGTAAGTCAACTTCAAGGTAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCT
 CCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAAATTACTAGCAGCTAAAGCAAAACAAG
 CACAACCTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACCTGCACCAGACA
 GAAGCCCTTAGCAGAGCAAGCCGCAGCCAGAGTGACAGCAGCTGGTGGCTAAAAAAGCTCATTTGCAATATCT
 AAGGGACTTTAAATTGAATCCTAACCCGCTTCAAGTGATACGTGAGCGCATTTGATAATACTAAGCAAGATT
 TGGCTAAAACCTACCTCATCTTTGTAAATGACACAGAAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGT
 CTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAACCTTAGCTAACGAAAAGGAATA
 TCGCCACTTAGACGAAGATATAGCTACTGTGCCGTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAAC
 CGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACATTTA
 GAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTGGCCAAACCTCTGA
 AATGGTAGCAAGTAATGCCATTCTGTCTAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTT
 ATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAAGAGCTCTTAAA
 GCAGGAGTCGTCATGTTGGCAGCTGTCCGCCCTCAGGATTTAGGTTCCGTAAGGAATCTAAGTGA

SEQ ID NO: 3 amino acid sequence comprising an N terminal leader sequence of GAS 40 MDLEQTKPNQVKQ-
 KIALTSTIALLSA

SEQ ID NO: 4 polynucleotide sequence encoding an N terminal leader sequence of GAS 40

ATGGACTTAGAACAAACGAAGCCAAACCAAGTTAAGCAGAAAATTGCTTTAACCTCAACAATTGCTTTATT
 GAGTGGC

SEQ ID NO: 5 amino acid sequence comprising a fragment of GAS 40 with N terminal leader sequence removed

SVGVSHQVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVEKTLSSQKAELTELATALTKTTAEINH
 LKEQQDNEQKALTSAGEIYTNILASSEETLLAQGAHQRELTAETELHNAQADQHSKETALSEQKASISA
 ETTRAQDLVEQVKTSEONIAKLNAMISNPDAITKAAQTANDNTKALSSELEKAKADLENQAKVKKQLTEE
 LAAQKAALAEKEAELSRKSSAPSTQDSIVGNNTMKAPQGYPLEELKKLEASGYIGSASYNNFYKEHADQI
 IAKASPGNQLNQYQDIPADRNRFVDPDNLTPVQNELAQFAAHMINSVRRQLGLPPVTVTAGSQEFARLLS
 TSYKKTGHNTSRPSFVYQPGVSGHYGVGPHDKTIIEDSAGASGLIRNDDNMYENIGAFNDVHTVNGIKRGI
 YDSIKYMLFTDHLHGNTYGHAINFLRVDKHNPAPVYLGFSTSNVGSLSNEHFVMFFESNIANHQRFNKTP
 KAVGSTKDYAQRVGTIVSDTIAAIKGVSSLENRLSAIHQEADIMAAQAKVSQLOQKLASTLQSDSLNLQV
 RQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRD
 FKLNPRLQVIRERIDNTKQDLAKTTSSLLNAQEALAAALQAKQSLEATIATTEHQLTLLKTLANEKEYRH
 LDEDIATVPDLQVAPPLTGKPLSYSKIDTTPLVQEMVKETKQLEASARLAAENTSLVAEALVGQTSSEM
 ASNAIVSKITSSITQPSKTSYSGSGSSTTSLNLSVDDESTQRLKAGVVMMLAAVGLTGFRFRKESK

SEQ ID NO: 6 polynucleotide sequence encoding a fragment of GAS 40 with N terminal leader sequence removed

AGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATACTCA
 CGACGATAGTTTACCRAAACAGAAACAATTCAAGAGGCAAGGCAACTATTGATGCAGTTGAAAAAAGCTC
 TCAGTCAACAAAAAGCAGAACTGACAGAGCTTGTCTACCGCTCTGCACAAAACTACTGCTGAAATCAACCAC
 TTAAGAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAAGAAATTTACACTAATACTCTTGC
 AAGTAGTGAGGAGACGCTATTAGCCCCAAGGAGCCGAACATCAAGAGAGTTAACAGCTACTGAAACAGAGC
 TTCATAATGCTCAAGCAGATCAACATTTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCATTTCAGCA
 GAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGCTCTGAACAAAATATTGCTAAGCTCAATGC
 TATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTAAAGCT
 CAGAAATTGGAGAAGGCTAAAGCTGACTTAGAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAAGAG
 TTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGAGCAACTTAGTCGCTCTTAAATCCTCAGCTCCGTC
 TACTCAAGATAGCATTTGTGGGTAATAATACCATGAAAGCACCAGCAAGGCTATCCTCTTGAAGAAGCTTAAAA
 AATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAAAAT
 AATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTTTGT
 TGATCCCGATAAATTTGACACCAGAAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAATAGTG
 TAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGATTACTTAGT
 ACCAGCTATAAGAAAACCATGTTAATACAAAGACCATCAATTTGCTTACGAGCAGCCAGGGGTATCAGGGCA
 TTATGGTGTGGGCTCATGATAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCATTCGAAATG
 ATGATAACATGTACGAGAATATCGGTGCTTTTAAAGATGTGCTACTGTGAATGGTATTAAACGTTGGTATT
 TATGACAGTATCAAGTATATGCTCTTTACAGATCATTTTACACGGAATAACATACGCGCCATGCTATTAACTT
 TTTACGTGTAGATAAACATAACCCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTAGGATCTT
 TGAATGAACACTTTTGAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCTATA
 AAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCAGCGATCAA
 AGGAAAAGTAAGCTCATTTAGAAAATCGTTTGTGCGCTATTTCATCAAGAAGCTGATATATGGCAGCCCAAG
 CTAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAAGTG
 AGACAATTAATGATACTAAAGGTTCTTTGAGAACAGAACTTACTAGCAGCTAAAGCAAAACAAGCACAACT
 CGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAAGCCT
 TAGCAGAGCAAGCCGACAGCAGAGTGACAGCACTGGTGGCTAAAAAGCTCATTTCGAATATCTAAGGGAC
 TTTAAATTTGAATCTTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATTGCGCTAA
 AACTACCTCATCTTTGTTAAATGCACAAGAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGTCTAGAAG
 CTACTATTGCTACCACAGAACACAGTTGACTTTGCTTAAACCTTAGCTAACGAAAAGGAATATCGCCAC
 TTAGACGAAGATATAGTACTGTGCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAACCGCTATC
 ATATAGTAAGATAGATACTACTCCGCTTGTTCAGAAATGGTTAAAGAAACGAAACAACCTATTAGAAGCTT
 CAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAAGCGCTTGTGGCCAAACCTCTGAAATGGTA
 GCAAGTAATGCCATTGTGTCTAAATTCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGGCTC

AGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAAGAGCTCTTAAAGCAGGAG
 TCGTCATGTTGGCAGCTGTGCGCCTCACAGGATTTAGGTTCCGTAAGGAATCTAAGTGA

SEQ ID NO: 7 amino acid sequence comprising a C terminal transmembrane region of GAS 40 ALKAGVVM-
 LAAVGLTGFRFRKESK

SEQ ID NO: 8 polynucleotide sequence encoding a C terminal transmembrane region of GAS 40

GCTCTTAAAGCAGGAGTCGTTCATGTTGGCAGCTGTGCGCCTCACAGGATTTAGGTTCCGTAAGGAATCTAA
GTGA

SEQ ID NO: 9 amino acid sequence comprising a fragment of GAS 40 with a C terminal transmembrane sequence removed

MDLEQTKPNQVKQKIALTSTIALLSASVGVSHQVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVE
KTLSSQQAELTELATALTKTTAEINHLKEQODNEQKAL TSAQEIYTN TLASSEETLLAQGAEHORELTATE
TELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKRAQTANDNTKA
LSSELEKAKADLENQKAKVKQLTEELAAQKAALAEKEAEL SRLKSSAPSTQDSIVGNNTMKAFQGYPLEE
LKKLEASGYIGSASYNNYKEHADQIIAKASPGNQLNQYQDIPADRNRFPVDPDNLTFEVQNELAQFAAAMI
NSVRRQLGLPPTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGPHDKTIIEDSAGASGLI
RNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAINFLVVDKHNPNAPVYLGFTSTNV
GSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGVSSLENRLSAIHQEADIMA
AQAKVSQLOGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLASLKAALHQ
EALAEQAAARVTALVAKKAHLQYL RDKLNFENRLQVIRERIDNTKQDLARTSSLLNAQELALALQAKQSS
LEATIATTEHQLTLLKTLANEKEYRHLDEDIATVPDLQVAPPLTGKPLSYSKIDTTPLVQEMVKETKQLL
EASARLAAENTSLVAELVGQTSSEMVASNAIVSKITSSITQPSKTSYSGSGSTTSLNLSVDVDESTQR

SEQ ID NO: 10 polynucleotide sequence encoding a fragment of GAS 40 with a C terminal transmembrane sequence removed

ATGGACTTAGAACAAACGAAGCCAAACCAAGTTAAGCAGAAAAATTGCTTTAACCTCAACAATTGCTTTATT
GAGTGCCAGTGTAGGCGTATCTCACCAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTA
ATACTCAGCAGATAGTTTACCAAAACCAAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGTCAGTTGAA
AAAACCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAAT
CAACCACTTAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATA
CTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAGAGAGTTAACAGCTACTGAA
ACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTTGTCAGAAACAAAAAGCTAGCAT
TTCAGCAGAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGCTCTGAACAAAAATATTGCTAAGC
TCAANTGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCA
TTAAGCTCAGAAATGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAGCTAAAGTTAAAGCAATTGAC
TGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGCTTTAAATCCTCAG
CTCCGCTACTCAAGATAGCATTGTGGCTAATAATACCATGAAAGCACCAGCAAGGCTATCCTCTTGAAGAA
CTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTTATTACAAAGAGCATGCAGA
TCAATTTATTGCCAAAGCTAGTCCAGGTAATCAATTAATCAATACCAAGATATTCCAGCAGATCGTAATC
GCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGACAGCTCACATGATT
AATAGTGTAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAGAATTTGCAAGATT
ACTTAGTACCAGCTATAAGAAAACCATGGTAATACAGACCATCATTTGTCTACGGACAGCCAGGGGTAT
CAGGGCATTATGGTGTGGGGCTCATGATAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCAT
CGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACG
TGGTATTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGCTA
TTAACTTTTTACGTGTAGATAAAACATAACCCCTAATGCGCCTGTTTACCCTGGATTTCACACAGCAATGTA
GGATCTTTGAATGAACACTTTGTAATGTTTCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGAC
CCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCGAG
CGATCAAAGGAAAAGTAAAGCTCATTAGAAAATCGTTTGTGCGCTATTCAATCAAGAAGCTGATATTATGGCA
GCCAAGCTAAAGTAAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCT
CCAAGTGAGACAATTAAATGATACTAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAG
CACAACCTGAAGCTACTCGTGATCAATCATTAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACA
GAAGCCTTAGCAGAGCAAGCCGAGCCAGAGTGACAGCACTGGTGGCTAAAAAGCTCATTTCGAATATCT
AAGGGACTTTAAATTTGAATCCTAACCGCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATT

TGGCTAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGT
CTAGAAGCTACTATTGCTACACAGAACACCAGTTGACTTTGCTTAAACCTTAGCTAACGAAAAGGAATA
TCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTGCAAGTAGCTCCACCTCTTACGGGGCTAAAC
CGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAGAAATGCTTAAAGAAACGAAACAACTATTA
GAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTGGCCAAACCTCTGA
AATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTT
ATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAAGA

SEQ ID NO: 11 amino acid sequence comprising a transmembrane region of GAS 40 as shown in Figures 1 and 2. ALKAGVVMLAAVGLTG

SEQ ID NO: 12 amino acid sequence comprising a first coiled-coil region of GAS 40

ETIQEAKATIDAVEKTLSSQKAELELATALTKTTAEINHLKEQDNEQKALTSAQEIYTNLTASSEETLL
AQGAHQRELATETELHNAQADQHSKETALSEQKASISAETTRAQDLVBQVKTSEQNIAKLNAMISNPDA
ITKAAQTANDNTKALSSLEKAKADLENQKAKVKQLTEELAAQKALAEKEALSRLKSSA

SEQ ID NO: 13 amino acid sequence comprising a second coiled-coil region of GAS 40

RLSAIHQEADIMAAQAKVSQLQGLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLETRDQSL
AKLASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLPNRLQVIRERIDNTKQDLAKTSSLLNA
QEALALQAKQSSLEATTIATTEHQLTLLKTLANEKE

SEQ ID NO: 14 amino acid sequence comprising a leucine zipper motif within the second coiled-coil region of GAS 40. QVIRFRIDNTKQDL2KTTSSLLNAQEALAAAL

SEQ ID NO: 15 amino acid sequence comprising SpA from *Streptococcus gordonii* Genbank reference G125990270

MNKRKEVFGFRKSKVAKTLCCGAVLGAALIAIADQOVLADEVTETNSTANVAVTTTGNPATNLPEAQGEATE
AASQSQAGSKEGALPVEVSADDLNQAVTDAAAGVNVVDQSDKGTATTAAENAKQKQAEIKSDYAKQA
EEIKKTEAYKKEVEAHQAEEDKINAENKAAEDKYQEDLKAHQAEVEKINTANATAKAEYEAQLAQYQKDL
AAVQKANEDSQLDYONKLSAYQAEARVQKANAEEAEYKAVKENTAKNAALQAENEAIKQNETAKANY
DAAMKQYEADLAATKKAKEDNDADYQAKLAAYQAEARVQKANADAKAAYEKAVEENTAKNTAIQAENEAI
KQRNAAAKATYEAALKQYEADLAAKKANEDSDADYQAKLAAYQTELARVQKANADAKAAYEKAVEDNKAK
NAALQAENEAIKQRNAAAKTDYEAKLAKYEADLAKYKKEAEYPAKLKAYEDEQAQIKAAALVELEKNKQD
GYLSKPSAQSLVYDSEPNQSLSTTNGKMLKASAVDEAFSHDTAQYSKKILQPDNLNVSYLQQADDVTSSM
ELYGNFGDKAGWTTTVGNNTVEKFPASVLLERQSVTATYTNLEKSYNGKKISKAVFKYSLSDSKFKNV
KAWLGVLDPPTLGVFASAYTGQEKDTSIFIKNEFTFYDENDQPINFDNALLSVAQLNRENNSIEMAKDYS
GTFVKISGSSVGEKDGKIYATETLNFQGGSSRWMTYKNSQPGSGWSSDAPNSWYGAGAISMSGPTNHV
TVGAI SATQVVPSPDPMVAVATGKRPNIWYSLNGKIRAVNVPKITKEKPTPPVAPTEPQAPTYEVEKPLEPA
PVAPTYENEPTPPVKTPDQPEPSKPEEPTYETEKPLEPAPVPTVENSPTPPVKTEPDQPEPSKPEEPTYET
EKPLEPAPVAPTYENEPTPPVKTPDQPEPSKPEEPTYDELPPTPPVAPTEPKQLPTPPVPTVHFHYSSLLAQ
PQINKEIKNEDGVIDIDRTLVAQSIKFKELKTEALTAGRPKTTSFVLVDPLPTGYKFDLDATKAASGTGFD
TYDEASHTVTFKATDETATYNADLTTPVETLHPTVVGRLVNDGATYINNFTLTVDNAYGIKSNVVRVTT
GKPNDFDNPMNTYIKPTKVNKNKEGLNIDGKEVLGASTNYEELTWDLQYKGDKSSKEAIQNGFYVDDYF
EEALDVRPDLVKVADEKGNQVSGVSVQYDSLEAAPKKVQDQLLKKANITVKGAFLFSADNPPEFYKQYVS
TGTSLVITDPMTVKSEFGKTGGKYENKAYQIDFGNGYATEVVVNNVVKITPKKDVTVSLDPTSENLDGQTV
QLYQTFNYRLIGGFIPQNHSELEEDYSFVDDYDQAGDQYTGNYKTFSSNLNLTMDGSGVIKAGTDLTSQTTA
ETDAANGIVTVRSKEDSLQKISLSDSPFQAEITYLQMRRIAIGTFENTYVNTVNVKAYASNTVTRTTPIPRTP
DKPTPIPTPKPKDPKPTPKPKPKVPSKVEDPSAPIPVSVGKELTLPKTGTNDSSYMPYLGLAALVGL
GLQQLKRKEDESN

SEQ ID NO: 16 Amino acid sequence comprising Streptococcal surface protein B precursor from *Streptococcus gordonii* Genbank reference G125055226 AAC44102.3

MQKREVFGRKSKVAKTLCGAVLGAALIAIADQQVLADDEVTTETNSTANVAVTTTGNPATNLPEAQGEATEA
 ASQSQAQAGSKDGCALPVEVSADDLNKAVTDAKAAGVNVVQDQTSKGTATTAAENAQKQAEIKSDYAKQAE
 EIKKTTEAYKKEVEAHQAEYDKINAENKAAEDKYQEDLKAHQAEVEKINTANATAKAAYEAKLAQYQKDLA
 AVQKANEDSQLDYQNKLSAYQAEARVQKANAEAKEAYEKAVKENTAKNAALQAEENAIKQRNETAKANYD
 AAMKQYEAADLAAIKKAKEDNDADYQAKLAAYQAEARVQKANADAKAAYEKAVEENTAKNTAIQAEENAIK
 QRNETAKATYEAAYKQYEAADLAAVQKANATNEADYQAKLAAYQTELARVQKANADAKATYKAVEDNKAKN
 AALQAEENEEIKQRNAAKTDYEAKLAKYEADLAKYKKDFAAYTAALAEAESKKFQDGYLSEPRSQSLNFKS
 EPNAIRTIDSSVHQYQGOELDALVKSWSGISPTNPDPRKKSTAYSFYFNAINSNNTYAKLVLEKDKPVDVYTG
 LKNSSFNGKKISKVVYTYTLKETGFDDGTKMTMFASSDPTVTAWYNDYFTSTNINVKVKFYDEEQQLMNLTK
 GGLVNFSSLNRRNGSGAIDKDAIESVRNFNGRYIPISSSSIKIHENNSAYADSSNAEKSRGARWDTSEWDT
 TSSPNMWYGAIVGEITQSEISFNMASSKSGNIWFAPNSNINAIGVPTKPVAPTAPTQPMYETEKPLEPAPV
 VPTYENEPTPPVKTPDQPEPSKPEEPTYETEKPLEPAPVAPTYPENEPTPPVKIPDQPEPSKPEEPTYETEK
 PLEPAPVAPTYPENEPTPPVKTPDQPEPSKPEEPTYEDPLPTPPLAPTQKQLPTPPVVPTVHFHYSSLLAQPO
 INKEIKNEDGVDLDITLVAKQSIGKPELKTEALTAGRPKTTSFVLVDPLPTGYKFDLDATKAASTGFDTTY
 DEASHTVTFKATDETALYNADLTKEPVELHPTVVGRLVNDGATYTNNTFTLVNDAYGIKSNVVRVTTGPK
 PNDPDNPNNNYIKPTKVNKNKEGLNIDGKEVLASTNYBELTWDLDQYKGDSSKEAIQNGFYVDDYPEE
 ALDVRPDLVKVADEKGNQVSGVSVQQYDSLEAAPKKVQDILLKKANTTVKGAPQLFSADNPEEFYKQYVSTG
 TSLVITDPMTVKSEFGKTKGKYENKAYQIDFGNGYATEVVVNNVPKITPKKDVTVSLDPTSENLDGQTVQL
 YQTFNYRLIGGFIPQNHSELEDYSPVDDYDQAGDQYTGNYKTFSSSLNLTMKDGSVIKAGTDLTSQTTAET
 DATNGIVTVRFKEDFLQKISLDSPPQABTYLQMRRIAIGTFENTYVNTVNVKAYASNTVRTTTPIPRTDPK
 PTPIPTPKPKDPDKPETPKPEKVPSPKVEDPSAPIFVSVGKELTTLPKTGTNDATYMPYLGLAALVGLGL
 GLAKRKED

SEQ ID NO: 17 amino acid sequence comprising PspA from *Streptococcus pneumonia* Genbank reference G1
 282335

MNKKRMILTSLSVAIILGAGFVASQPTVVRAEESPVASQSKAEKDYDAKKDAKNKAVEDAQKALDDAK
 AAQKKYDEDQKKTEEKAALKAASEEMDKAVAAVQQAYLAYQQATDKAAKDAADKMIDEAKKREHEARTKF
 NTVRAMVVPPEQLAETKKKSEEAQKAPELTKKLEEAQAKLEEAQKATEAKQKVDAAEVAPQAKIAELE
 NQVHRLQELKEIDESESEDYAKEGFRAPLQSKLDKAKKLSKLEELSDKIDELDAEIAKLEDQLKAAEEN
 NNVEDYFKEGLEKTIAAKKALEKTEADLKAVNEPEKPPAPAPETPAPEAPAEQPKPAPAPQAPAPKPEK
 PAEQPKPEKTDQQAEDYARRSEEEYNRLTQQQPPKAEKPPAPAPKTGWKQENGWYFYNTDGSMTGWLO
 NNGSWYYLNSNGAMATGWLYNGSWYYLNANGAMATGWAKVNGSWYYLNANGAMATGWLYNGSWYYLNAN
 GAMATGWAKVNGSWYYLNANGAMATGWLYNGSWYYLNANGAMATGWAKVNGSWYYLNANGAMATGWVKDG
 DTWYYLEASGAMKASQWFKVSDKWYYVNLGALAVNTTVDGYKVNANGENV

SEQ ID NO: 18 amino acid sequence comprising a portion of Se89.9 of *Streptococcus equi* Genbank reference G1
 2330384

ESDIVDATRFSTTEIPKSGQVIDRSASIQALTNDIASIKGKIASLESRLADPSSEAEVTAQAQAKISQLQH
 QLEAAQAKSHKLDQQVEQLANTKDSLRTQLLAKEEQAKLANKLDKALALLASSKATLHKLEAAMEEAKA
 RVAGLASQKAQLEDLLAFEKNPNRIELAQEKVAAAKKALADTEDKLLAAQASLSDLOAQRARLQLSIATI

SEQ ID NO: 19 polynucleotide sequence comprising GST-40-HIS

CTGGTTCCGCGTGGATCCCATATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGG
 AGAAACGAAGGCGAGTAATACTCAGCAGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAGGCCAA
 CTATTGATGTCAGTTGAAAAAATCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACA
 AAACTACTGCTGAAATCAACCCTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACA
 AGAAATTFACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAG
 AGTTAACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAGAGAGCTGCATTSTCA
 GAACAAAAAGCTAGCATTTTCAGCAGAACTACTCGAGCTCAAGATTTAGTGGAAACAAGTCAAACAGCTCTGA
 ACAAATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTA
 ATGATAATACAAAAGCATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAAATCAAAAAGCTAAA
 GTTAAAAAGCAATTGACTGAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCGAGAAGCTTAG
 TCGTCTTAAATCCTCAGCTCCGCTTACTCAAGATAGCATTGTGGGTAAATAACCATGAAAGCACCGCAAG
 GCTATCCTCTTGAAGAATTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTAT
 TACAAAGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATAT
 TCCAGCAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGT

TTGCAGCTCACATGATTAATAGTGTAAAGAAGACAATTAGGTTCTACCACCAGTTACTGTTACAGCAGGATCA
 CAAGAATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGTCTA
 CGGACAGCCAGGGGTATCAGGGCATTATGGTGTGGGCTCATGATAAACTATTATTGAAGACTCTGCCG
 GAGCGTCAGGGCTCATTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACT
 GTGAATGGTAAATTAACGTTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAA
 TACATACGGCCATGCTATTAACCTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGAT
 TTTCAACCAGCAATGTAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCAT
 CAACGCTTTAATAAGACCCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCCAAGAGTAGGCACCTGT
 ATCTGATACTATTGTCAGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTGCGCTATTTCATCAAG
 AAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAG
 TCAGACAGCTTAAATCTCCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGC
 AGCTAAAGCAAAACAAGCACAACCTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAG
 CCGCACTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGACCCAGAGTGACAGCACTGGTGGCTAAAAAA
 GCTCATTTGCAATATCTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTTGA
 TAATACTAAGCAAGATTTGGCTAAAACCTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTAC
 AAGCTAAACAAAGCAGTCTAGAAGCTACTATTGCTACCCACAGAACACCAGTTGACTTTGCTTAAACCTTA
 GCTAACGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACC
 TCTTACGGGCGTAAACCCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAGAAATGGTTAAAG
 AAACGAAACAACCTATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTT
 GTTGGCCAAACCTCTGAAATGCTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCC
 CTCATCTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTA
 CTCAAAGAGCTCTTAAAGCAGGAGTCGTATGTTGGCAGCTGTGCGCCTCACAGGATTTAGGTTCCGTAAG
 GAATCTAAGGCGGCGCACTCGAGCACCACCACCACCACCACCAC

SEQ ID NO: 20 amino acid sequence comprising GST-40-HIS

L V P R G S H Met S V G V S H Q V K A D D R A S G E T K A S N T H D D S
 L P K P E T I Q E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T
 K T T A E I N H L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E
 T L L A Q G A E H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E
 Q K A S I S A E T T R A Q D L V E Q V K T S E Q N I A K L N A Met I S N
 P D A I T K A A Q T A N D N T K A L S S E L E K A K A D L E N Q K A K V
 K K Q I T E E L A A Q K A A L A E K E A E L S R L K S S A P S T Q D S I
 V G N N T Met K A P Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y
 K E H A D Q I I A K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L
 T P E V Q N R L A Q F A A H Met I N S V R R Q L G L P P V T V T A G S Q
 E F A R L L S T S Y K K T H G N T R P S F V Y G Q P G V S G H Y G V G P
 H D K T I I E D S A G A S G L I R N D D N Met Y E N I G A F N D V H T V
 N G I K R G I Y D S I K Y Met L F T D H L H G N T Y G H A I N F L R V D
 K H N P N A P V Y L G F S T S N V G S L N E H F V Met F P E S N I A N H
 Q R F N K T P I K A V G S T K D Y A Q R V G T V S D T I A A I K G K V S
 S L E N R L S A I H Q E A D I Met A A Q A K V S Q L Q G K L A S T L K Q
 S D S L N L Q V R Q L N D T K G S L R T E L L A A K A K Q A Q L E A T R
 D Q S L A K L A S L K A A L H Q T E A L A E Q A A A R V T A L V A K K A
 H L Q Y L R D F K L N P N R L Q V I R E R I D N T K Q D L A K T T S S L
 L N A Q E A L A A L Q A K Q S S L E A T I A T T E H Q L T L L K T L A N
 E K E Y R H L D E D I A T V P D L Q V A P P L T G V K P L S Y S K I D T
 T P L V Q E Met V K E T K Q L L E A S A R L A A E N T S L V A E A L V G
 Q T S E Met V A S N A I V S K I T S S I T Q P S S K T S Y G S G S S T T
 S N L I S D V D E S T Q R A L K A G V V Met L A A V G L T G F R F R K E
 S K A A A L E H H H H H H

SEQ ID NO: 21 polynucleotide sequence comprising 40a-HIS

ATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATAC
 TCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAAAAAA
 CTCTCAGTCAACAAAAAGCAGAAGTACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAATCAAC
 CACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCT

TGCAAGTAGTGGAGGACGCTATTAGCCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAG
 AGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTGTCAGAACAAAAAGCTAGCATTTC
 GCAGAAACTACTCGAGCTCAAGATTAGTGGAAACAAGTCAAAACGCTCTGAACAAAATATTGCTAAGCTCAA
 TGCTATGATTAGCAATCGCTGATGCTATCACTAAAGCAGCTCAAAACGGCTAATGATAATACAAAAGCATTA
 GCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAGCTAAAGTTAAAAAGCAATTGACTGAA
 GAGTTGSCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGCTCTAAATCCTCAGCTCC
 GTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCAGCAAGGCTATCCTCTTGAAGAACCTA
 AAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAA
 ATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTT
 TGTGATCCCGATAATTTGACACCAGAAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAAAT
 GTGTAAGAAGACAAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGATTACTT
 AGTACCAGCTATAAGAAAACCTCATGGTAATACAAGACCATCATTGTCTACGGACAGCCAGGGGTATCAGG
 GCATTATGGTGTGGGCTCATGATAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCATTCGAA
 ATGATGATAACATGTACGAGAATATCGGTGCTTTTAAAGCATGTGCATCTGTGAATGGTATTAAACGTGGT
 ATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTACACGGAAATACATACGGCCATGCTATTAA
 CTTTTTACGTGTAGATAAACATAACCCTAATGCGCCTGTTTACCTTGGATTTCACACAGCAATGTAGGAT
 CTTTGAATGAACACTTTGTAATGTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCT
 ATAAAGCCGTTTGAAGTACAAAAGATTATGCCAAAGAGCTAGGCACTGTATCTGATACTATTGACCGCAT
 CAAAGGAAAAGTAAAGCTCATTAGAAAATCGTTTGTGCGCTATTTCATCAAGAAGCTGATATTATGGCAGCCC
 AAGCTAAAGTAAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAA
 GTGAGACAATTAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAGCACA
 ACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAA
 CCTTAGCAGAGCAAGCCGAGCCAGAGTGACAGCACTGGTGGCTAAAAAGCTCATTTGCAATATCTAAGG
 GACTTTAAATTGAATCCTTAACCGCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATTGGC
 TAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGCTAG
 AAGCTACTATTGCTACACAGAACACCAGTTGACTTTGCTTAAACCTTAGCTAACGAAAAGGAATATCGC
 CACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAACCGCT
 ATCATATAGTAAGATAGATACTACTCCGCTTGTTCAGAAATGGTTAAAGAAACGAAACAATATTAGAAG
 CTTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTGGCCAAACCTCTGAAATG
 GTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGG
 CTCAGGATCTTCTACAACGAGCAATCTCATTCTGATGTTGATGAAAGTACTCAAGCGTGGCGCCGCACTCG
 AGCACCACCACCACCACCACCAC

SEQ ID NO: 22 amino acid sequence comprising 40a-HIS

M S V G V S H Q V K A D D R A S G E T K A S N T H D D S L P K P E T I Q
 E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H
 L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E
 H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E
 T T R A Q D L V E Q V K T S E Q N I A K L N A Met I S N P D A I T K A A
 Q T A N D N T K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E L
 A A Q K A A L A E K E A E L S R L K S S A P S T Q D S I V G N N T M K A
 P Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I A
 K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L A
 Q F A A H M I N S V R R Q L G L P P V T V T A G S Q E F A R L L S T S Y
 K K T H G N T R P S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A
 G A S G L I R N D D N M Y E N I G A F N D V H T V N G I K R G I Y D S I
 K Y M L F T D H L H G N T Y G H A I N F L R V D K H N P N A P V Y L G F
 S T S N V G S L N E H F V M F P E S N I A N H Q R F N K T P I K A V G S
 T K D Y A Q R V G T V S D T I A A I K G K V S S L E N R L S A I H Q E A
 D I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T
 K G S L R T E L L A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L
 H Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L R D F K L N P N R
 L Q V I R E R I D N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q
 S S L E A T I A T T E H Q L T L L K T L A N E K E Y R H L D E D I A T V
 P D L Q V A P P L T G V K P L S Y S K I D T T P L V Q E M V K E T K Q L
 L E A S A R L A A E N T S L V A E A L V G Q T S E M V A S N A I V S K I

T S S I T Q P S S K T S Y G S G S S T T S N L I S D V D E S T Q R A A A
L E H H H H H H H

SEQ ID NO: 23 polynucleotide sequence comprising 40a-RR-HIS

ATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATAC
TCACGACGATAGTTTACCAAAACAGAAACAAATCAAGAGGCAAGGCAACTATTGATGCAAGTTGAAAAA
CTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACCTGCTGAAATCAAC
CACTTAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCT
TGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAG
AGCTTCATAATGCTCAAGCAGATCAACATTCAAAGAGAGCTGCATTGTCAGAACAAAAAGCTAGCATTTC
GCAGAAACTACTCGAGCTCAAGATTTAGTGGAAACAAGTCAAACCGTCTGAACAAAAATATTGCTAAGCTCAA
TGCTATGATTAGCAATCCTGATGCTATCACTAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTAA
GCTCAGAATTTGGAGAAGGCTAAAGCTGACTTAGAAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAA
GAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCTGCTTAAATCCTCAGCTCC
GTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAACTTA
AAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATTAATTATTACAAAGAGCATGCAGATCAA
ATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTT
TGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAATA
GTGTAGCTGCAATTAGGCTTACCACCAGTTACTGTTACAGCAGGATCACAAGAAATTTGCAAGATTACTT
AGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCATTTGCTACGGACAGCCAGGGGTATCAGG
GCATTATGGTGTGGGCCCTCATGATAAACTATTATTGAAGACTCTGCCGGAGCGCTCAGGGCTCATTCGAA
ATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACGTGGT
ATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGCTATTAA
CTTTTACGTGTAGATAAAACATAACCCCTAATGCGCCTGTTTACCTTGGATTTTCAACCAGCAATGTAGGAT
CTTTGAATGAACACTTTGTAAATGTTCCAGAGCTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCT
ATAAAAGCCGTTGGAAAGTACAAAAGATTATGCCCAAAGAGTAGGCAGCTGTATCTGATACTATTGCAGCGAT
CAAAGGAAAAGTAAGCTCATTTAGAAAATCGTTTGTGGCTATTTCATCAAGAAGCTGATATTATGGCAGCCC
AAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAA
GTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAAATTACTAGCAGCTAAAGCAAAACAAGCACA
ACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAAG
CCTTAGCAGAGCAAGCCGAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCTAAGG
GACTTTAAATGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATTGGC
TAAACTACCTCATCTTTGTTAARTGCACAGAAGCTTTAGCAGCCTTACAAGCTAAACAAGCAGTCTAG
AAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAACCTTAGCTAACGAAAAGGAATATCGC
CACTTAGACGAAGATATAGCTACTGTGCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAACCGCT
ATCATATAGTAAGATAGATACTACTCCGCTTGTCAAGAAATGGTTAAAGAAACGAAACAACATTTAGAA
CTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAAGCGCTTGTGGCCAAACCTCTGAAATG
GTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGG
CTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAGCTGCGGCCGCACTCG
AGCACCACCACCACCACCACC

SEQ ID NO: 24 amino acid sequence comprising 40a-RR-HIS

M S V G V S H Q V K A D D R A S G E T K A S N T H D D S L P K P E T I Q
E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H
L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E
H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E
T T R A Q D L V E Q V K T S E Q N I A K L N A Met I S N P D A I T K A A
Q T A N D N T K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E L
A A Q K A A L A E K E A E L S R L K S S A P S T Q D S I V G N N T N K A
P Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A
K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L A
Q F A A H M I N S V R R Q L G L P V T V T A G S Q E F A R L L S T S Y
K K T H G N T R P S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A
G A S G L I R N D D N M Y E N I G A F N D V H T V N G I K R G I Y D S I
K Y M L F T D H L H G N T Y G H A I N F L R V D K H N P N A P V Y L G F
S T S N V G S L N E H F V M F P E S N I A N H Q R F N K T P I K A V G S

T K D Y A Q R V G T V S D T I A A I K G K V S S L E N R L S A I H Q E A
D I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T
K G S L R T E L L A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L
H Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L R D F K L N P N R
L Q V I R E R I D N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q
S S L E A T I A T T E H Q L T L L K T L A N E K E Y R H L D E D I A T V
P D L Q V A P P L T G V K P L S Y S K I D T T P L V Q E M V K E T K Q L
L E A S A R L A A E N T S L V A E A L V G Q T S E M V A S N A I V S K I
T S S I T Q P S S K T S Y G S G S S T T S N L I S D V D E S T Q R A A A
L E H H H H H H H

SEQ ID NO: 25 polynucleotide sequence comprising 40a-RR (nat)

ATGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATAC
TCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTTGAAAAAA
CTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAACTACTGCTGAAATCAAC
CACTTAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCT
TGCAAGTAGTGCAGGAGACGCTATTAGCCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAG
AGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTGAGAACAAAAAGCTAGCATTTC
GCAGAACTACTCGAGCTCAAGATTTAGTGGAAACAAGTCAAAACGCTCTGAACAAATATTGGCTAAGCTCAA
TGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTA
GCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAAAGCTAAAGTTAAAAAGCAATTGACTGAA
GAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGCTCTTAAATCCTCAGCTCC
GTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAAGCTTA
AAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAA
ATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTT
TGTTGATCCCGATAATTTGACACCAGAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAAATA
GTGTACGtGtCAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGATTACTT
AGTACCAGCTATAAGAAAACCTCATGGTAATACAAGACCATCATTTGTCTACGGACAGCCAGGGGTATCAGG
GCATTATGGTGTGGGCTCATGATAAACTATTATTGAAGACTCTGCCGAGCGTCAGGGCTCATTCGAA
ATGATGATAACATGTACGAGAATATCGGTGCTTTTAAACGATGTGCATACTGTGAATGGTATTAAACGTGGT
ATTTATGACAGTATCAAGTATATGCTCTTTACAGATCAATTACACGGAAATACATACGGCCATGCTATTAA
CTTTTACGTTGTAGATAAACATAACCTAATGCGCCTGTTTACCTTGGATTTCACACAGCAATGTAGGAT
CTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACGCTTTAATAAGACCCCT
ATAAAAGCCGTTTGAAGTACAAAAGATTATGCCCAAAGAGTAGGCACTGTATCTGATACTATTGCAGCGAT
CAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTGCGCTATTATCAAGAAGCTGATATTATGGCAGCCC
AAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCAA
GTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAGCACA
ACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAAG
CCTTAGCAGAGCAAGCCGAGCCAGAGTGACAGCACTGGTGGCTAAAAAGCTCATTGTGCAATATCTAAGG
GACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATTGATAATACTAAGCAAGATTGGC
TAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCAGTCTAG
AAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAACCTTAGCTAACGAAAAGGAATATCGC
CACTTAGACGAAGATATAGCTACTGTGCTGATTGTGCAAGTAGCTCCACCTCTTACGGGCGTAAACCGCT
ATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACCTATTAGAAG
CTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTGGCCAAACCTCTGAAATG
GTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGG
CTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAGCt

SEQ ID NO: 26 amino acid sequence comprising 40a-RR (nat)

M S V G V S H Q V K A D D R A S G E T K A S N T H D D S L P K P E T I Q
E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H
L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E
H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E
T T R A Q D L V E Q V K T S E Q N I A K L N A M I S N P D A I T K A A Q
T A N D N T K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E L A
A Q K A A L A E K E A E L S R L K S S A P S T Q D S I V G N N T M K A P

Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A K
A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L A Q
F A A H M I N S V R R Q L G L P P V T V T A G S Q E F A R L L S T S Y K
K T H G N T R P S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A G
A S G L I R N D D N M Y E N I G A F N D V H T V N G I K R G I Y D S I K
Y M L F T D H L H G N T Y G H A I N F L R V D K H N P N A P V Y L G F S
T S N V G S L N E H F V M F P E S N I A N H Q R F N K T P I K A V G S T
K D Y A Q R V G T V S D T I A A I K G K V S S L E N R L S A I H Q E A D
I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T K
G S L R T E L L A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L H
Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L R D F K L N P N R L
Q V I R E R I D N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q S
S L E A T I A T T E H Q L T L L K T L A N E K E Y R H L D E D I A T V P
D L Q V A P P L T G V K P L S Y S K I D T T P L V Q E M V K E T K Q L L
E A S A R L A A E N T S L V A E A L V G Q T S E M V A S N A I V S K I T
S S I T Q P S S K T S Y G S G S S T T S N L I S D V D E S T Q R

SEQ ID NO: 27 polynucleotide sequence comprising His-40a NH

ATGGGATCGCATCACCATCACCATCACGCTAGTAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAG
AGCCTCAGGAGAAACGAAGGCGAGTAATACTCACGACGATAGTTTACCAAAACAGAAACATTCAGAGG
CAAAGGCAACTATTGATGCACTGAAAAAAGCTCTCAGTCAACAAAAGCAGAACTGACAGAGCTTGCTACC
GCTCTGACAAAAGCTACTGCTGAAATCAACCACTTAAAGAGCAGCAAGATAATGAACAAAAGCTTTAAC
CTCTGCACAAGAAATTTACACTAATACTCTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAAC
A¹CAAAGAGAGTTAACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACT
GCATTGTGACAAAAGCTAGCATTTCAGCAGAACTACTCGAGCTCAAGATTTAGTGGAAACAAGTCAA
AACGCTCTGAACAAAATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTC
AAACGGCTAATGATAATACAAAAGCATTAAAGCTCAGAATTGGAGAAGGC²TAAAGCTGACTTAGAAAATCAA
AAAGCTAAAGTTAAAAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAGAGGC
AGAACTTAGTCGCTCTAAATCCTCAGCTCCGCTACTCAAGATAGCATTTGTGGGTAATAATACCATGAAAG
CACCGCAAGGCTATCCTCTTGAAGAAGCTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTAC
AATAATTATTACAAAGAGCATGCAGATCAAAATATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATA
CCAAGATATTTCCAGCAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGC
TAGCGCAGTTTGCAGCTCACATGATTAATAGTGTGAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACA
GCAGGATCACAAAGAAATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACCTATGGTAATACAAAGACCATC
ATTTGTCTACGGCAGCCAGGGGTATCAGGGCATTATGGTGTGGGCCCTCATGATAAAACTATTATTGAAG
ACTCTGCCGGAGCGTCAGGGCTCATTCGAAATGATGATAACATGTACGAGAAATATCGGTGCTTTTAACGAT
GTGCATACTGTGAATGGTATTTAAACGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTT
ACACGGAAATACATACGGCCATGCTATTAACTTTTACGTTGATATAAACATAACCC³TAAATGCGCCTGTTT
ACCTTGGATTTTCAACCAGCAATGTAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGCTTAACATT
GCTAACCATCAACGCTTTAATAAGACCCCTATAAAAGCCGTTGGAAGTACAAAAGA⁴CTATGCCCAAAGAGT
AGGCACTGTATCTGATACTATTGACAGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTGGCTA
TTCATCAAGAAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACA
CTTAAGCAGTCAGACAGCTTAAATCTCCAAAGTGAGACAATTAAATGATACATAAGGTTCTTTGAGAACAGA
ATTACTAGCAGCTAAAGCAAAACAAGCACAACCTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCAT
CGTTGAAAGCCGCACTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGACCCAGAGTGACAGCACTGGTG
GCTAAAAAAGCTCATTGTGCAATATCTAAGGGACTTTAAATGAATCCTAACC⁵CGCTTCAAGTGATACGTGA
GCGCATTGATAATACTAAGCAAGATTTGGCTAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAG
CAGCCTTACAAGCTAAACAAAGCAGTCTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTT
AAAACCTTAGCTAACGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGT
AGCTTCAACCTTACGGGCGTAAACCCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAGAAA
TGGTTAAAGAAACGAAACAACCTATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAGCTCTTGTAGCA
GAAGCGCTTGTGGCCAAACCTCTGAAATGGTAGCAAGTAATGCCATTTGTGTCTAAAATCACATCTTCGAT
TACTCAGCCCTCATCTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTG
ATGAAAGTACTCAAC⁶gt

SEQ ID NO: 28 amino acid sequence comprising His-40a NH

N G S H H H H H H A S S V G V S H Q V K A D D R A S G E T K A S N T H D
 D S L P K P E T I Q E A K A T I D A V E K T L S Q Q K A E L T E L A T A
 L T K T T A E I N H L K E Q Q D N E Q K A L T S A Q E I Y H S K E T A S
 E E T L L A Q G A E H Q R E L T A T E T E L H N A Q A D Q H S K E T A L
 S E Q K A S I S A E T T R A Q D L V E Q V K T S E Q N I A K L N A M I S
 N P D A I T K A A Q T A N D N T K A L S S E L E K A K A D L E N Q K A K
 V K K Q L T E E L A A Q K A A L A E K E A E L S R L K S S A P S T Q D S
 I V G N N T M K A P Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y
 K E H A D Q I I A K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L
 T P E V Q N E L A Q F A A H M I N S V R R Q L G L P E V T V T A G S Q E
 F A R L L S T S Y K K T H G N T R P S F V Y G Q P G V S G H Y G V G P H
 D K T I I E D S A G A S G L I R N D D N M Y E N I G A F N D V H T V N G
 I K R G I Y D S I K Y M L F T D H L H G N T Y G H A I N F L R V D K H N
 P N A P V Y L G F S T S N V G S L N E H F V M F P E S N I A N H Q R F N
 K T P I K A V G S T K D Y A Q R V G T V S D T I A A I K G K V S S L E N
 R L S A I H Q E A D I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N
 L Q V R Q L N D T K G S L R T E L L A A K A K Q A Q L E A T R D Q S L A
 K L A S L K A A L H Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L
 R D F K L N P N R L Q V I R E R I D N T K Q D L A K T T S S L N A Q E
 A L A A L Q A K Q S S L E A T I A T T E H Q L T L L K T L A N E K E Y R
 H L D E D I A T V P D L Q V A P F L T G V K P L S Y S K I D T T P L V Q
 E M V K E T K Q L L E A S A R L A A E N T S L V A E A L V G Q T S E M V
 A S N A I V S K I T S S I T Q P S S K T S Y G S G S S T T S N L I S D V
 D E S T Q R

SEQ ID NO. 29 polynucleotide sequence comprising HIS-40a CH

ATGGCTAGTAGTGTAGGCGTATCTCACCAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGCGAG
 TAATACTCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCCAAAGGCAACTATTGATGCAGTTG
 AAAAAACTCTCAGTCAACAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACGAAAACCTACTGCTGAA
 ATCAACCACTTTAAAAGAGCAGCAAGATAATGAACAAAAGCTTTAACCCTCTGCACAAGAAATTTACACTAA
 TACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTG
 AAACAGAGCTTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACTGCATTGTCAGAACAAAAAGCTAGC
 ATTTTCAGCAGAACTACTCGAGCTCAAGATTTAGTGGAAACAAGTCAAACAGCTCTGAACAAAATATTGCTAA
 GCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAG
 CATTAAAGCTCAGAAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAGAGCTAAAGCTTAAAAAGCAATTG
 ACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTC
 AGCTCCGCTCTACTCAAGATAGCATTGTGGTAATAATACCATGAAAGCACCAGGCTATCCTCTTGAAG
 AACTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCA
 GATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAA
 TCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCAGATGA
 TTAATAGTGTAAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAGAATTTGCAAGA
 TTAATTAGTACCAGCTATAAGAAAACCTCATGGTAATACAAGACCATCATCTGTCTACGGACAGCCAGGGGT
 ATCAGGGCATTATGGTGTGGGCTCATGATAAACTATTATTGAAGACTCTGCCGGAGCGTCAGGGCTCA
 TTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAAACGATGTGCATACTGTGAATGGTAATAAA
 CGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGGAAATACATACGGCCATGC
 TATTAACCTTTTACGTGTAGATAAACATAACCTAATGCGCTGTGTTACCTTGGATTTTCAACCCAGCAATG
 TAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTTGCTAACCATCAACGCTTTAATAAG
 ACCCTATATAAAGCCGTTGGAGTACAAAAGATTATGCCCAAAGAGTAGGCAGTGTATCTGATACCTATTGC
 AGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTGGCTATTTCATCAAGAGCTGATATTATGG
 CAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAAT
 CTCCAAGTGAGACAATTAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACA
 AGCACAACTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCATCGTTGAAAGCCGCACTGCACCAGA
 CAGAAGCTTAGCAGAGCAAGCCGAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATAT
 CTAAGGGACTTTAAATTTGAATCCTAACCGCTTCAAGTATACGTGAGCGCATTGATAATCTAAGCAAGA
 TTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCTAAACAAAGCA
 GTCTAGAAGCTACTATTGCTACCACAGAACACAGTTGACTTTGCTTAAAACTTAGCTAACGAAAAGGAA

TATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGTAGCTCCACCTCTTACGGGCGTAAA
 ACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAATGGTTAAAGAAACGAAACAACATAT
 TAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCAGAAGCGCTTGTGGCCAAACCTCT
 GAAATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGATTACTCAGCCCTCATCTAAGACATC
 TTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTGATGAAAGTACTCAAGGtGCGGCCG
 CACTCGAGCACCACCACCACCACCAC

SEQ ID NO: 30 amino acid sequence comprising HIS-40a CH

M A S S V G V S H Q V K A D D R A S G E T K A S N T H D D S L P K P E T
 I Q E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I
 N H L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G
 A E H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S
 A E T T R A Q D L V E Q V K T S E Q N I A K L N A M I S N P D A I T K A
 A Q T A N D N T K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E
 L A A Q K A A L A E K E A E L S R L K S S A P S T Q D S I V G N N T M K
 A P Q G S P L E E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I
 A K A Y P N Q L N Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L
 A Q F A A H M I N S V R R Q L G L P V T V T A G S Q E F A R L L S T S
 Y K K T H G N T R P S E V Y G Q P G V S G H Y G V G P H D K T I I E D S
 A G A S G L I R N D D N M Y E N I G A F N D V H T V N G I K R G I Y D S
 I K Y M L F T D H L H G N T Y G H A I N F L R V D K H N P N A P V Y L G
 F S T S N V G S L N E H F V M F P E S N I A N H Q R F N K T P I K A V G
 S T K D Y A Q R V G T V S D T I A A I K G K V S S L E N R L S A I H Q E
 A D I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D
 T K G S L R T E L L A A K A K Q A Q L E A T R D Q S L A K L A S L K A A
 L H Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L R D F K L N P N
 R L Q V I R E R I D N T K Q D L A K T T S S L L N A Q E A L A A L Q A K
 Q S S L E A T I A T T E H Q L T L L K T L A N E K E Y R H L D E D I A T
 V P D L Q V A P P L T G V K P L S Y S K I D T T P L V Q E M V K E T K Q
 L L E A S A R L A A E N T S L V A E A L V G Q T S E M V A S N A I V S K
 I T S S I T Q P S S K T S Y G S G S S T T S N L I S D V D E S T Q R A A
 A L E H H H H H H

SEQ ID NO: 31 polynucleotide sequence comprising HIS-40a-RR NH

ATGGGATCGCATCACCATCACCATCAGCTAGTAGTGTAGGCGTATCTCACCAGTCAAAGCAGATGATAG
 AGCCTCAGGAGAAACGAAGCGAGTAATACTCAGCAGATAGTTACCAAAACCAGAAACAATTCAAGAGG
 CAAAGGCAACTATTGATGTCAGTTGAAAAAATCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACC
 GCTCTGACAAAACTACTGCTGAAATCAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAAC
 CTCTGCACAAGAAATTTACACTAATACTCTTGAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAAC
 ATCAAAGAGAGTTAACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGACT
 GCATTGTGCAACAACAAAGCTAGCATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGAACAAGTCAA
 AACGCTGAACAAATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTC
 AAACGGCTAATGATAATACAAAAGCATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAA
 AAAGCTAAAGTTAAAAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGC
 AGAAGCTAGTCTGCTTTAAATCCTCAGCTCCGCTACTCAAGATAGCATTGTGGGTAAATAATACCATGAAAG
 CACCGCAAGGCTATCCTCTTGAAGAATTAATAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTAC
 AATAATTATTACAAAGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATA
 CCAAGATATTCCAGCAGATCGTAATCGCTTTGTTGATCCCGATAATTGACACCAGAAAGTGCAAAATGAGC
 TAGCGCAGTTTGCAGCTCACATGATTAATAGTGTACGtGtGtCAATTAGGTCTACCACCAGTTACTGTTACA
 GCAGGATCACAGAATTTGCAAGATTACTTAGTACCAGCTATAGAAAATCATGGTAAATACAGACCATC
 ATTTGCTACGGACAGCCAGGGGTATCAGGGCATTATGGTGTGGGCCCTCATGATAAACTATTATTGAAG
 ACTCTGCCGGAGCGTCAGGGCTCATTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTACGAT
 GTGCATCTGTGAATGGTATTAAACGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTT
 ACACGGAAATACATACGGCCATGCTATTAACCTTTTACGTGTAGATAAACATAACCCATATGCGCCTGTTT
 ACCTTGATTTTCAACCAGCAATGTAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATT
 GCTAACCATCAACGCTTTAATAAGACCCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGT

AGGCACTGTATCTGATCTATTGCAGCGATCAAAGGAAAAGTAAGCTCATTAGAAAAATCGTTTGTCTGGCTA
TTCATCAAGAAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACA
CTTAAGCAGTCAGACAGCTTAAATCTCCAAGTGAGACAATTAATGATACTAAAGGTTCTTTGAGAACAGA
ATTACTAGCAGCTAAAGCAAAACAAGCACAACCTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCAT
CGTTGAAAGCCGCACTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGAGCCAGAGTGACAGCACTGGTG
GCTAAAAAAGCTCATTGCAATATCTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGA
GCGCATTGATAATACTAAGCAAGATTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAG
CAGCCTTACAAGCTAAACAAAGCAGTCTAGAAGCTACTATTGCTACCCACAGAACACCAGTTGACTTTGCTT
AAAACCTTAGCTAACGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTTGCAAGT
AGCTCCACCTCTTACGGGCGTAAAACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAAGAAA
TGTTTAAAGAAACGAAACAACCTATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTGTAGCA
GAAGCGCTTGTGTGGCCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGTCTAAAATCACATCTTCGAT
TACTCAGCCCTCATCTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCTGATGTTT
ATGAAAGTACTCAAGT

SEQ ID NO: 32 amino acid sequence comprising HIS-40a-RR NH

M G S H H H H H A S S V G V S H Q V K A D D R A S G E T K A S N T H D
D S L P K P E T I Q E A K A T I D A V E K T L S Q Q K A E L T E L A T A
L T K T T A E I N H L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S
E E T L L A Q G A E H Q R E L T A T E T E L H N A Q A D Q H S K E T A L
S E Q K A S I S A E T T R A Q D L V E Q V K T S E Q N I A K L N A M I S
N P D A I T K A A Q T A N D N T K A L S S E L E K A K A D L E N Q K A K
V K K Q L T E E L A A Q K A A L A E K E A E L S R L K S S A P S T Q D S
I V G N N T M K A P Q G V P L E E L K K L E A S G Y I G S A S Y N N Y Y
K E H A D Q I I A K A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L
T P E V Q N E L A Q P A A H M I N S V R R Q L G L P P V T V T A G S Q E
F A R L L S T S Y K K T H G N T R P S F V Y G Q P G V S G H Y G V G P H
D K T I I E D S A G A S G L I R N D D N M Y E N I G A F N D V H T V N G
I K R G I Y D S I K Y M L F T D H L H G N T Y G H A I N F L R V D K H N
P N A P V Y L G F S T S N V G S L N E H F V M F P E S N I A N H Q R F N
K T P I K A V G S T K D Y A Q R V G T V S D T I A A I K G K V S S L E N
R L S A I H Q E A D I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N
L Q V R Q L N D T K G S L R T E L L A A K A K Q A Q L E A T R D Q S L A
K L A S L K A A L H Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L
R D F K L N P N R L Q V I R E R I D N T K Q D L A K T T S S L L N A Q E
A L A A L Q A K Q S S L E A T I A T T E H Q L T L L K T L A N E K E Y R
H L D E D I A T V P D L Q V A P P L T G V K P L S Y S K I D T T P L V Q
E M V K E T K Q L L E A S A R L A A E N T S L V A E A L V G Q T S E M V
A S N A I V S K I T S S I T Q P S S K T S Y G S G S S T T S N L I S D V
D E S T Q R

SEQ ID NO: 33 polynucleotide sequence comprising 40N-HIS

ATGCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAGTAATACTCAGCAGCAGATAGTTTACC
 AAAACCAGAAACAATTCAAGAGGCAAAAGGCAACTATTGATGCAGTTGAAAAAACTCTCAGTCAACAAAAAG
 CAGAACTGACAGAGCTTGCTACCGCTCTGACAAAACTACTGCTGAAATCAACCCTTAAAAGAGCAGCAA
 GATAATGAACAAAAAGCTTTAACCCTCTGCACAAGAAATTTACACTAATACTCTTGCAAGTAGTGAGGAGAC
 GCTATTAGCCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAGAGCTTCATAATGCTCAAG
 CAGATCAACATTCAAAGAGACTGCATTGTCAGAACAAAAAGCTAGCATTTCAGCAGAACTACTCGAGCT
 CAAGATTTAGTGGAACAAGTCAAAACGCTCTGAACAAAATATTGCTAAGCTCAATGCTATGATTAGCAATCC
 TGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTAAGCTCAGAATTGGAGAAGG
 CTAAAGCTGACTTAGAAAAACAAAAGCTAAAGTTAAAAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAA
 GCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCTTAAATCCTCAGCTCCGTCTACTCAAGATAGCAT
 TGTTGGSTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTGAAGAACTTAAAAAATTAGAAGCTAGTG
 GTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAAATTATTGCCAAAGCTAGT
 CCAGGTAATCAATTAAATCAATACCAAGCGCGCCGCACTCGAGCACCACCACCACCACCACCAC

SEQ ID NO: 34 amino acid sequence comprising 40N-HIS

M Q V K A D D R A S G E T K A S N T H D D S L P K P E T I Q E A K A T I
 D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H L K E Q Q D
 N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E H Q R E L T
 A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E T T R A Q D
 L V E Q V K T S E Q N I A K L N A M I S N P D A I T K A A Q T A N D N T
 K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E L A A Q K A A L
 A E K E A E L S R L K S S A P S T Q D S I V G N N T M K A P Q G Y P L E
 E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A K A S P G N Q
 L N Q Y Q A A A L E H H H H H H H

SEQ ID NO: 35 amino acid sequence comprising GAS 117

MTLKKHYLLSLLALVTVGAAFNTSQSVSAQVYSNEGYPHQLTDEKSHLQYSKDNAQLQLRNILDGYQNDL
 GRHYSSYYYNLRTVMGLSSEQDIEKHYEELKNKLHDMYNHY

SEQ ID NO: 36 polynucleotide sequence encoding GAS 117

ATGACACTAAAAAACACTATTATCTTCTCAGCCTGCTAGCTCTTGTAACGGTTGGTGCTGCCTTTAACAC
 AAGCCAGAGTGTCAAGTGCACAAGTTTATAGCAATGAAGGGTATCACCAGCATTTCAGTATGAAAAATCAC
 ACCTGCAATATAGTAAAGACAACGCACAACCTCAATTGAGAAATATCCTTGACGGCTACCAAAATGACCTA
 GGGAGACACTACTCTAGCTATTATTACTACAACCTAAGAACCCTTATGGGACTATCAAGTGAGCAAGACAT
 TGA AAAACACTATGAAGAGCTTAAGAACAAGTTACATGATATGTACAATCATTATTAA

SEQ ID NO: 37 amino acid sequence comprising GAS 117 leader sequence TLKKHYLLSLLALVTVGA

SEQ ID NO: 38 amino acid sequence comprising fragment of GAS 117 where leader sequence is removed

AFNTSQSVSAQVYSNEGYPHQLTDEKSHLQYSKDNAQLQLRNILDGYQNDLGRHYSSYYYNLRTVMGLSS
 EQDIEKHYEELKNKLHDMYNHY

SEQ ID NO: 39 amino acid sequence comprising GAS 130

MSHMKRPEVLSPAGTLEKLKVAIDYGADAVFVGQAYGLRSRAGNFSMEELQEGIDYAHARGAKVYVAAN
 MVTHEGNEIGAGEWFRQLRDMGLDAVIVSDPALIVICSTEAPGLEIHLSTQASSTNYETPEFWKAMGLTRV
 VLAREVNMAELAEIRKRTDVEIEAPVHGAMCISYSGRCVLSNHMSHRDANRGGCSQSCRWKYDLYDMPFGG
 ERRSLKGEIPEYSSSVDMCMIDHIPDLIENGVDLSLKIEGRMKS IHVYSTVTNICYKAAVGYMESPEAFY
 AIKEELIDELWKVAQRELATGFYVGIPTENEQLFGARRKI PQYKFVGEVVAFDSASMTATIRQNVIMEGD
 RIECYGPGFRHFETVVKDLHDADGQKIDRAPNPMELLTISLPREVKPGDMIRACKEGLVNLVQKDGTSKTV
 RT

SEQ ID NO: 40 polynucleotide sequence encoding GAS 130

ATGTCACATATGAAAAACGTCCTCCGAGGTCTTATCACCTGCTGGAACACTTGAAAAATTAAAGTTGCGAT
 TGACTATGGCGCAGATGCTGTTTGTGTTGGAGGGCAGGCCCTATGGCCTAAGAAGCCGCGCTGGTAACTTCT
 CTATGGGAAGAATGCAAGAAGGCATGATTATGACATGCGCGTGGAGCTAAGGTCTATGTTGCTGCTAAC
 ATGGTTAACCACGAAGGGAACGAAATGGTGGCGGCGAGTGGTTTCGTCAACTGCGTGATATGGGGCTTGA
 TGCGGTCAATTGTTTCAGATCCAGCCTTGATTGTTATTTGTTCAACAGAAGCCCCAGGTTTGGAAATTCATT
 TGTCACGCAAGCTTCATCTACCAATTACGAGACCTTGAATTTTGGAAAGCCATGGGCTTGACCCGAGTT
 GTTTTAGCTCGCGAGGTTAATATGGCCGAGTTAGCAGAAATCCGCAAGCGGACAGATGTGGAAATGAAGC
 CTTTGTCCATGGAGCCATGTGTATCTCTTATTCAGGCCGCTGTGTTTGTCAAACACATGAGTCACCGTG
 ATGCCAACAGGGGCGGCTGCTCACAGTCTTGCCGCTGGGAAGTATGATTTGTATGACATGCCATTTGGAGGA
 GAGCGCGCTCTCTTAAAGGGGAAATTCAGAAGACTATTCTATGTCCTCTGTTGACATGTGTATGATTGA
 CCATATTCTCTGACCTGATTGAAATGGGTTGATAGCTTAAAAATTTGAAGGCCGAATGAAATCTATCCACT
 ACCTCTCAACCGTAACCAACTGTTACAAGGCGGCTGTAGGTGCTTACATGGAAAGCCAGAGCTTTTAT
 GCTATCAAAGAGGAATTGATTGACGAGTTGTGGAGGTTGCCAGCGCGAGTTGGCTACAGCTTTTACTA
 TGGTATCCCAACTGAAATGAACAATTATTTGGTGTCTGCCGCAAAATTCACAATATAAATTTGTCTGGAG
 AAGTAGTTGCCCTTGACTCAGCTAGCATGACAGCGACCATTCGTCAGCGTAATGTCATCATGGAAGGCGAT

CGGATTGAATGTTATGGACCAGGTTTCCGTCATTTTGAAACGGTTGTTAAGGACTTACATGATGCGGATGG
 CCAAAAGATTGACCGTGCCCAAAATCCAATGGAACCTTAACCATCTCTTTACCGAGAGAAGTTAAGCCAG
 GGGATATGATTAGGCTTGCAAGGAAGGTTCTGGTTAACCTCTATCAAAAAGATGGCACCAGTAAAACTGTT
 AGAACATAG

SEQ ID NO: 41 amino acid sequence comprising GAS 277

MTMQKTIISLLSLALLIGLLGTSGKAI SVYAQDQHTDNVIAESTISQVSVEASMRGTEPYIDATVTTDQPV
 RQPTQATITLKDASDNTINSWVYTMAAQQRRFTAWFDLTGQKSGDYHVT/VVHTQEKAVTGQSGTVHFDQN
 KARKTFTNMQQKDTSKAMTNSVDVDTKAQTNQSANQEI DSTSNPFRSATNHRSTSLKRSTKNEKI/PTASN
 SQKNGSNKTKMLVDKEEVKPTSKRGFPWVLLGLVVS LAAGLFIAIQKVSRRK

SEQ ID NO: 42 polynucleotide sequence encoding GAS 277

ATGACAACTATGCAAAAAACAATTAGCTTATTATCACTAGCTTTACTTATTGGTTTGTGGGGACTTCTGG
 CAAAGCCATATCTGTGTATGCACAAGATCAGCACACTGATAATGTTATAGCTGAATCAACTATTAGTCAGG
 TCAGTGTGTAAGCCAGTATGCGTGGAACAGAACCTTATATGATGCTACAGTCACCACAGATCAACCTGTC
 AGACAACCAACTCAGGCAACGATAACACTTAAAGACGCTAGTGATAATACTATTAAATAGTTGGGTATATAC
 TATGGCAGCGCAACAGCGTCGTTTTACAGCTTGTTGTTGATTAACTGGACAAAAGAGTGGTGACTATCATG
 TAACTGTCAACCGTTTACTACTCAAGAAAAGGCAGTAACGTGCAATCAGGAAGTGTTCATTTTGATCAAAAC
 AAAGCTAGAAAAACCAACTAATATGCAACAAAAGGATACTTCTAAAGCAATGACGAATTCAGTCGATGT
 AGACACAAAAGCTCAACCAATCAATCAGCTAACCAAGAAATAGATTCTACTTCAAATTCCTTTCAGATCAG
 CTACTAATCATCGATCAACTTCCTTAAAGCGATCTACTAAAAATGAGAACTTACACCAACTGCTAGTAAT
 AGUCAAAAAACCGTAGCAACAAGACAAAATGCTAGTGGACAAAGAGGAAGTAAACCTACTTCAAAAAG
 AGGATTCCCTTGGGTCTTATTAGGTCTAGTAGTCAGTTTACCTGCAGGTATTATTATAGCTATTCAAAAAG
 TATCTAGACGAAAATAA

SEQ ID NO: 43 amino acid sequence comprising N-terminal leader sequence of GAS 277 TTMQKTIISLLSLAL-
 LIGLLGTSGKAI SVYA

SEQ ID NO: 44 amino acid sequence comprising fragment of GAS 277 where N-terminal leader sequence is removed

QDQHTDNVIAESTISQVSVEASMRGTEPYIDATVTTDQPVVRQPTQATITLKDASDNTINSWVYTMAAQORR
FTAWFDLIGQKSGDYHVTVTVHTQEKAVTQSGSTVHFDQNKARKTPTNMQQKDTSKAMTNSVDVDTKAQTN
QSANQEIDSTSNPFRSATNHRSTSLKRSTKNEKLTPASNSQKNGSNKTKMLVDKEEVKPTSKRGFPWVLL
GLVVSALAAGLFIAIQVSRRK

SEQ ID NO: 45 amno acid sequence comprising GAS 236

MTQMNYTGKVKRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGMLLSAFHMYEKELD
KVRFGIHTGHLGFYTDYRDFEVDKLIIDNLRKDKGEQISYPILKVAITLDDGRVVKARALNEATVKRIEKT
MVADVIINHVKFESFRGDGIVSVSTPTGSTAYNKS LGGAVLHPTIEALQLTEISSLNNRVFRTLGSIIIPK
KDKIELVPRKRLGIYTTISIDNKTYQLKNVTKVEYFIDDEKIHVSSPSHTSFWERVKDAFIGEIDS

SEQ ID NO: 46 polynucleotide sequence encoding GAS 236

ATGACACAGATGAATTATACAGGTAAGGTAAAACGAGTTGCTATTATTGCAAAATGGTAAGTACCAAAGTAA
ACGCGTCGCTCCAACTTTTCTCCGTATTTAAAGATGATCCTGATTCTATCTTTCAAAGAAAAATCCGG
ATATTGTGATTTCTATTGGCGGAGATGGGATGCTCTTATCTGCTTTCACATGTATGAAAAAGAATTAGAT
AAGGTACGTTTTGTAGGAATCCACACCGGTCATCTTGGCTTTTATACCGATTATAGGGATTGGAAGTTGA
TAAATTAATTGATAATTTAAGARAAGACAAGGGAGAACAAATCTCTTATCCGATTTTAAAAGTTGCTATTA
CTTTAGATGATGGTCTGTGGTTAAAGCGCGTGCTTTGAATGAAGCGACGGTTAAGCGTATTGAAAAACG
ATGGTAGCAGATGTTATTATTAACCATGTCAAATTTGAAAGCTTCCGAGGTGATGGGATTTCAGTATCGAC
CCCGACAGGGAGCACAGCTACAATAAATCTTTAGGTGGTGCTGTCTTGCATCCGACGATTGAAGCGCTGC
AATTGACGGAATTTCCAGTCTTAATAACCGTGCTCTTTAGAACCTTGGGCTCATCAATCATTATCCCCAAA
AAAGATAAGATTGAGTTAGTGCCAAAACGATTAGGAATTTATACCATTTCCATTGATAATAAACCTATCA
GTTAAAAAATGTGACGAAGGTGGAGTATTTTATCGACGATGAGAAAATTCATTTGTTTCTCTCCGAGTC
ATACGAGCTTTTGGGAAAGGGTCAAGGATGCCTTTATTTGGAGAGATTGACTCATGA

SEQ ID NO: 47 amino acid sequence comprising N-terminus leader sequence of GAS 236 MTQM

SEQ ID NO: 48 amino acid sequence comprising a fragment of GAS 236 where the N-terminal leader sequence is removed

NYTGKVKRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGMLLSAFHMYEKELDKVRF
VGIHTGHLGFYTDYRDFEVDKLIIDNLRKDKGEQISYPILKVAITLDDGRVVKARALNEATVKRIEKT
MVADVIINHVKFESFRGDGIVSVSTPTGSTAYNKS LGGAVLHPTIEALQLTEISSLNNRVFRTLGSIIIPKDKI
ELVPRKRLGIYTTISIDNKTYQLKNVTKVEYFIDDEKIHVSSPSHTSFWERVKDAFIGEIDS

SEQ ID NO: 49 amino acid sequence comprising GAS 389

MRNEMAKIMNVTGEEVIALAATYMTKADVAFAKALAYATAAHFYQVRKSGEPYIVHPIQVAGILADLHLD
AVTVACGFLHDVVEDTDITLDEIEADFGHDARDIVDGVTKLGEVEYKSHEEQLAENHRKMLMAMSKDIRVI
LVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGISRIKWELEDLAFRYLNETEFYKISHMMKE
KRREREALVEAIVSKVPTYTQQLFGDVYGRPKHIYSIYRKMROKKRFDQIFDLIAIRCVMETQSDVYA
MVGYIHELWREMPGRFKDYIAAPKANGYQSIHTTVYGPKGFIQIRTKDMHQVAEYGVAAHWAYKKGVRG
KVNQAEQAVGMNWIKEVELQDASNGDAVDVDSVKEDIFSERIYVFTPTGAVQELPKESGPIDPAYAHT
QTGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVKLVKTNKARNKIROFFKNQDKELSVNK
GRDLIVSYFQROGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFGDISPISVFNKLTEKERREERAKA
KAEAEELVKGGEVKHENKDVLRSENGVVIQGASGLLMRIAKCCNPVPGDPIDGYITTKGRGIAIHRSDCH
NIKSQDGYQERLIEVEWDLNNSKDYQAEIDTYGLNRSGLLNDVLQILSNSTKSISTVNAQPTKDMKFANI
HVSFGIPNLTHLTTVVEKIKAVPDVYSVKRTNG

SEQ ID NO: 50 polynucleotide sequence encoding GAS 389

ATGAGGAACGAAATGGCAAAATAATGAACGTAACAGGAGAAGAAATCATTGCCTTAGCGGCCACCTATAT
 GACCAAGGCTGATGTGGCTTTTGTGGCAAAGGCTTTAGCATATGCAACAGCGGCCCATTTCTACCAAGTGA
 GAAAGTCAGGCGAACCCATATATCGTCCATCCGATTACAGGTGGCGGGGATTCTGGCTGATTGCATCTGGAT
 GCTGTGACAGTTGCTTGTGGCTTTTACATGATGTGCTAGAAAGATACGGATATTACCTTAGATGAGATCGA
 AGCAGACTTTGGCCATGATGCTCGTGATATCGTTGATGGTGTCAACCAAGTTAGGTGAAGTTGAGTACAAAT
 CTCATGAGGAGCAACTCGCCGAAAACCATCGCAAAATGCTGTATGGCTATGTCCAAAGATATTCCGCTGATT
 TTGGTGAAATTGGCTGACCGCTGCATAATATGCGCACCCCTCAAACATTTGCGCAAGGACAAACAAGAGCG
 CATTTCGCGCGAAACCATGGAAATCTATGCCCCCTTGGCGCATCGTTTGGGGATTAGTCGCATCAAATGGG
 AACTAGAAGATTGGCTTTTCTGTACCTCAATGAAACCGAATTTTACAAAATTTCCCATATGATGAAAGAA
 AAACGTCGCGAGCGTGAAGCTTTGGTACAGGCTATTGTCACTAAGCTCAAAACCTATACGACACAACAAGG
 GTTGTGGAGATGTGTATGGCGGACCAAAACACATTTTATTCGATTTATCGGAAAATGCGGGACAAAAGA
 AACGATTCGATCAGATTTTGTATCTGATTGCCATTCTGTTGTGTATGGAAACGCAAGCGATGTCTATGCT
 ATGGTTGGCTATATTATGAGCTTTGGCGTCCCATGCCAGGCGCTTCAAGGATTATATTGCAGCTCCTAA
 AGCTAATGGCTACCACTTATTCATACCACCGTGTATGGGCGCAAAAGGACCTATTGAGATTCAAATCAGAA
 CTAAGGACATGCATCAAGTGGCTGAGTACGGGGTGTGCTGCTCACTGGGCTTATAAAAAGGCGTGCCTGGT
 AAGGTCAATCAAGCTGAGCAAGCGGTTGGCATGAACGGATCAAAGAGCTGGTAGAATTGCAAGATGCCCTC
 AAATGGCGATCGAGTGGACTTTGTGGATTCCGCTCAAAGAACATTTTTTCTGAACGGATTATGTCTTTA
 CACCGACAGGGGCGCTTACAGGAGTTACCAAAAGAAACAGGTCCATTGATTGTGCTTATCGGATCCATACG
 CAAATCGGTGAAAAAGCAACAGGTGCCAAAGTCAATGGACGTATGGTTCTCTCACTGCCAAGTTAAAAAC
 AGGAGATGTGGTTGAAATCATCACCAATGCCAATTCCTTTGGCCCTAGTCGAGACTGGGTAAACTGGTCA
 AAACCAATAAGGCTCGCAACAAAATTCGTGAGTTCTTTAAAAATCAAGACAAGGAATTGTCAGTGAATAAA
 GGCCGTGATTGTGGTGTCTTATTTTCAAGAGCAGGGCTACCTTGCCAATAAATACCTTGACAAAAACG
 CATTGAAGCCATCCCTTCCAAAAGTCAGTGTGAAGAGCGAAGAATCACTCTATGCAGCCGTTGGGTTGGTGTG
 ACATTAGCTCTATCAGTGTCTTTAACAAGTTAACCAGAAAAAGAGCGCCGTGAAGAAGAAAGGGCCAGGCT
 AAAGCAGAGCTGAAGAATTGGTTAAGGGCGGTGAGGTCAAACACGAAAAACAAGATGTGCTCAAGGTTCTG
 CAGTGAATAAGGAGTCATTATCCAAGGAGCATCAGGCCCTCTGTATGCGGATTGCCAAGTGTGTGAATCCTG
 TACCTGGTGATCTTATGACGGCTACATTACCAAGGGCGTGGCATTGCGATTACAGATCGGACTGTCTAT
 AACATTAAAGAGTCAAGATGGCTACCAAGAACGCTTGATTGAGGTGAGTGGGATTGGGACAATTGAGATAA
 AGATTATCAGGCTGAAATTGATATCTATGGGCTCAATCGTAGTGGTCTGCTTAATGATGTGCTCCAAATTT
 TATCAAATCAACCAAGAGCATATCGACAGTCAATGCTCAGCCGACCAAGGACATGAAGTTTGCTAATAATT
 CACGTGAGCTTTGGCATTCCAAATCTGACGCATCTGACCCTGTGTGCGAAAAATCAAGGCAGTTCAGA
 TGTTTATAGCGTGAAGCGGACCAATGGCTAA

SEQ ID NO: 51 amino acid sequence comprising GAS 504

MKTRITELLNIDYPIFOGGMAWVADGDLGAVSNAGGLGIIGGNAPKEVVKANIDRVKAITDRPFVGNIM
 LLSFFADDIVDLVIEGVKVVTTGAGNPGKYMERLHQAGIIVPVVPSVALAKRMEKLGVDVIAEGNEAG
 GHIGKLTMSLVRFVRAVSLPVLIAAGGIADGHGAAAFMLGAEAVQIGTRFVVAKESNAHQNFKDILAA
 KDIDTVISAQVVGHPVRSIKNKLTSAYAKAEKAFLLIGQKTATDIEBMGAGSLRHAVIEGDVNVGSVMAGQI
 AGLVRKEESCETILKDIYYGAARVIQNEAKRWQSVSIEK

SEQ ID NO: 52 polynucleotide sequence encoding GAS 504

ATGAAACACGATATTACAGAATTACTTAATATTGATTACCCCATTTTCAAGGAGGAATGGCTTGGGTTGC
 TGATGCTGATTTAGCAGCTGCAGTTTCTAATGCTGGTGGTTTAGGCATTATAGGTGGTGGCAATGCTCCCA
 AAGAAGTCGTTAAAGCTAATATTGATCGTGTCAAAGCTATTACTGATAGACCTTTTGGGGTTAATATCATG
 CTTTTATCTCCTTTTGGCTGATGATATCGTGTGATCTGGTCATTGAAGAAGGTGTTAAAGTAGTAACAACAGG
 CGCAGGAAATCCAGCAAGTATATGGAAAGACTGCACAGGCGGGTATAATCGTTGTTCTGTTGTCCCAA
 GCGTTGCGCTAGCCAAACGATATGGAAAAGCTTTGGGGTAGATGCTGTTATTGCTGAGGGTATGGAAGCTGGA
 GGACATATTGGCAAGTTAACGACTATGCTTTAGTAAGACAAGTTGTTGAAGCGGTTTCGATTCTCTGTCTAT
 TGCGGCAGGTGGTATAGCTGATGGTTCATGGTGCAGCAGCAGCATTATGTTAGGAGCAGAGGCTGTTCAAA
 TTGGAACCTCGCTTTGTTGTTGCTAAAGAATCCATGCTCACCAAAATTTTAAAGATAAAATCTTAGCAGCA
 AAAGATATTGATACGGTGATTCTTGGCGCAGGTTGTGGGCCACCTGTCCGTTCTATTAAAAATAAATTGAC
 CTCAGCTTACGCTAAAGCAGAAAAAGCATTTTAAATTGGTCAAAAAACAGCTACTGATATTGAAGAAATGG
 GAGCAGGATCGCTTCGACACGCTGTTATTGAAGGCGATGTAGTCAATGGATCTGTTATGGCTGGCCAAATTT
 GCAGGGCTTGTGAGAAAAGAAGAAAGCTGTGAAACGATTTTAAAGATATTTATTATGGTGCAGCTCGTGT
 TATTCAAATGAAGCTAAGCGCTGGCAATCTGTTTCAATAGAAAAGTAG

SEQ ID NO: 53 amino acid sequence comprising GAS 509

MTKIYKTTITELVGQTPPIIKLNRLLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAEGLISPGDVIIIE
PTSGNTGIGLAWVGAAGKYRVIIIVMPETMSLERRQIIQAYGAELVLTPGAEGMKGAIKARTLAIELGAW
MPMQFNPNANPSIHEKTTAQEILEAFKEISLDAFVSGVGTGGTSLGVSHVLKKNANPETVIYAVEAEESAV
LSGQEPGPHKIQGISAGFIENLTDTKAYDQIIIRVKS KDALETARLTGAKEGFLVGISSGAALYAAIEVAK
QLGKCKHVLITLPDNGERYLSTELYDVPVIKTK

SEQ ID NO: 54 polynucleotide sequence encoding GAS 509

ATGACTAAAATTACAAAACATAACAGAATTAGTAGGTCAAACACCTATTATCAAACCTAACCGTTTAA
TTCCAAACGAAGCTGCTGACGTTTATGTAAATTAGAAGCTTTTAAACCCAGGATCTTCTGTTAAAGATCG
TATTGCTTTATCGATGATTGAAGCTGCTGAAGCTGAAGCTCTGATAAGTCTGGTGACGTTATTATCGAA
CCAACAAGTGGTAATACAGGTATTGGTCTTGCCATGGGTAGGTGCTGCTAAAGGGTATCGAGTCATTATG
TTATGCCCGAAACTATGAGCTTGGAAAGACGGCAAATCATTCAGGCTTATGGTGCAGAGCTTGTCTTAAC
ACCTGGAGCAGAAGGTATGAAGGGGGCTATTGCAAAAGCTGAAACTTTAGCAATAGAACTAGGTGCTTGG
ATGCCTATGCAATTTAATAACCCCTGCCAATCCAAGCATCCATGAAAAAACACAGCTCAAGAAATTTTGG
AAGCTTTTAAGGAGATTTCTTTAGATGCATTCTGCTATCTGGTGTGGTACTGGAGGAACACTTTCTGGTGT
TTACATGCTCTTGAAGAAAGCTAACCCCTGAACTGTTATCTATGCTGTTGAAGCTGAAGAAATCTGCTGTC
TTATCTGGTCAAGAGCCTGGACCACATAAAATCAAGGTATATCAGCTGGATTTATCCCAAACACGTTAG
ATACCAAAGCCTATGACCAATTTATCCGTGTAAATCGAAAGATGCTTTAGAACTGCTCGACTAACAGG
AGCTAAGGAAGGCTTCCTGGTTGGGATTTCTTCTGGAGCTGCTCTTTACGCCGCTATTGAAGTCGCTAAA
CAGTTAGGAAAGGCAACATGTGTTAACTATTTTACCAGATAATGGCGAACGCTATTTATCGACTGAAC
TCTATGATGTACCAGTAATTAAGACGAAATAA

SEQ ID NO: 55 amino acid sequence comprising C-terminus transmembrane region of GAS 509
FLVGISSGAALYAAIEVAKQLGKCKHVLITLPDNGERYLSTELYDVPVIKTK

SEQ ID NO: 56 amino acid sequencing comprising a fragment of GAS 509 where the C-terminal transmembrane region is removed

MTKIYKTTITELVGQTPPIIKLNRLLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAEGLISPGDVIIIEP
TSGNTGIGLAWVGAAGKYRVIIIVMPETMSLERRQIIQAYGAELVLTPGAEGMKGAIKARTLAIELGAWMP

MQFNPNANPSIHEKTTAQEILEAFKEISLDAFVSGVGTGGTSLGVSHVLKKNANPETVIYAVEAEESAVLSG
QEPGPHKIQGISAGFIENLTDTKAYDQIIIRVKS KDALETARLTGAKEG

SEQ ID NO: 57 amino acid sequence comprising GAS 366

MKVISNFQNKILILGLAKSGEAAAKLLTKLGAIVTVNDSKPFQNPAAQALLEEGIKVICGSHPVELLDE
NFEYMKVKNPGIPYDNPVYKRALAKEIPILTEVELAYFVSEAPIIGITGSNGKTTTMTIADVLNAGGQSAL
LSGNIGYPASKVVQKAIAGDTLVMELSSFQLVGVNAFRPHIAVITNLMPTHLDYHGSFEDYVAAKWMIQAO
MTESDYLLILNANQEISATLAKTTKATVIPFSTQKVVDGAYLKDGLYFKEQAIIAATDLGVPGSHNIENAL
ATIAVAKLSGIADDIIAQCLSHFGGVKHLQRVGQIKDITFYNSKSTNIIATQKALSGFDNSRLILIAGG
LDRENEFPDDLVPDLLGLKQMIILGESAEEMKRAANKAEVSYLEARNVAEATELAFKLAQTGDTILLSPAN
ASWDMYPNFEVRGDEFLATFDCLRGDA

SEQ ID NO: 58 polynucleotide sequence encoding GAS 366

ATGAAAGTGATAAGTAATTTTCAAAACAAAAAATATTAATATTGGGGTTAGCCAAATCGGGCGAAGCAGC
AGCAAAATPATTTGACCAAACTTGCTGCTTTAGTGACTGTTAATGATAGTAAACCATTTGACCAAAATCCAG
CGGCACAAGCCTTGTTGGAAGAGGGGATTAAAGTTCATTTGTGGTAGCCACCCAGTAGAATTATTAGATGAG
AACTTTGACTACATGGTTAAAAACCTTGGGATTTCCTTATGATAATCCTATGGTTAAACGCGCCCTTGCAAA
GGAAATFCCCATCTTGACTGAAGTAGAATTGGCTTATTTTCGTATCTGAAGCGCCTATTATCGGGATTACAG
GATCAAAACGGAAGACAACCACAACGACAATGATTGCCGATGTTTTGAATGCTGGCGGGCAATCTGCACTC
TTATCTGAAACATTGGTTATCCTGCTTCAAAAGTTGTTCAAAAAGCAATTGCTGGTGATACTTTGGTGAT
GGAAATGTCTCTTTTCAATTAGTGGGAGTGAATGCTTTTCGCCCTCATATTGCTGTCACTACTAATTTAA
TGCCGACTCACCTGGACTATCATGGCAGTTTGGAGATTATGTTGCTGCTAAATGGATGATTCAAGCTCAG
ATGACAGAATCAGACTACCTTATTTTAAATGCTAATCAAGAGATTTCAGCAACTCTAGCTAAGACCACCAA
AGCAACAGTGATTTCCTTTTCAACTCAAAAAGTGGTTGATGGAGCTTATCTGAAGGATGGAATACTCTATT
TTAAAGAACAGGCGATTATAGCTGCAACTGACTTAGGTGTCCAGGTAGCCACAACATTGAAATGCCCTA
GCAACTATTGCAGTTGCCAAGTTATCTGGTATTGCTGATGATATTATTGCCAGTGCCCTTTCACATTTTGG
AGGCGTTAAACATCGTTTGCACGGGTTGGTCAAATCAAGATATACCTTCTACAATGACAGTAAGTCAA
CCAATATTTTAGCCACTCAAAAAGCTTTATCAGGTTTGTATAACAGTCGCTTGATTTTGATTGCTGGCGGT
CTAGATCGTGCCAATGAATTTGACGATTGGTGCCAGACCTTTTAGGACTTAAGCAGATGATTATTTTGGG
AGAATCCGCAGAGCGTATGAAGCGAGCTGCTAACAAGCAGAGGCTCTCTTATCTTGAAGCTAGAAATGTGG
CAGAAGCAACAGAGCTTGCTTTTAAGCTGGCCCAACAGGCGATACTATCTTGCTTAGCCCAAGCAATGCT
AGCTGGGATATGTATCCTAATTTTGAAGTTCGTGGGGATGAATTTTGGCAACCTTTGATTGTTTAAGAGG
AGATGCCCTAA

SEQ ID NO: 59 amino acid sequence comprising N-terminal leader sequence of GAS 366
 MKVISNFQNKILILGLAKSGEAAA

SEQ ID NO: 60 amino acid sequence comprising a fragment of GAS 366 where the N-terminal leader sequence is removed

KLLTKLGLVTVNDSKPFQNPAAQALLEEGIKVICGSHFVELLDENFEYMKVKNPGIPYDNPVVKRALAKE
 IPILTEVELAYFVSEAPIIGITGSGNKTFTTMIADVLNAGGQSALLSGNIGYPASKVVQKAIAGDTLVME
 LSSFQLVGVNAFRPHIAVITNLMPHLDYHGSFEDYVAAKWMIIAQMTESDYILNANQEISATLAKTTKA
 TVIPIFSTQKVVDGAYLKDGIILYFKEQAIIAATDLGVPGSHNIENALATIAVAKLSGIADDIIAQCLSHFGG
 VKHRLQRVGQIKDITFYNDSKSTNILATQKALSGFDNSRLILIAAGGLDRGNEFDDLVPDLLGLQMIILGE
 SAERMKRAANKAEVSYLEARNVAEATELAFKLAQTGDTILLSPANASWDMYPNFEVRGDEFLATFDCLRGD
 A

SEQ ID NO: 61 amino acid sequence comprising GAS 159

MRKLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLVIYNWGPYIDPALLKKFTKETGLEVQYETFDNS
EAMYTKIKQGGTTYDIAVPSDYITDKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPPQNDYSLPYFWGTV
GIVYNDQLVDKAPMHWEDLWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAERKLQQLTPNV
KAIVADEMKGYMIQGDAAIGITPSGEASEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKAYAFNLF
INRPENAAQNAAYIGYATPNKKAKALLPDEIKNDPAFYPTDDIIKKLEVYDNLGSRWLGIYNDLYLOFKMY
RK

SEQ ID NO: 62 polynucleotide sequence encoding GAS 159

ATGCGTAAACTTTATTCCTTTCTAGCAGGAGTTTGGGTGTTATTGTTATTTTAAACAAGTCTTTCTTTTCAT
CTTGCGAGAAAAATCGGTTCTGCTAGTCAATCGGATAAAATTAGTTATTTATAACTGGGAGATTACATTG
ATCCAGCTTTGCTCAAAAAATTCACCAAGAAACGGGCATTGAAGTGCAGTATGAAACTTTTCGATTCCAAT
GAAGCCATCTACACTAAAAATCAAGCAGGGCGGAACCACTTACGACATTGCTGTTCTTAGTGATTACACCAT
TGATAAAATGATCAAAAGAAACCTACTCAATAAGCTTGATAAGTCAAAATTAGTTGGCATGGATAATATCG
GGAAAGAATTTTATAGGGAAAAGCTTTGACCCACAAAACGACTATTCTTTGCCCTATTTCTGGGGAACCGTT
GGGATGCTTTATAATGATCAATTAGTTGATAAGGCGCCTATGCACTGGGAAGATCTGTGGCGTCCAGAATA
TAAAAATAGTATTATGCTGATTGATGGAGCGCGTGAAATGCTAGGGGTTGGTTTAACAACTTTTGGTTATA
GTGTGAATTTCTAAAAATCTAGAGCAGTTGCAGGCAGCCGAGAGAAAACCTGCAGCAGTTGACGCCGAATGTT
AAAGCCATTGTAGCAGATGAGATGAAAGGCTACATGATTCAAGGTGACGCTGCTATTGGAATTACCTTTTC
TGGTGAAGCCAGTGAGATGTTAGATAGTAACGAACACCTTCACTACATCGTGCCTTCAGAAGGGTCTAACCC
TTTGGTTTGATAATTTGGTACTACCAAAAACCATGAAACACGAAAAAGAAGCTTATGCTTTTTTGAACCTTT
ATCAATCGTCTGAAAATGCTGCGCAAAATGCTGCAATATTTGGTTATGCGACACCAAAATAAAAAAGCCAA
GGCCTTACTTCCAGATGAGATAAAAAATGATCCTGCTTTTATCCAACAGATGACATTATCAAAAAATTGG
AAGTTTATGACAATTTAGGGTCAAGATGCTTGGGGATTATAATGATTTATACCTCCAATTTAAATGTAT
CGCAATAA

SEQ ID NO: 63 amino acid sequence comprising N-terminal leader sequence of GAS 159
MRKLYSFLAGVLGVIVILTSLSFI

SEQ ID NO: 64 amino acid sequence comprising a fragment of GAS 159 where the N-terminal leader sequence is removed

LQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSEAMYTKIKQGGTTYDIAVPSDYTI
DKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPPQNDYSLPYFWGTGIVYNDQLVDKAPMHWEDLWRPEY
KNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLQQLTENVKAIIVADEMKGYMIQGDAAGITTF
GEASEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFNFINRPNENAAQNAAYIGYATPNKKAK
ALLPDEIKNDPAFYPTDDIIEKKLEVYDNLGSRWLGIYNDLYLQFKMYRK

SEQ ID NO: 65 amino acid sequence comprising C-terminal hydrophobic sequence of GAS 159
WLGIYNDLYLQFKMYRK

SEQ ID NO: 66 amino acid sequence comprising a fragment of GAS 159 where the C-terminal hydrophobic region is removed

MRKLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSE
EAMYTKIKQGGTTYDIAVPSDYTIIDKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPPQNDYSLPYFWGT
GIVYNDQLVDKAPMHWEDLWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNV
KAIIVADEMKGYMIQGDAAGITTFSGEASEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFNFIN
RPNENAAQNAAYIGYATPNKKAKALLPDEIKNDPAFYPTDDIIEKKLEVYDNLGSR

SEQ ID NO: 67 amino acid sequence comprising a fragment of GAS 159 where the N-terminal leader sequence and the C-terminal hydrophobic region is removed

LQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSEAMYTKIKQGGTTYDIAVPSDYTI
DKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPPQNDYSLPYFWGTGIVYNDQLVDKAPMHWEDLWRPEY
KNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNVKAIIVADEMKGYMIQGDAAGITTF
GEASEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFNFINRPNENAAQNAAYIGYATPNKKAK
ALLPDEIKNDPAFYPTDDIIEKKLEVYDNLGSR

SEQ ID NO: 68 amino acid sequence comprising GAS 217

MAQRITVITGASGGLAQAIKQPKEDSLILLGRNKRLEHCYQHIDNKECLELDITNPVAIEKMVAQIYQ
RYGRIDVLINNAGYGAFKGFEEFSAQEIADMFQVNTLASIHFACLIGQKMAEQGQHLINIVSMAGLIASA
KSSIYSATKFAIGFSNALRLELADKGVVTVTPNGRIATKFPDQADPSGHYLESVVGKFTLQPNQVAKRLV
SIIGKNKRELNLFFSLAVTHQFYTLFPKLSDYLARKVFNYK

SEQ ID NO: 69 polynucleotide sequence encoding GAS 217

ATGGCCACAAAGAATCATTTGTTATCACGGGAGCTTCTGGAGGACTGGCTCAGGCAATTGTTAAGCAGTTACC
CAAGGAAGACAGCTTGATTTTATTAGGACGTAACAAAGAACGCCTAGAACACTGTTATCAGCATATTGACA
ACAAAGAATGCCTCGAGTTGGATATTACCAATCCAGTAGCCATTGAGAAAATGGTCGCCCAGATTTACCAG
CGCTATGGCCGTTATGATGTCCTTGATTAATAATGCTGGCTACGGAGCTTCAAAGGCTTTGAAGAGTTTTC
TGCCCAAGAAATAGCTGATATGTTTCAGGTTAACACCTAGCGAGCATTCACCTTTGCTTGCTTGATTGGTC
AGAAAATGGCAGAGCAGGGGCAAGGTCACCTTATTATATTGTGTCCATGGCAGGCTTGATTGCGTCAGCC
AAATCGAGCATTTATTCAGCCACCAAGTTTGGCCCTTATCGGATTTTCCAATGCCCTTCGCTTAGAATTAGC
GCATAAAGGGGTTTACGTGACCACCGTGAATCCAGGTCCCATTTGCCACCAAGTTTGTGACCAAGCTGACC
CGTCTGGACATTTATTTGGAAAGCGTTGGTAAATTTACTCTCCAACCAATCAAGTGGCTAAGCGTTTGGTT
TCTATTATCGGGAATAAACGAGAATTGAATTTGCCCTTTAGTTTAGCGGTGACCCATCAATTTTACAC
CCTTTTCCCTAAATATCTGATTATCTTGAAGAAAGGTATTTAATTATAAATGA

SEQ ID NO: 70 amino acid sequence comprising GAS 309

MIEKYLESSIESKQQLIVLFFPKTSYLPITEVAEKTGLTFLQLNHYCEELNAFFPGSLSMITQKRMISCOFT
HPFKETYLYQLYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNSSAYRMREALIPLLRNFELKLSKNKI
VGEEYRIRYLIALLYSKFGIKVYDLTQQDKNTIHSFLSHSSHLKTSFWLSEFSFYDILLALSWKRHOFS
VTIPQTRIFQQLKLFVYDSLKKSSHDIIETCYQLNFSAGDLDYLYLIYITANNSFASLQWTPEHIROYCQ
LPEENDYFRLLLNPIITLLPNLKEQKASLVKALMFFSKSFLNLQHFIPETNLFVSPYYKGNQKLYTSLKL
IVEEMAKLPGKRDNLNKKHFLFCHYVEQSLRNIQPLVVVFVASFNAHLITDSFPFRYFSDKSIDPHSY
YLLQDNVYQIPDLKPLDLVITHSQLIPFVHHELTGKIABEISFDESILSIQELMYQVKEEFQADLTQOLT

SEQ ID NO: 71 polynucleotide sequence encoding GAS 309

TTGATAGAAAAATACCTTGAATCATCAATCGAATCAAAATGTCAGTTAATTGCTTGTGTTTTTAAGACATC
TTATTTGCCAATAACTGAGGTAGCAGAAAAAAGTGGCTTAACCTTTTACAACTAAACCATTATTGTGAGG
AACTGAATGCCTTTTCCCTGGTAGTCTGTCTATGACCATCCAAAAAAGGATGATATCTTGCCAAATTTACA
CATCCTTTTAAAGAACTTATCTTTACCAACTCTATGCATCATCTAATGTCTTACAATTACTAGCCTTTT
AATAAAAAATGGTTCCCACTCTCGTCCCTTACGGATTTTGCAAGAAGTCATTTTATCAAACTCCTCAG
CTTATCGGATGCGCGAAGCATTGATTCCTTTATTAAGAACTTTGAATTAAGAACTCTCTAAGAACAGATT
GTCGGTGAGGAATATCGCATCCGTTACCTCATCGCTCTGCTATATAGTAAGTTTGGCATTAAGTTTATGA
CTTGACCCAGCAAGACAAAACACTATTATAGCTTTTATCCCATAGTCCACCCACCTTAAACCTCTC
CTTGGTTATCGGAATCGTTTCTTCTATGACATTTATTAGCTTTATCGTGGAAAGCGGCATCAATTTTCG
GTAACATATCCCCAAACAGAAATTTTCAACAATTAAGAACTTTTGTCTACGATTCCTTGAAAAAAG
TAGCCATGATATTATCGAACTTACTGCCAACTAACTTTTCAGCAGGAGATTGGACTACCTCTATTAA
TTATATACACCGCTAATAATTCTTTGCGAGCTTACAATGGACACCTGAGCATATCAGACAAATATGTCAA
CTTTTGAAGAAATGATACTTTTCGCTGCTTTTAAATCCTATCATCACTCTTTTACCTAACCTAAAAGA
GCAAAAGGCTAGTTTAGTAAAAGCTCTTATGTTTTTTCAAAATCATTCTGTGTTAATCTGCAACATTTTA
TTCCTGAGACCACTTATTCGTTTCTCCGTACTATAAAGGAAACCAAAACTCTATACGCTCTTAAAGTTA
ATTGTGGAAGAGTGGATGGCCAACTTCCTGGTAAGCGTGACTTGAACCATAAGCATTTTCATCTTTTTTG
CCACTATGTCGAGCAAGTCTAAGAAATATCCAACCTCCTTTAGTTGTTGTTTTCGTAGCCAGTAATTTTA
TCAATGCTCATCTCCTAACGGATCTTTTCCAAGGTATTTCTCGGATAAAAGCATTGATTTTCATTCCTAT
TATCTATTGCAAGATAATGTTTATCAAAATCCTGATTTAAAGCCAGATTGGTCATCACTCACAGTCAACT
GATTCCTTTTGTTCACCATGAACCTACAAAAGGAATGCTGTTGCTGAAATATCTTTTGATGAATCGATTC
TGTCTATCCAAGATTTGATGTATCAAGTTAAAGAGGAAAAATCCAAGCTGATTTAACCAGCAATTAACA
TAA

SEQ ID NO: 72 amino acid sequence comprising GAS 372

MIQIGKLFAGRYRILKSIGRGGMADVYLANDLILDNEDVAIKVLRITNYQTDQVAVARFQREARAMAELNHP
NIVAIRDIGEEDGQQFLVMEYVDGADLKRYIQNHAPLSNNEVVRIMEEVLAMTLAHQKGIHVHDLKPQNI
LLTKBGVVVKTDFGIAVAFATSLTQTNMGLSVHYLSPEQARGSKATIQSDIYAMGLMFLFEMI/TGHI
GDSAVTIALQHFQKPLPSIIENHNVPQALENVIRATAKKLSDRYGSTPEMSRDLMTALSYNRSRERKII
FENVESTKPLPKVASGPTASVKLSPTPTVLTQESRLDQTNQTDALQPTKKKSGRFLGTLFKILFSFFI
VGVALFTYLILTKPTSVKVPNVAGTSLKVAQELYDVGLKVGKIRQIESDTVAEGNVVRTPDKAGTAKRQG
SSITLYVSIKNGFDMENYKGLDYQEAMNSLIETGVKPSKIKIERIVTNEYPTNTVSIQSPSAGDKFNEN
GKSKITLSVAVSDTITMPMVTEYSYADAVNTLTALGIDASRIKAYVPSSSSATGFVPIHSPSSKAIIVSGQS
PYYGTSLSLSDKEISLYLYPEETHSSSSSSSSSTSSSSSSSINDSTAPGSNTELSPSETTSQTP

SEQ ID NO: 73 polynucleotide sequence encoding GAS 372

ATGATTGAGATTGGCAAATATTTGCTGGTCGTTATCGCATTCTGAAATCTATTGGCCGCGCTGGTATGGC
GGATGTTTATTTAGCAAAATGACTTGATCTTGGATAATGAAGACGTTGCAATCAAGGCTTGGCGTACCAATT
ATCAACAGATCAGGTAGCAGTTGCGCGTTTCCAAAGAGAGCGCGGGCCATGGCTGAATTGAACCATCCC
AATATTGTTGCCATCCGGGATATAGGTGAAGAAGACGGACAGCAATTTTAGTAATGGAATATGTGGATGG
TGCTGACCTAAAGAGATACATTCAAAATCATGCTCCATTATCTAATAATGAAGTGGTTAGAATTATGGAAG
AAGTCCFTTCTGCTATGACTTTAGCCACCAAAAAGGAATTGTACACAGAGATTTAAACCTCAAAATATC
CTACTAACTAAGGAGGGTGTGTCAAAGTAAGTATTTGCGCATCGCAGTAGCCTTTGCAGAAACAGCTT
GACACAACTAATTCGATGTTAGGCAGTGTTCATTTACTTGTCTCCAGAACAGGCTCGCGGCTCCAAAGCGA
CGATTCAAAAGTGATATTTATGCGATGGGGATTATGCTCTTTGAGATGTTGACAGGCCATATCCCTTATGAC
GGCGATAGTGCTGTTACGATTGCTTGCACATTTTCAAAAGCCTCTTCCATCTATTATCGAGGAGAACCA
CAATGTGCCACAAGCTTTGGAGAATGTTGTTATTCGAGCAACAGCCAAGAAATTAAGTGATCGTTACGGGT
CAACCTTTGAAATGAGTCTGACTTAATGACGCGCCTTAGTTATAATCGTAGTCGGGAGCGTAAGATTATC
TTTGAGAATGTTGAAAGTACCAAAACCCCTCCCCAAAGTGGCCTCAGGTCCACCGCTTCTGTAAATTTGTC
TCCCCCTACCCCAACAGTGTAAACACAGGAAAGTCGATTAGATCAAACTAATCAACAGATGCTTTACAGC
CCCCCACCACAAAAGAAAAAAGTGCTCGTTTTTAGGTACTTTATTCAAAATTTCTTTTCTTTCTTTATT
GTAGGTGTAGCACTCTTTACTTATCTTATACTAACTAAACCAACTTCTGTGAAAGTTCCTAATGTAGCAGG
CCTAGTCTTAAAGTTGCCAAACAAGAACTGTATGATGTTGGGCTAAAAGTGGGTAAAATCAGGCAAAATG
AGACTGATACGGTTGCTGAGGGAAATGTAGTTAGAACAGATCTTAAAGCAGGAACAGCTAAGAGGCAAGGC
TCAAGCATTACGCTTTATGTCTCAATTGGAACAAGGTTTGTGACATGGAAAACACTACAAAGGACTAGATTA
TCAAGAAGCTATGAATAGTTTGATAGAACTTTAGTGTTCCTCAAAATCAAAAATCAAAATGAGCGCATTTG
TAATTAATGAATATCCTGAAAATACAGTCATCAGTCAATCGCCAAAGTGGGGTGATAAATTAATCCAAAC
GGAAAGTCTAATAATTACGCTCAGTGTGCTGTTAGTGATACGATCACTATGCCTATGGTAACAGAAATAG
TTATGCAGATGCAGTCAATACCTTAACAGCTTTAGGTATAGATGCATCTAGAATAAAAGCTTATGTGCCAA
GCTCTAGCTCAGCAACGGGCTTTGTGCCAATTCATTCTCTCTAGTTCTAAAGCTATTGTGAGTGGTCAATCT
CCTTACTATGGAACGCTCTTGAGTCTGTCTGATAAAGGAGAGATTAGTCTTTACCTTTATCCAGAAGAAAC
ACACTCTTCTAGTAGCTCATCGAGTTCAACGTCAAGTTCAACAGTTCTTCAATAAATGATAGTACTGCAC
CAGGTAGCAACACTGAATTAAGCCCATCAGAACTACTTCTCAACACCTTAA

SEQ ID NO: 74 amino acid sequence comprising GAS 39

MDLILFLVLVLGLGAYLLFKVNLQHQLAQTLEGNADNLSQDMTYQLDTANKQQLLELTQLMNRQQACL
YQQLTDIRDVLHRSLSDSRDRSDKRLKINQQVNQSLKNMQESNEKRLKMRQIVEEKLLETLKNRLHASF
DSVSKQLESVNKGLGEMRSVAQDVGTNLKVLNTRKTRGILGELQLGQIIEDIMTSSQYEREFVTVSGSSER
VEYAIKLPNGQGQGYIYLPIDSKFPLEDYRYLEDAYEVDKLAIEASRKALLAAIKRPAKDIHKKYLNPPPE
TTNFGVMFLPTEGLYSEVVRNASFFDSLRRREENIVVAGPSTLSALLNSLSVGFKTLNIOKNADDISKILGN
VKLEFDKFGLLAKAQKQMNNTANNTLDQLISTRNIVRALNTVETVYQDQATKSLNMLPLEEENEN

SEQ ID NO: 75 polynucleotide sequence encoding GAS 39

ATGGACCTTATCTTGTTCCTTTTGGTCTTGGTCTCTTAGGTTTAGGGGCTTATCTGTTGTTCAAAGTCAA
 CGGCCTTCAACATCAGCTTGCCCAAAACCCTAGAAGGCAACGCGGATAATTTGTCTGACCAATGACCTACC
 AGTTGGATACAGCTAACAAACAACAATTGTAGAGCTAACACAGCTGATGAACCGACAACAAGCAGGCCCTT
 TACCAACAATTAAACAGATATTGCTGACGCTTTGCAACCGTAGTTTGTCTGATAGTAGGGACCGGCTGACAF
 ACGCTTAGAAAAAATTAACCAGCAGGCTCAACCAATCGCTCAAAAAATATGCAAGAATCTAACGAAAAACGT
 TGGAGAAAAATGCGCCAGATCGTTGAAGAAAAATTTGAAGAAACCTTAAAAAATCGCTGACAGCCCTCTTTC
 GATTCTGTATCCAAGCAACTAGAAAGTGTCAATAAAGGCTTGGGAGAAATGCGTAGCGTGGCTCAAGATGT
 GGGTACTTTAAATAAGGTTTGTCCAAATACCAAAACACGAGGCATTTTAGGCGAATCTCAACTAGGCCAAA
 TCATTGAGGATATCATGACATCAAGCCAGTACGAAAGAGAATTTGTAACGGTTAGTGGTCTAGTGAACGC
 GTAGAATATGCGATTAAAGCTCCCAGGAAATGGTCAAGGCGGTTATATTACCTACCGATTGACTCAAAAT
 CCCTCTTGAAGATTATTACCGATTAGAGATGCTTACGAAGTTGGTGATAAACTGGCCATCGAGGCTAGCC
 GAAAAGCACCTTCTGGCAGCTATCAAACGCTTTGCCAAAGACATTATATAAAAGTACTTGAACCCCCAGAG
 ACGACCAATTTCCGAGTTATGTTCTTACCAACAGAGGCTTTTATTACAGAAGTGGTCAGAAATGCGTCTTT
 CTTTGATAGCCCTTCGTCGGGAAGAAAATATTGTGGTTGCAGGCCCTTCGACCCCTGCTGCTTTGCTGAAT
 CCTTATCTGTTGGTTTCAAGACCCCTTAATATCAAAAAAATGCTGATGACATCAGTAAATTTTAGGCAAT
 GTCAAGTTAGAAATTCGATAAATTTGGCGGCCCTGCTTGCCAAGGCTCAAAAACAAATGAATACAGCTAATAA
 TACGCTGGATCAGCTCATTTCAACAAGGACAAATGCCATTGTTGAGGCTTGAATACCGTTGAAACTTATC
 AAGACCAAGCAACAAAATCTCTCTTGAACATGCCCTTATTAGAAGAGGAAAATAATGAAAATTAA

SEQ ID NO: 76 amino acid sequence comprising GAS 42

MTKEKLVAFSQAHAEPAWLQERRLALEAIPNLELPTIERVKFHRWNLGDGTLTENESLASVPDFIAIGDN
 PKLVQVGTQTVLEQLPMALIDKGVVFSDFYTALEEIPEVIEAHFGQALAFDEDKLAAYHTAYFNAAVLV
 PDHLEITTPIEAIFLQSDSDVPFNKHVLVIAGKESKFTYLERFESIGNATQKISANISVEVIAQAGSQIK
 FSAIDRLGPSVTPIYSRRGRLEKDANIDWALAVMNEGNVIADFDSDLIQGQSQADLKVVAASSGRQVQID
 TRVTNYGQRTVGHILQHGVLERGLTFNGIGHILKDAKGADAQQESRVLMLSDAQARADANPILLIDENEV
 TAGHAASIGQVDPEDMYLMSRGLDQETAERLVIRGFLGAVIAEIPISVRQBIKVLDEKLLNR

SEQ ID NO: 77 polynucleotide sequence encoding GAS 42

ATGACAAAAGAAAACTAGTGGCTTTTTTCGCAAGCCACGCTGAGCCTGCTTGGCTGCAAGAACGGCGTTT
 AGCGGCATTAGAAGCCATTCCAAATTTGGAATTACCAACCATCGAAAGGTTAAATTTACCCGTTGGAATC
 TAGGAGATGGTACCTTAACAGAAAATGAAAGTCTAGCTAGTGTTCAGATTTTATAGCTATTGGAGATAAC
 CCAAAGCTTGTTCAGGTAGGCACGCAACAGTCTTAGAAGAGTTACCAATGGCGTTAATTGACAAGGGAGT
 TGTTTTTCAGTGATTTTATACGGCGCTTGAGGAAATCCAGAAAGTAATTGAAGCTCATTTTGGTCAGGCAI
 TAGCTTTTTCATGAAGACAAACTAGCTGCTTACCACACTGCTTATTTTAAATAGCGCAGCGGTGCTCTACGT
 CCTGATCAGTTGGAAATCACAACCTCTATTGAAGCTATTTTCTTACAAGATAGTGACAGTGACGTTCTCTT
 TAACAAGCATGTTCTAGTGATTGCAGGAAAGAAAGTAAGTTCACCTATTTAGAGCGTTTGAATCTATTG
 GCAATGCCACTCAAAAGATCAGCGCTAATATCAGTGTAGAAGTGATTGCTCAAGCAGGCAGCCAGATTAAA
 TTCTCGGCTATCGACCGCTTAGGTCTTTCAGTGACAACCTATATTAGCCGTCGAGGACGTTTAGAGAAGGA
 TGCCAACTTGATGGGCTTAGCTGTGATGAATGAAGGCAATGTCATTGCTGATTTTGACAGTGATTTGA
 TTGGTCAGGGCTCACAAAGCTGATTTGAAAGTTGTTGCAGCCTCAAGTGGTGGTCAGGTACAAGGTATTGAC
 ACGCGCGTGACCAACTATGGTCAACGTACGGTCGGTTCATATTTACAGCATGGTGTGATTTTGAACGTGG
 CACCTTAACGTTTAAACGGGATTGGTCAATTTCTAAAAGACGCTAAGGGAGCTGATGCTCAACAAGAAAGCC
 GTGTTTTGATGCTTTCTGACCAAGCAAGAGCCGATGCCAATCCAATCTCTTAATTGATGAAAATGAAGTA
 ACAGCAGGTCATGACGCTTCTATCGGTCAGGTTGACCTGAAGATATGTATTACTTGATGAGTCGAGGACT
 GGATCAAGAAACAGCAGAACGATTGGTTATTAGAGGATTCTAGGAGCGGTTATCGCTGAAATTCCTATTTC
 CATCAGTCCGCCAAGAGATTATTAAGGTTTATAGATGAGAAATTGCTTAATCGTTAA

SEQ ID NO: 78 amino acid sequence comprising GAS 58

MKWSGFMKTKSKRFLNLATLCLALLGTTLLMAHPVQAEVISKRDYMTFRGLGDLEDDSANYPNLEARYKG
 YLEGYEKGLKGGDIIPERPKIQVPEDVQPSDHGDYRDGYEEGFGEGQHKRDPLETEAEDDSQGGRRQGRQH
 QEGADSSDLNVEESDGLSVIDEVVGVIYQAFSTIWTYLSGLF

SEQ ID NO: 79 polynucleotide sequence encoding GAS 58

ATGAAATGGAGTGGTTTTATGAAAACAAAATCAAAACGCTTTTAAACCTAGCAACCCTTTGCTTGGCCCT
 ACTAGGAACAACCTTTGCTAATGGCACATCCCGTACAGGCGGAGGTGATATCAAAAAGAGACTATATGACTC
 GCTTCGGGTTAGGCGATTAGAAGATGATTTCAGCTAAGCTATCCTTCAAATTTAGAAGCTAGATATAAAGGA
 TATTTAGAGGGATATGAAAAAGGCTTAAAAGGAGATGATATACCCGAACGGCCCAAGATTCAGGTTCTCTGA
 GGATGTTACGCCATCTGACCATGGCGACTATAGAGATGGTTATGAGGAAGGATTTCGAGAAGGACAACATA
 AACGTGATCCATTAGAAACAGAACGAGAAGATGATTCTCAAGGAGGACGTCAGAAGGACGTCAGGACAT
 CAAGAAGGAGCAGATTCTAGTGATTGAACGTTGAAGAAAGCGACGGTTTGTCTGTTATTTGATGAAGTAGT
 TGGAGTAATTTATCAAGCATTAGTACTATTTGGACATACTTAAGCGGTTTGTCTAA

SEQ ID NO: 80 amino acid sequence comprising N-terminal leader sequence of GAS 58
 MKWSGFMKTKSKRFLNLATLCLALLGTTLLMA

SEQ ID NO: 81 amino acid sequence comprising a fragment of GAS 58 where the N-terminal leader sequence is removed

HPVQAEVISKRDYMTFRFLGLDLEDDSANYPNLEARYKGYLEGYEKGLKGGDIPERPKIQVPEDVQPSDHG
 DYRDGYEEGFGEQHKRDPLETEAEDDSQGGRRQEGRQGHQEGADSSDLNVEESDGLSVIDEVVGVIYQAFS
 TIWTFYLSGLF

SEQ ID NO: 82 amino acid sequence comprising GAS 290

MKHILFTVGSLSREGSFNHQLAAQAQKALEHQAVVSYNLWVKDVPVLNQDIEANAPLPVVDARQAVQSADAIW
 IFTPVYNFSIPGSVKNLLDWLSRALDLSDPTGSPAIGGKVTVSSVANGGHDQVFDQFKALLPFIRTSVAG
 EFTKATVNPDAWGTGRLEISKETKANLLSQAEALLAAI

SEQ ID NO: 83 polynucleotide sequence encoding GAS 290

ATGAAACATATTTTATTATTATTGTTGGCTCGCTTCGTGAAGGGTCTTTTAACCATCAATTAGCGGCTCAAGC
 AAAAAAGCTCTGGAACATCAAGCAGTTGTATCTTACTTAAATTGGAAAGACGTTCCCTGTTTGAATCAAG
 ATATCGAAGCTAATGCACCTTTACCAGTTGTTGACGCTCGTCAAGCTGTTTCAGTCAGCGGATGCTATCTGG
 ATTTTACACCAAGTTTACAACCTCTCTATTCCAGGTTCTGTGTTAAAAAOCCTGCTAGACTGGTTGTCTCGTGC
 TCTTGATTGTCTGATCCGACGGGCCCCATCTGCTATTGGCGGTAAGGTGGTTACGGTCTCTTCAGTTGCAA
 ATGGCGGGCATGATCAAGTATTTGATCAGTTTAAAGCACTATTGCCGTTTATCCGAACCTTCAGTAGCAGGA
 GAGTTTACAAAAGCAACTGTGAATCCTGATGCCGCGGGAACAGGAAGGCTTGAGATTTCAAAAGAGACAAA
 AGCAAACCTTGCTATCTCAGGCAGAGGCTCTTTTAGCGGCTATTATAG

SEQ ID NO: 84 amino acid sequence comprising GAS 511

MTDVSRIKLEARDQGRLTTLIDYANLIFDDFMELHGDHRHFSDDGAIVGGLAYLAGQPVTVIGIQKGNLQDN
 LARNFGQPNPEGYRKALRLMKQAEKFRPVVTFINTAGAYPCVGASERGGGEAIKLNLMEMSDLKVPILAI
 IIEGGSSGALALAVADQVWMLENTMYAVLSPEGFASILWKDGRATEAAELMKITAGELYKMGIVDRIIP
 EHGYSFSEIVDIIKANLIEQITSLQAKPLDQLLDERYQRFKRY

SEQ ID NO: 85 polynucleotide sequence encoding GAS 511

ATGACAGATGTATCAAGAATTTTAAAGAAGCGCGTGATCAAGGGCGTTTAAACAACCTTTGGATTACGCCAA
 CCTTATTTTCGATGACTTTATGGAACATGATGGCGATCGCCATTTTTCAGATGATGGTGCCATTGTAGGTG
 GCCTAGCTTATTTGGCGGGACAACCTGTTACGGTCATTGGTATTCAAAAAGGTAAGAATTTACAGGATAAT
 TTGGCAGGAATTTTGGCCAGCCCATCCAGAAGGTATCGTAAAGCTTTGCGCCTTATGAAACAGGCAGA
 AAAATTTGGACGACCCAGTTGTTACGTTTATCAATACTGCAGGAGCCTATCCAGGTGTCGGTGCGGAAGAAC
 GAGGACAGGGTGAGGCCATTGCTAAAAATTTGATGGAATGAGTGATCTCAAGGTTCCCATTTATCGCCATC
 ATTATTTGGTGAAGGAGGCTCTGGTGGTGCAATTAGCCTTAGCGGTTGCCGATCAGGTCTGGATGCTTGAAAA
 TACTATGTATGCGGTTCTTAGCCCCAGAAGGCTTTGCTTCTATTTTATGGAAGGATGGTTCAAGGCGACCG
 AGGCCGCTGAATTGATGAAAATCACAGCGGGTGAACCTCTACAAAATGGGAATAGTAGACCGTATTATTCCA
 GAACATGGTTATTTTCAAGTGAAATCGTTGACATCATCAAGCTAACCTCATCGAACAAATAACCAGTTT
 GCAAGCTAAGCCATTAGACCAATTATTAGATGAGCGCTACCAACGCTTTCGTAAATATTAA

SEQ ID NO: 86 amino acid sequence comprising GAS 533

MAITVADIRREVKEKNVTFLRLMFTDIMGVMNVEIPATKEQLDKVLSNKVMFDGSSIEGTVRINESDMYL
YPDLDTWIVFPWGDENGAVAGLICDIYTAEGKPFAGDPRGNLKRALKHMEIIGYKSFNLGPEPEFFLTKMD
DKGNPTLEVNDNGGYFDLAPIDLADNTRREIVNLTGMGFVEASHHEVAVGQHEIDFKYADVLRKACDNIQ
IFKLIVKTIAREHGLYATFMAKPKFGIAGSGMHCNMSLFDNQGNNAFYDEADKRGMQLSedayyFLGGLMK
HAYNYTAITNPTVNSYKRLVPGYEAPVYVWAGSNRSPLIRVPASRGMGTRLELRSDPTANPYLALAVLL
EAGLDGIINKIEAPEPVEANIYTMTEERNEAGIIDLPSTLHNAKALQKDDVVQKALGYHIYTNFLEAKR
IEWSSYATFVSQWEIDHYTHNY

SEQ ID NO: 87 polynucleotide sequence encoding GAS 533

ATGGCAATAACAGTAGCTGACATTCGTCGTGAAGTCAAAGAAAAAATGTAACGTTTCTTCGCTTGATGTT
CACTGATATCATGGGCGTTATGAAAAATGTGGAGATTCTCGCAACTAAAGAACAGTTAGACAAAGTATTGT
CTAACAAGGTTATGTTTGATGGTTCATCTATCGAAGGTTTTGTACGGATCAATGAGTCAGATATGTACCTT
TACCCCGATTAGACACTTGGATTGTTTTTCCCTGGGGAGATGAAAATGGAGCAGTTGCAGCTTTAATTTG
TGATATTTATACAGCAGAAGGAAAGCCTTTTCAGGAGATCCTAGAGGAAATTTAAAAAGAGCCCTGAAAC
ACATGRACGAGATCGGCTACAAATCATTAACTCTTGGACCAGAACCAAGATTTTTCCTTTTAAAGATGGAT
GATAAAGGTAATCCGACACTTGAAGTTAACGATAATGGTGGTTATTTTGATTAGCGCCAATTGACTTAGC
AGACAACACCGCGCCGTGAAATTTGTGAATATTTTAACGAAAATGGGTTTTGAAGTGGAAGCTAGTCATCATG
AAGTGGCTGTGGTCAACATGAGATTGATTTTAAATATGCAGATGTTTTGAAAGCTTGTGATAATATTCAA
ATTTTTAAGCTAGTTGTAAAAACGATTGCCCGTGAACATGGACTTATGCTACTTTTCATGGCTAAACCAAA
ATTTGGAATAGCTGGATCAGGGATGCACTGTAAATGTCTTTGTTTGATAACCAAGGTAATAATGCTTTTT
ATGATGAAGCTGATAAGCGAGGGATGCAGTTATCAGAAGATGCTTATTATTTCTTGGGAGGACTAATGAAG
CATGCTTATAACTACACTGCTATCACTAACCTACAGTGAATTCCTATAAACGATTAGTTCAGGTTATGA

GGCACCTGTTTATGTCGCTTGGGCTGGAAGTAATCGTTACCGCTTATCCGTGTTCCAGCATCACGTGGTA
TGGGAACCGCTTGGAGTTACGTTGCGGTTGATCCGACAGCTAATCCTTATTTAGCCTTGGCTGTTCTCTTG
GAAGCTGGATTAGATGGTATCATTAACAAATTTGAAGCTCCAGAACCCGTTGAAGCTAACATTTATACCAT
GACAATGGAAGAACGAAATGAAGCAGGCATTATTTGATTTGCCATCAACGCTTCATAATGCCTTAAAAGCTC
TTCAAAAAGATGATGTGGTACAAAAGGCAC TAGTTTACCATATCTACACTAATTTCTTAGAAGCAAAACGA
ATTGAATGGTCTTCCTATGCAACTTTTGTTCCTCAATGGGAAATTGACCATTATATTCATAATTATTAG

SEQ ID NO: 88 amino acid sequence comprising GAS 527

MTEISILNDVQKIIVLDYGSQYNQLIARRIREFGVFSSELKSHKITAQELREINPIGIVLSGGPNNSVIADNA
FGIDPEIFELGIPILGICYGMQLITHKLGGKVVPAGQAGNREYQSTLHLRETSKLFSGTPQEQVLVMSHG
DAVTEIPEGFHLVGDSNDPCYAAIENTEKNLYGIQFHPFVRHSVYGNLILKNFAISICGARGDWSMDNFID
MBIAKIRETVGDRFVLLGLSGGVDSSVGVLLQKAIGDQLTCIFVDHGLLRKDEGDQVMGMLGGKFGLNII
RVDASKRFLDLLADVEDPEKKRKIIIGNEFVYVFDDEASKLKGVDPLAQGTLYTDIIESGTETAQTIRSHHN
VGGLPEDMQFELIEPLNTLFKDEVRLGLALGMPREIVWRQFPFGPLAIRVMGAITEEKLETVRESDAILL
REEIAKAGLDRDVWQYFTVNTGVRVSVGMGDGRTYDYTTIAIRAITSIDGMTADFAQLPVDVLKKISTRIVN
EVDHVNRIYVDITSKPPATVEWE

SEQ ID NO: 89 polynucleotide sequence encoding GAS 527

ATGACTGAAATTTCAATTTTGAATGATGTTCAAAAAATTATCGTTCTTGATTATGGTAGCCAGTACAATCA
 GCTTATTGCTAGACGTATTCGAGAGTTTGGTGTGTTTCTCCGAACTAAAAAGCCATAAAATCACCCGCTCAAG
 AACCTCGTGAGATCAATCCCATAGGTATCGTTTATCAGGAGGGCCTAACTCTGTTTACGCTGATAACGCC
 TTTGGCATTGACCCCTGAAATCTTTGAACTAGGGATTCCGATTCTTGGTATCTGTTACGGTATGCAATTAAT
 CACCCATAAATTAGGTGGTAAAGTTGTTCTCTGCTGGACAAGCTGGTAATCGTGATACGGTCAGTCAACCC
 TTCATCTTCGTGAACCGTCAAAATATTTTTCAGGCACACCTCAAGAACAACCTCGTTTGTATGAGCCATGGT
 GATGCTGTACTGAAATTCAGAAAGGTTTCCACCTTGTGGAGACTCAAAATGACTGTCCCTATGCAGCTAT
 TGAATAACTGAGAAAAACCTTTACGGTATTCACTTCCACCCAGAAGTGAGACACTCTGTTTATGGAATG
 ACATTCTTAAAAACCTTGCTATATCAATTTGTGGCGCGCGTGGTGATGGTCAATGGATAAATTTATTGAC
 ATGGAAATTCGTAAATTCGTGAAACTGTAGGCGATCGTAAAGTCTCTCTAGGTCTTTCTGGTGGAGTTGA
 TCTTTCAGTTGTGGTGTCTTACTTCAAAAAGCTATCGGTGACCAATTAACCTGTATTTTCGTGTGATCAG
 GTCTTCTCGTAAAGACGAGGCGCATCAAGTTATGGGAATGCTTGGGGCAAAATTTGGCCCTAAATATATATC
 CGTGTGGATGCTTCAAAACGTTCTTAGACCTTCTTGCAGACGTTGAAGATCCTGAGAAAAACGTAAAT
 TATTGGTAATGAATTTGCTCTATGTTTGTATGATGAAGCCAGCAAAATAAAGGTGTTGACTTCCTTGCCC
 AAGGAACACTTTTACTGATATCATTTGAGTCAGGAACAGAACTGCTCAAAACATCAAAATCACATCACAAT
 GTGGGTGGTCTCCCGAAGACATGCAGTTTGAATTGATTGAGCCCTTAAACACTCTTTTCAAAAGATGAAGT
 TCGAGCGCTTGGAAATCGCTCTTGGAAATGCTGAAGAAATTTGTTTGGCGCCAACCATTTCCAGGTCTGGAC
 TTGCTATCCGTGCTATGGGAGCAATTACTGAAGAAAACTTGAAACCGTTTCGCGAATCAGACCGCTATCCTT
 CGTGAAGAAATTGCTAAGGCTGGACTTGATCGTGACGTGTGGCAATACTTTACAGTTAACACAGGTGCTCG
 TTCTGTAGGCGCTCATGGGAGATGGTCTGACTTATGATTATACCATCGCCATTCTGTCTATTACGCTTATTG
 ATGGTATGACAGCTGACTTTGCTCAACTTTCCTTGGGATGCTTGAAGAAAAATCTCAACACGTATCGTAAAT
 GAAGTTGACCACTTAACCGTATCGTCTACGACATCAAGTAAACCACCCGCAACAGTTGAATGGGAATA
 A

SEQ ID NO: 90 amino acid sequence comprising GAS 294

MSQSTATYINVIGAGLAGSEAYQIAKRGI PVKLYEMRGVKATPQHKTNFPAELVCSNSFRGDSLTNAVGL
 LKEEMRRLDSIIMRNGEANKRVPAGGAMAVDREGYAESVTAELNHPLEIVIRGETEIPDDAITVIATGFL
 TSDALAEKIHALLNGGDGFYFYDAAAPIIDKSTIDMSKVYLKSRDKGEAAVLYNC PMTKKEEFMAFHEALTTA
 EEAPLNAFEKEKYFEGCMPFIEVMAKRGIKTMLYGPMKEVGLYFDDYTGPDRDGEPTPYAVVQLRQDNAAG
 SLYNIVGFQTHLKWGEQKRVFQMI PGLENAEFVR YGV MHRNSYMDSPNLLTETPQSRSNPNLFFAGQMTGV
 EGYVESAAAGSLVAGINAAFLFKREALIFPQTTAIGSLPHYVTHADSKHFQPMNVNFGI IKELEGPRIRDK
 KERYEALASRALADLDTCLASL

SEQ ID NO: 91 polynucleotide sequence encoding GAS 294

TTGTCTCAATCAACTGCAACTTATATTAATGTTATTGGAGCTGGGCTAGCTGGTTCTGAAGCTGCCATCA
 GATTGCTAAGCGCGGTATCCCGTTAAATTGTATGAAATGCGTGGTGTCAAAGCAACACCGCAACATAAAA
 CCACTAATTTTCCGAATTGGTCTGTTCCAACCTCATTTTCGTGCTGATAGCTTAACCAATGCAGTCGGTCTT
 CTCAAAGAGAAATGCGGCGATTAGACTCCATTATTATGCGTAATGGTGAAGCTAACCGCGTACCTGCTGG

GGGAGCAATGGCTGTTGACCGTGAGGGGTATGCAGAGAGTGTCACTGCAGAGTTGGAAAAATCATCCTCTCA
 TTGAGGTCATTCTGTGTTGAAATTACAGAAATCCCTGACGATGCTATCACGGTTATCCCGACGGGACCGCTG
 ACTTCGGATGCCCTGGCAGAAAAATTCACGCGCTAAATGGTGGCGACGGATTCTATTTTTACGATGCAGC
 AGCGCTTATCATTTGATAAATCTACCATTTGATATGAGCAAGGTTTACCTTAAATCTCGCTACGATAAAGGCG
 AAGCTGCTTACCTCAACTGCCCTATGACCAAGAAGAATTCATGGCTTCCATGAAGCTCTGACAACCGCA
 GAAGAAGCCCCGCTGAATGCCCTTTGAAAAAGAAAAGTATTTTGAAGGCTGTATGCCGATTGAAGTTATGGC
 TAAACGTGGCATTAAACCATGCTTTTATGGACCTATGAAACCCGTTGGATTGGAATATCCAGATGACTATA
 CAGGTCTCCGCGATGGAGAAATTTAAACGCCATATGCCGTCGTGCAATTGCGTCAAGATAATGCAGCTGGA
 AGCCTTTATAATATCGTTGGTTTCCAAACCCATCTCAAAATGGGGTGAGCAAAAACCGGTTTTCAAATGAT
 TCCAGGGCTTTGAAATGCTGAGTTTGTCCGCTACGGCGTCATGCATCGCAATTCTATATGGATTACACAA
 ATCTTTTAAACCGAAACCTTCCAAATCTCGGAGCAATCCAAACCTTTTCTTTGCAGGTCAGATGACTGGAGTY
 GAAGGTTATGTCGAATCAGCTGCTTCAGGTTTGTAGCAGGAATCAATGCTGCTCGTTTGTTCAAAAGAGA
 AGAAGCACTTATTTTCTCTCAGACAACAGCTATTGGGAGTTTGCCTCATTATGTGACTCATGCCGACAGTA
 AGCATTTCCAACCAATGAACGTCAACTTTGGCATCATCAAAGAGTTAGAAGGCCACGCAATTCTGTGACAAA
 AAAGAACGTTATGAAGCTATTTGCTAGTCTGCTTTGGCAGATTTAGACACCTGCTTAGCGTCCGCTTTAA

SEQ ID NO: 92 amino acid sequence comprising GAS 253

MPKKILFTGGGTVGHVTLNLILIPKFIKDGWEVHYIGDKNGIEHTEIEKSGLDVTFPHAIATGKLRRYFSWQ
 NLADVFKVALGLLQSLFIVAKLRPQALFSKGGFVSVPVVAAKLLGKPVFIHESDRSMGLANKIAYKFATT
 NYTTFEQEDQLSKVKHLGAVTKVFKDANQMPFESTQLEAVKEYFSRDLKTLFFIGGSAGAHVFNQFISDHPE
 LKQRYNIINITGDPHLNELSSHLRYVDYVTDLYQPLMAMADLVVTRGGSNTLFELLAMAKLHLIVPLCKEA
 SRGDQLENATYFEKRGYAKQLQEPDLTLHNFDQAMADLFEHQADYEATMLATKEIQSPDFFYDLLRADISS
 AIKEK

SEQ ID NO: 93 polynucleotide sequence encoding GAS 253

ATGCCTAAGAAGATTTTATTACAGGTGGTGGAACTGTAGGTCATGTCACCTTGAACCTCATTCATACACC
 AAAATTTATCAAGGACGGTTGGGAAGTACATTATATTGGTGATAAAAATGCCATTGAACATACAGAAATTG
 AAAAGTCAGGCCCTTGACGTGACCTTTTCATGCTATCGCGACAGGCAAGCTTAGACGCTATTTTTCATGGCAA
 AATCTAGCTGATGTTTTTAAGGTTGCACTTGGCCTCCTACAGTCTCTCTTTATTGTTGCCAAGCTTCGCCC
 TCAAGCCCTTTTTCCAAAGGTGGTTTGTCTCAGTACCGCCAGTTGTGGCTGCTAAATTGCTTGGTAAAC
 CAGTCTTTATTCATGAATCAGATCGGTCAATGGGACTAGCAAACAAGATTGCCATACAAATTTGCAACTACC
 ATGTATACCACTTTTGAGCAGGAAGACCAGTTGCTCTAAAGTTAAACACCTTGGAGCGGTGACAAAGGTTTT
 CAAAGATGCCAACCAATGCCTGAATCAACTCAGTTAGAGGCGGTGAAAGAGTATTTTAGTAGAGACCTAA
 AAACCTCTTGTATTATTGGTGGTTGGCAGGGGCGCATGCTTTAATCAGTTTATTAGTGATCATCCAGAA
 TTGAAGCAACGTTATAATATCATCAATATTACAGGAGACCTCACCTTAATGAATPGAGTTCTCATCTGTA
 TCGAGTAGATTATGTTACCGATCTCTACCAACCTTTGATGGCGATGGCTGACCTTGTAGTGACAAGAGGGG
 GCTCTAATACACTTTTGTAGCTACTGGCAATGGCTAAGCTACACCTCATCGTCTCTCTTGGTAAAGAAGCT
 AGCCGTGGCGATCAGTTAGAAAATGCCACTTATTTTGAGAAGAGGGGCTACGCTAAACAATTACAGGAACC
 TGATTTAACTTTGCATAATTTTGATCAGGCAATGGCTGATTTGTTTGAACATCAGGCTGATTATGAGGCTA
 CTATGTTGGCAACTAAGGAGATTCAGTCACCGGACTTCTTTATGACCTTTTGAGAGCTGATATTAGCTCC
 GCGATTAAGGAGAAGTAA

SEQ ID NO: 94 amino acid sequence comprising GAS 529

MCGIVGVGNRNATDILMQGLEKLEYRGYDSAGIFVANANQTNLIKSVGRIADLRKIGIDVAGSTGIGHT
 RWATHGQSTEDNAHRHTSQTRFVLVHNGVIENYLHIKTEFLAGHDFKQTDTEIAVHLIGKFVEEDKLSV
 LEAFKKLSIIIEGSYAFALMDSQATDTIYVAKNKSPLLIIGLGGYNMVCSDAMAMIRETSEFMEIHDKELV
 ILTKDKVTVDYDGKELIRDSYAELELSDIGKGYPFYMLKEIDEQPTVMRQLISTYADETGNVQVDPAI
 ITSIEADRLYILAAGTSYHAGFATKNMLEQLTDPVELGVASEWGYHMPLLSKKPMFILLSSQGETADSR
 QVLVKANAMGIPSLTVTNVFGSTLSREATYTMLIHAGPEIAVASTKAYTAQIAALAFKAVGEANGKQRA
 LDFNLVHELISLVAQSIETLSEKDLVAEKVQALLATTRNAPYIGRGNDYYVAMEAALKLKEISYIQCEGFA
 AGELKHGTISLIEEDTFVIALISSQLVASHTRGNIQEVAAARGAHVLTVVEEGLDREGDDIIVNKVHPFLA
 PIAMVIPTQLIAYYASLQRLDVKPRNLAKAVTVE

SEQ ID NO: 95 polynucleotide sequence encoding GAS 529

ATGTGTGGAATTGTTGGAGTTGTTGGAAATCGCAATGCAACGGATATTTTAAATGCAAGGCCTTGAAAAGCT
 TGAATACCGGGGTATGATTCAGCAGGAATTTTGTGGCTAATGCCAATCAACAAACTTGATTAAATCAG

TGGGGCGGATTGCTGATTTGCGTGCCAAAGATTGGCATTGATGTTGCTGGTTCAACAGGGATTGGTCACACC
 CGTTGGGCAACGCATGGCCAAATCAACAGAGGATAATGCCCATCCTCACACGTCACAACTGGACGTTTGT
 ACTTGTTTCATAATGGTGTGATTGAAAATTACCTTCACATTAAAAACAGAGTTCCCTAGCTGGACATGATTTTA
 AGGGGCAGACAGATACCTGAGATGTCAGTACACTTGATTGGAAAAATTTGTGGAAGAGACAAGTTGTCAGTA
 CTGGAAGCTTTTAAAAAATCTTTAAGCATTATTGAAGGTTCTACGCCCTTTGCATTAAATGGATAGCCAAGC
 AACTGATACTATTTATGTGGCTAAAAACAAGTCTCCATTGTTGATTGGACTTGGTGAAGGTTACAACATGG
 TTTGTTTCAGATGCCATGGCCATGATTCGTGAAACCAGTGAATTTATGGAAATTCATGATAAGGAGCTAGTT
 ATTTTAACCAAGATAAGGTAAGTGTACAGACTACGATGGTAAAGAGCTGATACGAGATTCCCTACACTGC
 TGAATTAGACTTATCTGATATTGGCAAAGGGAATTATCCCTTCTATATGCTGAAAGAAATTTGATGAGCAAC
 CAACCGTAATGCGTCAATTAATTTCAACTTATGCAGATGAAACTGGTAAAGTACAGGTTGATCCGGCTATC
 ATTACCTCTATCCAGAGGGCTGACCGTCTTTATATTTTAGCGGCAGGGACTTCCCTACCATGCTGGTTTTGC
 AACAAAAATATGCTTGAGCAATTGACAGATACACCAGTTGAGTTGGGCGTGGCTTCTGAGTGGGGTTACC
 ACATGCTCTGCTTAGCAAGAAACCAATGTTTATTCTACTAAGCCAATCAGGAGAAACCGCAGATAGTCGT
 CAAGTTTGTAGTAAAGGCAAAATGCTATGGGCATTCCGAGTTTGACAGTAACCTAACGTTCCAGGATCAACCTT
 ATCAGCTGAAGCAACATACACCATGTTGATTCATGCTGGACCTGAAATGCTGTTGCGTCTACAAAAGCTT
 ACACGTCACAAATGCTGCCCTTGCCCTTTTGGCTAAGGCAGTTGGTGAGGCAAAATGGTAAGCAAGAAGCT
 CTTGACTTTAACTTGGTACATGAGTTGTCTATGGTTGCCCAATCTATTGAGGCGACTTTGTCTGAAAAAGA
 TCTCGTGGCAGAAAGGTTCAAGCTTTGCTAGCTACTACTCGTAATGCTTTTACATCGGGCGTGGCAATG
 ATTATACGTTGCGATGGAAGCTGCTTTGAAATTAAGAGAGATTTCATTATTAATGCGAAGGCTTTGCG
 GCTGGTGAATTTGAACATGGAACCATTTTCATTAATTGAGGAGGACACGCCAGTAATCGCTTTAATATCGTC
 TAGTCAGTTGGTTGCCCTCTCATACGCGTGGTAATATTCAAGAAGTTGCTGCCCGTGGGGCTCATGTTTAA
 CAGTTGTGGAAGAAGGGCTTGACCGTGAGGAGATGACATTATTGTCAATAAGGTTCACTCCTTCCCTAGCC
 CCGATTGCTATGGTCATTCCAACTCAACTGATTGCTTACTACGCTTCATTACAACCTGGACTTGATGTTGA
 TAAGCCACGTAATTTGGCTAAAGCTGTAACAGTAGAATAA

SEQ ID NO: 96 amino acid sequence comprising GAS 45

VTFMKKSEKWLAAVSVAILSVSALAACGNKNASGGSEATKTYKYVFNPKSLDYILTNNGGTTDVTQMDV
 GLEENDEYGNLVPFLAKDKVSKDGLTYTYTLRDGVSWYTAGDEEYAPVTAEDFVTGLKHAVIDDKSDALYV
 VEDSIKLNKAYQNGEVDKFEVGVKALDDKTVQYTLNKPESYWNSTTYSVLFPVNAKFLKSKGKDFGTTDP
 SSILVNGAYFLSAFTSKSSMEPHKNENYWDKNGVIESVKLTYSDDGSDPGSFYKNFDKGEFSVARLYPNDF
 TYKSAKNYADNITTYGMLTGDIRHLTNLNRHSFKNTKKDPAQQDAGKKALNNKDFRQAIQFAPDKASFQA
 QTAGQDAKTKALRNMLVPPFTVTIGESDFGSEVEKEMAKLGDEWKDVNLADAQDGFYNPEKAKAEFAKAKE
 ALTAEGYTFPVQLDYPVDQANAATVQEAQSFQSVESLKGKENVIVNLETETSTHEAQGFYATPEQQDY
 DIISSWNGPDYQDPRTYLDIMSPVGGSVIQKLGIKAGQNKDVVAAAGLDYQTLLEDEAAITDDNDARYK
 AYAKAQAYLTDNAVDIPVVALGGTFRVTKAVPFSGGPSWAGSKGPLAYKGMKLQDKPVTVRQYKAKAKEMM
 KAKAKSNKVAEKLADIVERK

SEQ ID NO: 97 polynucleotide sequence encoding GAS 45

GTGACTTTTATGAAGAAAAGTAAATGGTTGGCAGCTGTAAGTGTTCGATCTTGTCTAGTATCCGCTTTGGC
 AGCTTGTGGTAATAAAAAATGCTTCAGGTGGCTCAGAAGCTACAAAACTTACAAGTACGTTTGTGTTAACG
 ATCCAAAATCATTGGATTATATTTTGACTAATGGCGGTGGAACGACTGATGTGATAACACAAATGGTTGAT
 GGTCTTTTGGAAAACGATGAGTATGGTAATTTAGTACCATCACTTGTCTAAGATTGGAAGGTTTCAAAGA
 CGGTCTGACTTATACTTATACTCTTCGCGATGGTGTCTCTTGGTATACGCTGATGGTGAAGAAATATGCC
 CAGTAACAGCAGAAAGATTTTGTGACTGGTTTGAAGCACCGGTTGACGATAAATCAGATGCTCTTTACGTT
 GTTGAAGATTCAATAAAAACTTAAAGGCTTACCAAAATGGTGAAGTAGATTTTAAAGAAGTTGGTGTCAA
 AGCCCTTGACGATAAACTGTTTCACTATACCTTGAACAAGCCTGAAAGCTACTGGAATTCAAAAACAAT
 ATAGTGTGCTTTTCCAGTTAATGCCAAATTTTGAAGTCAAAAGGTAAAGATTTTGGTACAACCGATCCA
 TCATCAATCCTTGTAAATGGTGTACTTCTTGAGCGCTTCACCTCAAAATCATCTATGGAATTCATATA
 AAATGAAAACCTACTGGGATGCTAAGAATGTTGGGATAGAATCTGTTAAATTGACTTACTCAGATGGTTTCA
 ACCAGGTTTCTTACAAGAACTTTGACAAGGGTGAGTTTCAAGCTTGACGACTTTACCCAAATGACCTT
 ACCTACAAATCAGCTAAGAAAACTATGCTGATAACATTAATACGGAATGTTGACTGGAGATATCCGCTCA
 TTTAATCATGGAATTTGAACCGTACTTCTTTCAAAAACCTAAGAAAGACCTTGCACAAACAGATGCCCGTA
 AGAAAGCTCTTAAACAACAAGGATTTTCTCAAGCTATTCAAGTTTGTCTTTGACCGAGCGTCATTCCAAGCA
 CAAACTGCAGGTCAGATGCCAAAACAAAAGCCTTACGTAACATGCTTGTCCCAACATTTGTGACCAT
 TGGAGAAAGTGATTTTGGTTTCAGAAAGTTGAAAAGGAAATGGCAAACTTGGTGTGAATGGAAAGACGTTA
 ACTTAGCTGATGCTCAAGATGGTTTCTATAATCTTCAAAAAGCAAAAGCTGAGTTGCAAAAGCCAAAGAA

GCTTTAACAGCTGAAGGTGTAACCTTCCAGTTCAATTAGATTACCCGTGTGACCAAGCAAACGCAGCAAC
 TGTTTCAGGAAGCCCAGTCTTTCAAACAATCTGTTGAAGCATCTCTTGGTAAAAGAGAATGTCATTGTCAATG
 TTCTTGAAACAGAAACATCAACTCAGGAAGCCCCAAGGCTTCTATGCTGAGACCCAGAACAAAGACTAC
 GATATCATTTTCATCATGGTGGGGACCAGACTATCAAGATCCACGGACCTACCTTGACATCATGAGTCCAGT
 AGGTGGTGGATCTGTTATCCAAAACTTGAATCAAAGCAGGTCAAATAAGGATGTTGTGGCAGCTGCAG
 GCCTTGATACCTACCAAACTCTTCTTGATGAAGCAGCAGCAATTACAGACGACAACGATGCCCGCTATAAA
 GCTTACGCAAAAGCAAGCCTACCTTACAGATAATGCCGTAGATATTCAGTTGTGGCATTTGGGTGGCAC
 TCCACGASTTACTAAAGCCGTTCCATTTAGCGGGGGCTTCTCTTGGGCAGGGTCTAAAGGTCTCTAGCAT
 ATAAAGGAATGAACTTCAAGACAAACCTGTACAGTAAACAATAACGAAAAAGCAAAAGAAAAATGGATG
 AAAGCAAAGGCTAAGTCAATGCAAAATATGCTGAGAAGTTAGCTGATCACGTTGAAAAA

SEQ ID NO: 98 amino acid sequence comprising an N-terminal leader sequence of GAS 45
 VTFMKKSKWLAAVSVAILSVSALAA

SEQ ID NO: 99 amino acid sequence comprising a fragment of GAS 45 where the N-terminal leader sequence is removed

CGKNKASGGSEATKYKYVFNDDPKSLDYILTNGGGTTDVITQMVDGLENDEYGNLVPFLAKDWKVS KDG
 LTYTYTLRDGVSWYTADGEEYAPVTAEDFVTGLKHAVDKSDALYVVEDSIKNLKAYQNGEVDFKEVGKVA
 LDDKTQYITLTKPESYWNSTTYSVLPVNAKFLKSKGKDFGTTDPSSILVNGAYFLSAFTSKSSMEFHKN
 ENYWDANKVGIESTVLTYSDGSDPGSPYKFNDFKGEFSVARLYPNDPYKSAKKNYADNITYGMLTGDIRHL
 TWNLNRTSFKNTKKDPAQQDAGKKALNNKDFRQAIQFAFDRASFOAQTAGQDAKTALRNMLVPPTFVTIG
 ESDPGSEVEKEMAKLGDEWKDVNLADAQDGFYNPEKAKAEFAKAKEALTAEGVTFPVQLDYPVDQANAATV
 QEAQSPKQSVESLKGENVIVNLETETSTHEAQGFYARTPEQQDYDIISWVGPDYQDPRTYLDIMSPVG
 GGSVIQKLGKAGQNKDVVAAAGLDYQTLLEDAATDDNDARYKAYAKAQAYLTDNAVDIFVVALGGTP
 RVTKAVPFGSGFSWAGSKGPLAYKGMKLQDKPVTVKQYKAKKWKAKAKSNKYAEKLADHVEK

SEQ ID NO: 100 amino acid sequence comprising GAS 95

MKIGKKIVLMFTAIVLTTLVLAALGVLTSAITFTSTGELSKTFKDFSTSSNKSDAIKQTRAFSILLMGVDTGS
 SERASKEGNSDSMILVTVPKTKRTTMTSLERDTLTLSQPKNNEMNGVEAKLNAAYAAGGAQMAIMTVQ
 DLLNITIDNIVQINMQGLIDLNVAVGGITVTNEFDFFISIAENEPEYQATVAPGTHKINGEQALVYARMRY
 DDPEGDYGRQKRQREVIQKVLKKILALDSISSYKILSAVSSNMQTNIEISSRTIPSLLYRDLALRTIKTY
 QLKGEDATLSDGGSYQIVTSNHLLEIQNRIRTELGLHKVNQLKTNATVYENLYGSTKSQTVNNYDSSGQA
 PSYSDSHSSYANYSSGVDTGQSASTDQDSTASSHRPATPSSSSDALADESSSSSGSLVPPANINPQT

SEQ ID NO: 101 polynucleotide sequence encoding GAS 95

ATGAAAAATTGAAAAAAATAGTTTAAATGTTACACAGCTATTGTGTTAACTGCTTTGGCATTAGGTGT
 CTATCTAACTAGTGTCTTATACCTTCTCAACAGGAGAATTATCAAGACCTTTAAAGATTTTCGACATCTT
 CAACAAAAGTGATGCCATTAAACAAACAAGAGCTTTTCTATCTTGTGATGGGTGTTGATACAGGCTCT
 TCAGAGCGTGCCCTCCAAGTGGGAAGGAAACAGTGATTCGATGATTTTGGTTACGGTTAAATCCAAAGACCAA
 GAAACAACTATGACTAGTTTGAACAGAGATACCTTAACACGTTATCTGGACCCAAAAATAATGAAATGA
 ATGGTGTGAAGCTAAGCTTAACGCTGCTTATCCAGCAGGTGGCGCTCAGATGGCTATTATGACCGTGCAA
 GATCTTTTGAATATCACCATTGATAACTATGTTCAAATTAATATGCAAGGCCTTATTGATCTTGTGAATGC
 AGTTGGAGGGATTACAGTTACAAATGAGTTTGATTTTCTATCTCGATTGCTGAAAACGAACCTGAATATC
 AAGCTACTGTTGCGCCTGGAACACACAAAATTAACGCTGAACAAGCTTTGGTTTATGCTCGTATGCGTTAT
 GATGATCCTGAGGGAGATTATGGTCGACAAAAGCTCAACGCTGAAGTCATTCAAAAGGTATTGAAAAAAAT
 CCTTGTCTCTTGATAGCATTAGCTCTTATCGGAAGATTTTATCTGCTGTAAGTAGTAATATGCAAAACGAATA
 TCGAAATCTCTTCTCGCACTATCCCTAGTCTATTAGGTTATCGTGACGCACTTAGAATATTAAAGACTTAT
 CAACATAAAGGAGAGATGCCACTTTATCAGATGGTGGATCATACCAAATTGTTACCTTAATCATTGTGTT
 AGAAATCCAAATCGTATCCGACAGAATTAGGACTTCATAAGGTAAATCAATTAATAACAAATGCTACTG
 TTTATGAAAATTTGTATGGGTCAACTAAGTCTCAGACAGTAACAACAACATATGACTCTTCAGGCCAGGCT
 CCATCTTATTTGATAGTCATAGCTCTTACGCTAATTAATCAAGTGGAGTAGATACCGCCAGAGTGCTAG
 TACAGACCAGGACTCTACTGCTTCAAGCCATAGGCCAGCTACGCCGCTTCTTTCATCAGATGCTTTAGCAG
 CTGATGAGTCTAGCTCATCAGGGTCTGGATCATTAGTTCCTCTCTGCTAATATCAACCCCTCAGACCTAA

SEQ ID NO: 102 amino acid sequence comprising N-terminal leader sequence of GAS 95
MKIGKKIVLMFTAIVLTTVLALGVYLTSAITFS

SEQ ID NO: 103 amino acid sequence comprising a fragment of GAS 95 where the N-terminal leader sequence is removed.

TGELSKTFKDFSTSSNKSDAIKQTRAFSILLMGVDTGSSERASKWEGNSDSMILVTVPKTKKTMTSLER
DTLFTLSGPKNNEMNGVEAKLNAAYAAGGAQMAIMTVQDLLNITIDNYVQINMQGLIDLNVAVGGITVTNE
FDFPLISIAENEPEYQATVAPGTHKINGEQALVYARMRYDDPEGDYGRQKQREVIQKVLKKILALDSISSY
RKILSAVSSNMQTNIBTSSRTIPSLGYRDALRTIKTYQLKGEDATLSDGGSVQIVTSNHLLEIQNRIRTE
LGLHKVYQLKTNATVYENLYGSTKSQTVNNYDSSGQAPSYSDSHSSYANYSSGVDTGQSASTDQDSTASS
HRPATPSSSSDALAADESSSSSGSLVPPANINPQT

SEQ ID NO: 104 amino acid sequence comprising GAS 193

MKKRKILLAVTLLSTILLNSAVPLVVADTSLRNSTSSSTDOPTTADTDTDESETPKKDKKSKETASQHDYTK
DHKPSHTHTPTSPSNDTKQTDQASSEATDKPNKDKNDTKQPDSSDQSTPSFKDQSSQKESQNKDGRPTPSFD
QQKDQTPDKTPEKSADKTEKGPKEATDKTPEPNRDAPKPIQPPPLAAAPVFIWRESDEKDLKSLKPSRSS
AAYVVRHWGDSAYTHNLLSRRYGITABQLDGLNSLGIHYDKERLNGKRLLEWEKLTGLDVRAITVAIAME
SSLGTQGVAKKEGANNMFGYGAFFDFNPNNAKKYSDEVAIRHNVEDTIIANKNQTFERQDLKAKKWSLQGLDT
LIDGGVYFTDTSGSGQRRADIMTKLDQWIDDHGSTPEIPEHLKITSGTQFSEVPVGYKRSQPNVLTYSKSE
TYSFGQCTWYAYNRVKELGYQVDRYMNGGDNQKPGFVTTHKPKVGVVVSFAPGQAGADATYGHVAVVEQ
IKEDGSILISESNVMGLGTISYRTTAEQASLLTVVVGDKLRP

SEQ ID NO: 105 polynucleotide sequence encoding GAS 193

ATGAAGAAAAGGAAATTTGTTAGCAGTAACACTATTAAGTACCATACTCTTAAACAGTGCAGTGCCATTAGT
TGTTGCTGATACCTCCTTGCGTAATAGCACATCATCCACTGATCAGCCTACTACAGCAGATAGTATGATACGG
ATGACGAGAGTGAACACCAAAAAAGACAAAAAAGCAAGGAAACAGCGTCGACGACGACACCCAAAAA
GACCATAAGCCATCACACACTCACCAACCCCCCTTCAAATGATACTAAGCAGACCGATCAGGCATCATC
TGAAGCTACTGACAAACCAATAAAGACAACAAACGACACCAAGCAACGACAGCAGTGAATCAATCCACCC
CATCTCCCAAGACAGTCTGCTCTCAAAAAGAGTCACAAACAAAGACGCGCGACCTACCCCATCACTGAT
CAGCAAAAAGATCAGACACCTGATAAAACACCAAGAAAAATCAGCTGATAAAACCCCTGAAAAAGGACCA
AAAAGCAACTGATAAAACACCAAGACCAAAATCGTGACGCTCCAAACCCATCCAAACCTCTTTAGCAGCTG
CTCCTGTCTTTATACCTTGGAGAGAAAGTGAACAAAGACCTGAGCAAGCTAAACCAAGCAGTCTGCTCATCA
GCGGCTTACGTCAGACACTGGACAGGTGACTCTGCCTACACTCACAACTGTTGTACGCGCGTTATGGGAT
TACTGCTGAACAGCTAGATGGTTTTTTGAACAGTCTAGGTATTCACTATGATAAAGAACGCTTAAACGGAA
AGCGTTTATTAGAAATGGCAAAACCTAACAGGACTAGACGTTTCGAGCTATCGTAGCTATTGCAATGGCAGAA
AGCTCACTAGGTACTCAGGGAGTTGCTAAAGAAAAAGGAGCCAAATATGTTTGGTTATGGCGCCCTTTGACTT
CAACCCAAACAATGCCAAAAAATACAGCGATGAGGTTGCTATTCGTCACATGGTAGAAGACACCATCATTG
CCAACAAAAACCAACCTTTGAAAGACAAGACCTCAAAGCAAAAAAATGGTCACTAGGCCAGTTGGATACC
TTGATTGATGGTGGGTTTACTTTACAGATACAAGTGGCAGTGGGCAAGACGAGCAGATATCATGACCAA
ACTAGACCAATGGATAGATGATCATGGAAGCACACCTGAGATTCAGAACATCTCAGATAACTTCCGGGA
CACAAATTAGCGAAGTGCCCGTAGGTTATAAAAGAAGTCAGCCACAAACGTTTTGACCTACAAGTCAGAG
ACCTACAGCTTTTGGCCAATGCACCTTGGTACGCCCTATAATCGTGTCAAAGAGCTAGGTTATCAAGTCGACAG
GTACATGGGTAAACGGTGGCGACTGGCAGCGCAAGCCAGGTTTGTGACCACCCATAAACCTAAAGTGGGCT
ATGTCGCTCATTTGCAACAGGCCAAGCAGGAGCAGATGCAACCTATGGTCACGTTGCTGTTGTAGAGCAA
ATCAAAGAAGATGGTTCTATCTTAATTTCAAGAGTCAAATGTTATGGGACTAGGCACCATTTCCCTATCGGAC
GTTACAGCTGAGCAGGCTAGTTTGTGACCTATGTCGTAGGGGACAACTCCCAAGACCATAA

SEQ ID NO: 106 amino acid sequence comprising GAS 137

MSDKHINLVIVTGMGAGKTVAIQSFEDLGYFTIDNMPALVPKFLELIEQTNENRRVALVVDMSRLFFK
EINSTLDSIESNPSIDFRILFLDATDGLVSRKETRSHFLAADGRVLDGIRLERELSLPLKMSQHVVD
TTKLTPQRLKTTISDQFSEGSNQASFRIEVMFSGFKYGLFLDADLVFDVRFLPNFYQVELREKTGLDEDV
FNYVMSPHESEVFYKHLNMLIVPILPAYQKEGKSVLTVAIGCTGGQHRSVAFHCLAESLATDWSVNESHR
DQNRKETVNR

SEQ ID NO: 107 polynucleotide sequence encoding GAS 137

ATGTCAGACAAACACATTAATTTAGTTATTTGTGACAGGAATGAGCGGCGCTGGAAAAACAGTTGCCATTCA
GTCTTTTGGAGGATCTAGGCTACTTTTACCATTTGATAATATGCCCCAGCCTTGGTTCCAAAAATTTTGAAT
TAATTGAACAAACCAATGAAATCGTAGGCTGGCTTTGGTTGTGATATGAGAAGTCGTTTGTTCATCAAG

GAAATTAATTTCTACCTTAGATAGTATTTGAAAGCAATCCTAGCATTGATTTTCGGATTCTTTTTTTGGATGC
AACGGATGAGAAATTGGTGTACGCTATAAAGAAACCAGACGGAGCCACCCTTTGGCTGCGGACGGTCGTG
TGCTTGATGGTATTCGATTGGAAAGAGAACTCCTATCTCCTTTGAAAAGCATGAGCCAACATGTGGTGGAT
ACAACAAAATTGACCCCTAGACAATTGCGTAAACCATTTCAGACCAGTTTCTGAAGGGTCTAATCAAGC
CTCTTTCCCTATTGAAGTGATGAGCTTTGGGTTCAAATATGGTCTTCCTTTGGATGCGGATTGGTTTTTG
ATGTGCTTTTCTACCCAATCCTTATTATCAGGTAGAGCTTCGTGAAAAAACAGGACTAGATGAGGACGTT
TTTAATTATGTGATGTCTACCCAGAATCAGAGGTGTTTTACAAGCATTGTGTTAAACCTTATGTCCCTAT
CTTACCGGCTTACCAAAAAGAAGGGAAGTCTGTCTTGACGGTGGCTATTGGCTGCACAGGAGGCCAACACC
GCAGCGTTCCTTTGCCCATTTGCTTGSCAGAAAGTCTGGCAACAGATTGGTCGGTTAATGAAAGCCATCGT
GATCAAAATCGTCTGTAAGGAAACGGTGAATCGTTCATGA

SEQ ID NO: 108 amino acid sequence comprising GAS 84

MIKKRTVAILAIASSFFLVACQATKSLKSGDAWGVYQKQKSITVGFNDTFVPMGYKDESGRCKGFDIDLAK
KEVFHQYGLKVNFPQAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAFTDSYMRNEQIIVVKRSDIKTI
SDMKHKVLGAQSASSGYDSLRLTPKLLKDFIKNKDANQYETFTQAFIDLKSDRIDGILIDKVYANYLLAKE
GQLENYRMIPTTFENEAFSVGLRKEDKTLQAKINRAFRVLYQNGKFPQALSEKWFQDDVATANIKS

SEQ ID NO: 109 polynucleotide sequence encoding GAS 84

ATGATTATAAAAAAAGAACCGTAGCAATTTTAGCCATAGCTAGTAGCTTTTCTTGGTAGCTTGTCAAGC
TACTAAAGTCTTAAATCAGGAGATGCTTGGGGAGTTTACCAAAGCAAAAAGTATTACAGTTGGTTTTG
ACAATACGTTTGTTCCTATGGGCTATAAGGATGAAAGCGGCAGATGCAAAGGTTTGTATATGATTGGCT
AAAGAAGTTTTCACCAATATGGACTCAAGGTTAACTTTCAAGCTATTAATGGGACATGAAAGAAGCAGA
ACTAAACAATGGTAAAAATGATGTAATCTGGAATCGTTATTCAATAACTAAGGAGCGTCAGGATAAGGTTG
CCTTTACTGATTCTTACATGAGAAATGAACAAATTATTGTTGTCAAAAAAGATCTGATATTAACAATA
TCAGATATGAACATAAAGTGTAGGAGCACAAATCAGCTTCATCAGGTTATGACTCCTTGTTAAGAACTCC
TAAACTGCTGAAAGATTTTATTAAAAATAAAGACCGCTAATCAATATGAACCTTTACACAGCTTTTATTG
ATTTAAATCAGATCGTATCGATGGAATATTGATTGACAAAGTATATGCCAATTAATTTAGCAAAAGAA
GGGCAATTAGAGAATTATCGGATGATCCCAACGACCTTTGAAATGAAGCATTTTCGGTTGGACTTAGAAA
AGAAGACAAAACGTTGCAAGCAAAATTAATCGTGCTTTTCAGGGTGCTTTATCAAAATGGCAATTTCAAG
CTATTTCTGAGAAATGTTTGGAGATGATGTTGCCACTGCCAATATTAAATCTTAA

SEQ ID NO: 110 amino acid sequence comprising N-terminal leader sequence of GAS 84

MIKKRTVAILAIASSFFLVA

SEQ ID NO: 111 amino acid sequence comprising a fragment of GAS 84 where the N-terminal leader sequence is removed

CQATKSLKSGDAWGVYQKQKSITVGFNDTFVPMGYKDESGRCKGFDIDLAKKEVFHQYGLKVNFPQAINWDMK
EALNNGKIDVIWNGYSITKERQDKVAFTDSYMRNEQIIVVKRSDIKTISDMKHKVLGAQSASSGYDSLRL
TPKLLKDFIKNKDANQYETFTQAFIDLKSDRIDGILIDKVYANYLLAKEGQLENYRMIPTTFENEAFSVG
LRKEDKTLQAKINRAFRVLYQNGKFPQALSEKWFQDDVATANIKS

SEQ ID NO: 112 amino acid sequence comprising GAS 384

MKTLAFDTSNKTLSLAILDDETLADMTLNIQKKHSVSLMFAIDFLMTCTDLKPDLERIVVAKGPGSYTG
LRVAVATAKTLAYSLNIALVGISSLYALAASTCKQYPNTLVVPLIDARRQNAYVGYRQGKSVMPQAHASL
EVIIEQLVEEGQLIFVGETAPFAEKIQKKLPQAILLPTLPSAYECGLLGQSLAPENVDAFVPQYLKRVEAE
ENWLKDNEIKDDSHYVKRI

SEQ ID NO: 113 polynucleotide sequence encoding GAS 384

ATGAAGACACTTGCATTTGATACCTCAAATAAAACCTTGTCCCTTGCTATACTTGATGATGAGACACTTCT
AGCAGATATGACCCCTTAACATTCAGAAAAACATAGTGTAGCCCTTATGCCCTGCTATTGATTTTTCATGA
CTTGTACTGATCTTAAACCTCAAGATTTAGAAAGAATAGTGGTTGCAAAAGGCCCTGGATCTTACACAGGT
TTACGAGTGGCAGTTGCTACTGCAAAAACGTTAGCGTACAGTTTAAATATTGCATTGGTCCGGGATTTCGAG
TCTATATGCTTTGGCTGCGTCTACTTGTAAACAGTATCCAAATACTTTGGTGGTGGCCATTGATTGATGCTA
GAAGGCCAAATGCCTATGTAGGTATTATTCGGCAAGGAAAATCAGTGATGCCACAAGCCCATGCTTCACTA
GAAGTTATTATAGAACAATTAGTAGAAGAAGGACAGCTGATTTTGTGGGGAGACTGCTCCTTTTCGTGA

GAAAATTCAAAGAAACTACCTCAGGCGATACTACTTCCAACCCTTCCTTCTGCTTACGAATGTGGTCTTT
TGGGGCAAAGTTTGGCACCAGAAAATGTAGACGCCCTTTGTCCCTCAATATCTCAAGAGAGTGGAGCTGAA
GAAAACCTGGCTCAAAGATAATGAGATAAAAGATGATAGTCACTACGTTAAGCGAATCTAA

SEQ ID NO: 114 amino acid sequence comprising GAS 202

MLKRLMLILGPLLIAPVLVVITIFSFPTQLDHSIAQEKANAVAITDSSFNGLIKRQALSDETCRFVPPFG
SSEWSRMDSMHPSVLAERYKRSYRPFLLGKRGSAASLSHYYGIIQOITNEMQKKKAFVVSPOWFTAQGINPS
AVQMYLSNTQVIEPLKARTDKESQFAAKRLLELNPVSKSNLLKKVSKGKSLRLDRAILKQHOVALRE
ESLFSFLGKSTNYEKRIILPRVKGLPKVPFSYKQLNALATKRGQLATTNNRFGIKNTFYRKRIAPKYNLYKNP
QVNYSYLASPEYNDPQLLLSEFAKRKTDVLFVITPVNKAWADYTGELNQDKYQAAVRKIKFQLKSQGFHRIA
DFSXDGGESEYFMQDTIHLGWNGLAFDKKVQPFLETKQVPVNYKMPYPYFSKIWANRRKDLQ

SEQ ID NO: 115 polynucleotide sequence encoding GAS 202

ATGCTTAAGAGACTCTGGTTAATTCTAGGTCCCTCTCTTATTGCCCTTGTGTTTAGTAGTGATTACTATTTT
TAGTTTTCTACACAACCTTGATCATTCATAGCTCAGGAAAAAGCAAATGCCGTTGCCGATCACAGATAGTT
CTTTTAAAAATGGTTTGATTAAAAGACAAGCTTTATCAGATGAGACTTGTCTGTTTGTGCCCTTTTGTGGT
TCTAGCGAATGGAGTCGAATGGATAGTATGCACCCCTTCGGTGCTTGCAAGAGCGCTACAAGCGGAGCTATAG
ACCATTTTTTAATTGGTAAGAGAGGATCAGCATCTTTGTGCGCATTATTATGGTATACAACAAATTACCAATG
AAATGCAAAAGAAAAAGCCATCTTTGTAGTATCTCCTCAATGGTTTACTGCTCAAGGGATTAACTCTAGT
GCCGTTTCAAGTACTTGTCTAACAACCTCAAGTGATGAATTTTTACTAAAAGCTAGAACTGATAAAGAATC
ACACTTTGCGAGCAAAGCGTTTGTCTGAGCTTAACCCCTGGTGTGCTTAAATCAAACCTATTGAAAAAGTAA
GTAAGGGTAAGTCTCTTAGTTCGGTTAGACAGAGCTATTTTGAAATGTCAACATCAAGTAGCATTGAGAGAA
GACTCCCTTTTGTAGTTTTTTAGGCAAATCTACTAATATGAAAAAAGAATTTTGCCCTCGCGTTAAGGGATT
ACCTAAAGTATTTTGTATATAACAATGAATGCATTAGCAACTAAGAGAGGCCAATTAGCAACAACCAACA
ACCGTTTTTGGGATTAAAAATACATTTTATCGTAAACGAATAGCACCTAAATACAATCTTTATAAGAATTTT
CAAGTTAATTTATAGTTACCTGGCGTCACCAGAAATACAATGATTTTCAGCTTTTATATCAGAAATTTGCTAA
ACGAAAAACAGATGTACTCTTTGTATAACTCCTGTTAATAAAGCTTGGGCGGATTATACCGGCTTAAATC
AAGATAAGTATCAAGCGGCGATTGCTAAAATAAAATTCAGTTAAAGTCACAAGGATTTTCATCGCATTGCT
GACTTCTCAAAGATGGTGGTGAATCCTACTTTTATGCAAGATACCATCCATCTCGGTGGGAATGGCTGGTT
AGCTTTTGATAAGAAAGTGCAACCATTCTAGAAACGAAGCAGCCAGTGCCCAACTATAAAATGAACCCTT
ATTTTATAGTAAAATTTGGGCAAATAGGAAGACTTGCAATAG

SEQ ID NO: 116 amino acid sequence comprising GAS 057

MEKKQRFSLRKYKSGTFVSLIGSVFLVMTTVAADELSTMSEPTITNHAQQQAQHL/TNTELSSAESKSQDT
SQITLKTNRKEQSQDLVSEPTTELADTDAASMAN TGSDATQKSASLPPVNTDVHDVVKTKGAWDKGYKG
QGKVVAVIDTGIDPAHQSMRI SDVSTAKVKSKE DMLARQKAAGINYG SWINDKVVF AHNYVENS DN IKENQ
FEDFDEDMENFEFDAAEPKAIKKHKIYRPQSTQAPKETVIKTEETDGGSHDIDWTQTDDDDTKYESHGMHVT
GIVAGNSKEAAATGERFLGIAPEAQVMFMRV FANDIMGSAESLFIKAI EDAVALGADV INLSLGTANGAQL
SGSKPLMEATEKAKKAGVSVVVAAGNERVYGS DHDPLATNP DYGLV GSPSTGRTP TSVAALNSK WVIQRL
MTVKEL ENRADLNHGKALYSESVD FKDIKDSLGYDKSHQFAYVKEST DAGVNAQDVKGKIALIERDPNKTY
DEMIALAKKHGALGVLI FNNKPGQSNRSMRLTANGMGIPSAFISHEFGKAMSQ LNCNGTGSLEFDSVVS KA
PSQKGNEMNHFSNWGLTSDGYLKPDITAPGGDIYSTYNDNHYGSQTGTSMAS PQIAGASLLVKQYLEKTQP
NLPKEKIADIVKNLLMSNAQI HYNPETKTTTSPRQQAGLLNIDGAVTSGLYVTGKDNYGSI SLGNITDTM
TFDVTVHNL SNKDKTTLRYDTELLTDHVD PQKGRFTLTSHSLKTYQGGEVTV PANGKVTVRV TMDVSQFTKE
LTKQMPNGYYLEGFVRFRDSQDDQLNRVNI PFVGFKGQFENLAVAEESIYRLKSQKGTGFYFDES GPKDDI
YVGKHFTGLVTLGSETNVSTKTI SDNGLHTLGT FKNADGKFILEKNAQGNPVLAISPNGDNNQDFAAFKGV
FLRKYQGLKASVYHASDKEHK NPLWVSPESF KGDKNFNSDIRFAKSTTLLGTAFSGKSLTGAELPDGHYHY
VVSYPDVVGAKRQEMTFDMILDRQKPVLSQATFDPETNRFKPEPLKDRGLAGVRKDSVFYLERKDNKPYT
VTINDSYKYVSVEDNKTFVERQADGSPILPLDKAKLGDFY YMVEDFAGNVAI AKLGDHLPQTLGKTFIKLK
LTDGNYQTKETLKNLEMTQSDTGLVTNQAQLAVVHRNQFQSOLTKMNQDFFISPNE DGNKDFVAPKGLKN
NVYNDLTVNVYAKDDHQKQTFI WSSQAGASVSAIESTAWYGITARGSKVMPGDYQYVVV TYRDEHGKEHQKQ
YTISVNDKKPMITQGRFDTINGVDHFTPDKTKALDSSGIVREEVFYLA KKNGRKFDVTEGKDGITVSDNKV
YIPKNPDGSGYTI SKRDGVTLS DYYYLVEDRAGNVSFATLEDL KAVGKDKAVVNFGLDLPVPEDKQIVNFTY
LVRDADGKPIENLEYNNSCNSLILPYGKYTVELLTYDTNAAKLES DKIVSPTLSADNNFQQVTFKITMLA
TSQITAHFDHLLPEGSRVSLKTAQDQLI PLEQSLVVPKAYGKTVQEGTYEVV VSLPKGYRIEGNTKVM TLP

NEVHEL SLRLVKVG DASDSTGDHKVMSKNNSQALTASATPTKSTTSATAKALPSTGEKMGLKLRI VGLVLL
GLTCVFSRKKSTKD

SEQ ID NO: 117 polynucleotide sequence encoding GAS 057

GTGGAGAAAAAGCAACGTTTTCCTTGTAGAAAATACAAATCAGGAACGTTTTCGGTCTTAATAGGAAGCGT
TTTCTTGGTGTAGACAACAACAGTAGCAGCAGATGAGCTAAGCACAATGAGCGAACCACAATCAGGAATC
ACGCTCAACAACAAGCGCAACATCTCACCATAACAGAGTTGAGCTCAGCTGAATCAAAATCTCAAGACACA
TCACAAATCACTCTCAAGACAAATCGTGAAGAAAGAGCAATCACAAGATCTAGTCTCTGAGCCAACCACAAC
TGAGCTAGCTGACACAGATGACAGATCAATGGCTAATACAGGTTCTGATGCGACTCAAAAAAGCGCTTCTT
TACCGCCAGTCAATACAGATGTTTACGATTTGGGTAAAAACCAAAGGAGCTTGGGACAAGGGATACAAAGGA
CAAGGCAAGGTTGTCTGAGTTATTGACACAGGATCGATCCGGCCCATCAAAGCATGCGCATCAGTGATGT
ATCAACTGCTTAAGTAAAAATCAAAAGAAGACATGCTAGCAGCCCAAAAAAGCCCGCTATTAAATTATGGGA
GTTGGATAAAATGATAAAGTTGTTTTTGCACATAATTATCTGGAAAAATAGCGATAATATCAAAAGAAAAATCAA
TTGAGGATTTTGTATGAGGACTGGGAAACTTTGAGTTTGTATGACAGGCGAGAGCCAAAAAGCCATCAAAAA
ACACAAGATCTATCGTCCCCAATCAACCCAGGACCCGAAAGAACTGTTATCAAAACAGAAGAACAGATG
GTTACATGATATTGACTGGACACAAACAGACGATGACACCAAATACGAGTACACCGGTATGCATGTGACA
GGTATTTGTAGCCGTAATAGCAAGAAGCCGCTGCTACTGGAGAACGCTTTTATAGGAATTGCACCAGAGGC
CCAAGTCATGTTTATGCGTGTTTTGGCAACGACATCATGGGATCAGCTGAATCAC'TCTTTATCAAAGCTA
TCGAAGATGCCGTGGCTTTAGGAGCAGATGTGATCAACCTGAGTCTTGGAAACCGCTAATGGGGCACAGCTT
AGTGGCAGCAAGCCTCTAATGGAAGCAATTGAAAAAGCTAAAAAGCCCGGTGTATCAGTTGTTGTAGCAGC
AGGAAATGAGCGCGTCTATGGATCTGACCATGATGATCCATTGGCGACAAATCCAGACTATGGTTTGGTCTG
GTTCTCCCTCAACAGGTCGAACACCAACATCAGTGGCAGCTATAAACAGTAAGTGGGTGATTCAAAGCTCTA
ATGACGGTCAAAGAATTAGAAAAACCGTCCCGATTTAAACCATGGTAAAGCCATCTATTTCAGAGTCTGTGCA
CTTTAAAGACATAAAAGATAGCC'TAGGTTATGATAAATCGCATCAATTTGCTTATGTCAAAGAGTCAACTG
ATGCGGGTTATAACCCACAAGACGTTAAAGGTAAATTTGCTTTAATTGAACGTGATCCCAATAAAACCTAT
GACGAAATGATTGCTTTGGCTAAGAAACATGGAGCTCTGGGAGTACTTATTTTAAATAACAAGCCCTGGTCA
ATCAAACCGCTCAATGCGTCTAACAGCTAATGGGATGGGGATACCATCTGCTTTTCATATCGCACGAAATTG
GTAAGGCCATGTCCCAATTAAATGGCAATGGTACAGGAAGTTTAGAGTTTGACAGTGTGGTCTCAAAGCA
CCGAGTCAAAGAGCAATGAAATGAATCATTTTCAAATTTGGGGCCTAACCTCTGATGGCTATTTAAACCC
TGACATTACTGCACCAGGTGGCGATATCTATTCTACCTATAACGATAACCACTATGGTAGCCAAACAGGAA
CAAGTATGGCCTCTCTCAGATTGCTGGCGCCAGCCTTTTGGTCAAACAATACCTAGAAAAGACTCAGCCA
AAGTTGCCAAAAAGAAAAATTTGCTGATATCGTTAAGAACCTTATTGATGAGCAATGCTCAAATTCAGTTAA
TCCAGAGACAAAAACGACCACCTCACCGCGTCAGCAAGGGGCAGGATTACTTAATATTGACGGAGCTGTCA
CTAGCGSCCTTTATGTGACAGGAAAAGACAACATATGGCAGTATATCATTAGGCAACATCACAGATACGATG
ACGTTTGATGTGACTGTTTCAACCTAAGCAATAAAGACAAAACATTACGTTATGACACAGAATTGCTAAC
AGATCATGTAGACCCACAAAAGGGCCGCTTCACCTTTGACTTCTCACTCCTTAAAAACGTACCAAGGAGGAG
AAGTTACAGTCCCAGCCAATGGAAGGTGACTGTAAAGGTTACCATGGATGTCTCACAGTTTCAAAAAGAG
CTAACAAAAACAGATGCCAATGGTTACTATCTAGAAGGTTTGTCCGCTTTAGAGATAGTCAAGATGAOCCA
ACTAAATACAGTAACATTCCTTTTGTGGTTTTTAAAGGGCAATTTGAAAACCTTAGCAGTTGCAGAAGAGT
CCATTTACAGATTAAATCTCAAGGCAAAACTGGTTTTTACTTTTGATGAATCAGGTCCAAAAGACGATATC
TATGTCGGTAAACACTTTTACAGGACTTGTCACTCTTGGTTTCAGAGACCAATGTGTCAACCAAAACGATTTT
TGACAAATGGTCTACACACACTTGGCACCTTTAAAAATGCAGATGGCAAATTTATCTTAGAAAAAAATGCC
AAGGAAACCCCTGTCTTAGCCATTTCTCCAAATGGTGACAACAACCAAGATTTTGCAGCCTTCAAAGGTGTT
TCTTTGAGAAAAATATCAAGGCTTAAAGCAAGTGTCTACCATGCTAGTGAACAAGGAACACAATAATCCACT
GTGGGTCAGCCCAGAAAGCTTTAAAGGAGATAAAAACTTTAATAGTGACATTAGATT'TGCAAAATCAACGA
CCCTGTTAGGCACAGCATTTTCTGGAAATCGTTAACAGGAGCTGAATTACCAGATGGGCATTATCATTTAT
GTGGTGTCTTATTACCCAGATGTGGTGGTGCCAAACGCTCAAGAAATGACATTTGACATGATTTTAGACCG
ACAAAAACCGGTACTATCACAAGCAACATTTGATCCTGAACAAACCGATTCAAACAGAACCCCTAAAAG
ACCGTGGATTAGCTGGTGTTCGCAAGACAGTGTCTTTATCTAGAAAGAAAAGACAACAAGCCCTTATACA
GTTACGATAAACGATAGCTACAAATATGTCTCAGTAGAAGACAATAAAACATTTGTGGAGCGACAAGCTGA
TGGCAGCTTTTATCTTGGCGCTTGATAAAGCAAAATTAAGGGGATTTCTATTACATGGTTCGAGGATTTTGCAG
GGAACGTGGCCATCGCTAAGTTAGGAGATCACTTACCACAAACATTAGGTAAAAACCAATTAACTTAAAG
CTTACAGACGGTAATTTATCAGACCAAGAAACGCTTAAAGATAATCTTGAAATGACACAGTCTGACACAGG
TCTAGTCAAAATCAAGCCAGCTAGCAGTGGTGCACCGCAATCAGCCGCAAAGCCAGCTAACAAAGATGA
ATCAGGATTTCTTTATCTCACCAACGAAGATGGGAATAAAGACTTTTGTGGCCTTTAAAGGCTTGAAAAAT
AACGTGTATAATGACTTAACGTTAAGCTATACGCTAAGATGACCACCAAAAACAAACCCCTATCTGGTC

TAGTCAAGCAGGCGCTAGTGTATCCGCTATTGAAAGTACAGCCTGGTATGGCATAACAGCCCGAGGAAGCA
 AGGTGATGCCAGGTGATTATCAGTATGTTGTGACCTATCGTACGAACATGGTAAAGAACATCAAAAGCAG
 TACACCATATCTGTGAATGACAAAAACCAATGATCACTCAGGGACGTTTGTATACCATTAAATGGCGTTGA
 CCACCTTTACTCTCTGACAAGACAAAAGCCCTTGACTCATCAGGCATTGTCCGCGAAGAAGTCTTTTACTTTGG
 CCAAGAAAAATGGCCGTAAATTTGATGTGACAGAAGGTAAAGATGGTATCACAGTTAGTGACAATAAGGTG
 TATATCCCTAAAAATCCAGATGGTTCCTTACACCATTTCAAAAAGAGATGGTGTACACTGTTCAGATTATTA
 CTACCTTGTGGAAGATAGAGCTGGTAAATGTGTCTTTTGTCTACCTTGGGTGACCTAAAAGCGGTTCGGAAAAG
 ACAAAGCAGTAGTCAATTTTGGATTAGACTTACCGGTCCCTGAAGACAAACAAATAGTGAACCTTTACCTTAC
 CTTGTGCGGGATGCAGATGGTAAACCGATTGAAAACCTAGAGTATTATAATAAATCAGGTAACAGTCTTAT
 CTTGCCATACGGCAANTACACGGTCGAATTTGTGACCTATGACACCAATGCAGCCAACTAGAGTCAGATA
 AAATCGTTTCTTTTACCTTGTGAGCTGATAACAACCTCCAACAAGTTACCTTTAAGATAACGATGTTAGCA
 ACTTCTCAAATAACTGCCCACCTTTGATCATCTTTTGGCAGAAGGCAGTCGCGTTAGCCTTAAAACAGCTCA
 AGATCAGCTAATCCCGCTTGAACAGTCTTGTATGTGCTTAAAGCTTATGGCAAAACCGTTCAAGAAGGCA
 CTTACGAAGTTGTTGTGAGCCTGCTTAAAGCTACCGTATCGAAGGCAACACAAAGGTGAATACCTTACCA
 AATGAAGTGCACGAATATCATTACGCCCTTGTCAAAGTAGGAGATGCCCTCAGATTCAACTGGTGATCATAA
 GGTATGTCAAAAAATAATTCACAGGCTTTGACAGCCTCTGCCACACCAACCAAGTCAACGACCTCAGCAA
 CAGCAAAAGCCCTACCATCAACGGGTGAAAAATGGGTCTCAAGTTGCGCATAGTAGGTCTTGTGTTACTC
 GGACTTACTTGGCTCTTTAGCCGAAAAAAATCAACCAAGATTGA

SEQ ID NO: 118 amino acid sequence comprising N-terminal leader sequence of GAS 57
 MEKKQRFSLRKYKSGTFSVLIGSVFLVMTTVA

SEQ ID NO: 119 amino acid sequence comprising a fragment of GAS 57 where the N-terminal leader sequence is removed

ADELSTMSEPTITNHAQQQAQHLTNTELSSAESKSQDTSQITLKTNRKEBQSQDLVSEPTTTELADTDAA
 MANTGSDATQKSASLPPVNTDVHDWVKTKGAWDKGYKQGKVVAVIDTGIDPAHQSMRISDVSTAKVKSKE
 DMLARQKAAGINYGSWINDKVFAHNYVENS DN I KENQFEDFDEDWENFEDAEAEPAIKKHKIYRPOST
 QAPKETVIKTEFTDGS HD IDWTQTD DDTKYESHGMHVTGIVAGNSKEAAATGERFLGIAPEAQVMFMRVFA
 NDIMGSAESLFIKATEDAVALGADVINLSLGTANGAQLSGSKPLMRAIEKAKKAGVSVVVAAGNERVYGS
 HDDPLATNPDYGLVGSPTGRTPTSVAAINSKWVIOQLMTVKELENRADLMHGKAIYSESVDPKDIKDSL
 YDKSHQFAYVKESTDAAGYNAQDVKGKIALIERDPNKTYDEMIALAKKHGALGVLI FNNKPGQSNRSMRLTA
 NGMGIPSAFISHEFGKAMSQNLNGNGTGSLEFDSVVS KAPSQKGNEMNHFSNWGLTSDGYLKPDITAPGGDI
 YSTYNDNHYGSQTGTSMA SPQIAGASLLVQKYLEKTQPNLPKEKLADIVKNLLMSNAQIHVNPETKTTTSP
 RQQCAGLLNIDGAVTSGLYVTGKDNYSISLGNITDTMTFDVTVHNLSNKDKTLRYDTBLLTDHVDPQKGR
 FTLTSHSLKTYQGGEVTVPANGKVTVRVMTDVSQFTKELTKQMPNGYYLEGFVRFRDSQDDQLNRVNTPFV
 GFKGQFENLAVAEESIYRLKSQGTGFYFDES GPKDDIYVGKHFTGLVTLGSETNVSTKTISDNGLHTLGT
 FRNADGKFILEKNAQGNFVLAISPNGDNNQDPAAFKGVFLRKYQGLKASVYHASDKEHKNPLWVSPESFKG
 DKNFNSDIRFAKSTTLTGTA FSGKSLTGAELPDGHYHYVVSYPYDVVGAKRQEMTFDMILDRQKPVLSQAT
 FQFETNRFKPEPLKDRGLAGVRKDSVFYLERKDNKPYVTINDSYKYVSVEDNKT FVERQADGGSFILPLDK
 AKLGDFYIMVEDFAGNVAIAKLGDLHPQTLGKTPIKLKLTDGNYQTKETLKDNLMTQSDTGLVTNQAQLA
 VVHRNQPSQLTKMNQDFFISPNEGDKDFVAFKGLKNNVYNDLTVNVYAKDDHQKQTFIWSSQAGASVSA
 IESTAWYGITARGSKVMPGDYQYVVVYRDEHGKEHQKYTISVNDKKPMITQGRFDITNGVDHFTPDKTKA
 LDSSGIVREEV FYLAKKNGRKFVTEGKGITVSDNKVYIPKNPDGSYTI SKRDGVTLSDYLYLVEDRAGN
 VSFATLRDLKAVGKDKAVVNFGLDLVPVPEDKQIVNFYLVLRDADGKPIENLEYNNSGNSLILPYGKYTVE
 LLTYDTNAAKLES DKIVSFTLSADNNFQQVTFKITNLATSQITAHFDHLLPEGSRVSLKTAQDQLI PLEQS
 LYVPKAYGKTVQEGTYEVVVSLEPKGYRIEGNTKVNLTLPNEVHEL SLRLVKVGDA SDSTGDHKVMSKNNSQA
 LTASATPTKSTTSATAKALPSTGEKMGKLRIVGLVLLGLTCVFSRKKSTKD

SEQ ID NO: 120 amino acid sequence comprising C-terminal hydrophobic region
 LPSTGEKMGKLRIVGLVLLGLTCVFSRKKSTKD

SEQ ID NO: 121 amino acid sequence comprising a fragment of GAS 57 where the C-terminal hydrophobic region is removed

MEKKQRFSLRKYKSGTFSVLIGSVFLVMTTVAADELSTMSEPTITNHAQQQAQHLTNTELSSAESKSQDT
SQITLKTNRKQBQSQDLVSEPTTTELADTDAASMANTGSDATQKSASLPPVNTDVHDWVKTKGAWDKGYKG
QGGKVVAVIDTGLDPAHQSMRISDVSTAKVKSKEMLARQKAAGINYGSWINDKVVFAHNYVENS DN I KENQ

FEDFEDWENFEFDAEAEPKAIKKHKIYRPQSTQAPKETVIKRETDGSHDIDWTQTD DDTKYESHGMHVT
GIVAGNSKEAAATGERFLGIAPEAQVMFMRVFAINDIMGSAESLFIKAIEDAVALGADV INLSLGTANGAQL
SGSKPLMEAIEKAKKAGVSVVVAAGNERVYGSDDHDDPLATNPDYGLVGSPTGRTPTSVAAINSKWVIQRL
MTVKELNENRADLNHGKAIYSESVDFFKDIKDSLGYDKSHQFAYVKESTADAGYNAQDVKGKIALIERDPNKTY
DEMIALAKKHGALGVLIFFNNKPGQSNRSMRLTANGMGIPSAFISHEFGKAMSQNLNGTGSLEFDSVVS KA
PSQKGNEMNHFSNWGLTSDGYLKPDIITAPGGDIYSTYNDNHYGSQTGTSMASPOIAGA SLLVKQYLEKTQP
NLPKEKIADIVKNLLMSNAQIHVNPETKTTTSPRQOGAGLLNIDGAVTSGLYVTGKDN YGSI SLGNITDTM
TFDVTVHNL SNKDKTLRYDTELLTDHVDPPQGRFTLTSLSLKYQGGEVTVPANGKVTVRVTMDVSQPTKE
LTKQMPNGYYLEGFVFRFRDSQDDQLNRVNI PFVGFPGQFENLAVAEESIYRLKSQGKTGFYFDES GPKDDI
YVGKHTGLVTLGSETNVSTKTI SDNGLHTLGTFFKNADGKFILEKNAQGNPVLAI SPNGDNNDFAAFKGV
FLRKYQGLKASVYHASDKEHKNPLWVSPESFKGDKNFNSDIRFAKSTTLTGTA FSGKSLTGAELPDGHVHY
VVSYPDVVGAQRQEMTFDMILDRQKPVLSQATFDPETNRFKPEPLKDRGLAGVRKDSVFYLERKDNKPYT
VTINDSYKYVSVEDNKT FVERQADGSFILPLDKAKLGDFYVMVEDFAGNVAIAKLGDLHPQTLGKTPIKLK
LTDGNYQTKETLKNLEMTQS D TGLVTNQAQLAVVHRNQPSQLTKMNQDFFI SPNEDGNKDFVAFKGLKN
NVYNDLTVNVYAKDDHQKQTP IWSSQAGASVSAIESTAWYGITARGSKVM PGDYQYVV TYRDEHGKEHQKQ
YTI SVNDKKPMITQGRFDTINGVDHFTPDKTKALDSSGIVRBEV FYLAKKNGRKFVDT EGGKDGITVSDNKV
YIPKNPDGSYTI SKRDGVTLS DYYLVEDRAGNVSFATLRDLKAVGKDKAVVNFGLDL PVPEDKQIVNFTY
LVRDADGKPIENLEYYNNSGNLSILPYGKYTVELLTYDTNAAKLES DKIVSFTLSADNNFQQVTFKITMLA
TSQITAHFDHLLPEGSRVSLKTAQDQLI PLEQSLYVPKAYGKTVQEGTYEVVVS LPKGYRIEGNTKVNTLP
NEVHEL SLRLVKVGDA SDSTGDHKVMSKNNSQAL TASATPTKSTTSATAKA

SEQ ID NO: 122 amino acid sequence comprising a fragment of GAS 57 where both the N-terminal leader sequence and the C-terminal hydrophobic region are removed

ADELSTMSEPTITNHAQQQAQHLTNTELSSAESKSQDTSQITLKTNRKQBQSQDLVSEPTTTELADTDAAS
MANTGSDATQKSASLPPVNTDVHDWVKTKGAWDKGYKGQGGKVVAVIDTGLDPAHQSMRISDVSTAKVKSKE
DMLARQKAAGINYGSWINDKVVFAHNYVENS DN I KENQFEDFEDWENFEFDAEAEPKAIKKHKIYRPQST
QAPKETVIKRETDGSHDIDWTQTD DDTKYESHGMHVTGIVAGNSKEAAATGERFLGIAPEAQVMFMRVFA
NDIMGSAESLFIKAIEDAVALGADV INLSLGTANGAQLSGSKPLMEAIEKAKKAGVSVVVAAGNERVYGS D
HDDPLATNPDYGLVGSPTGRTPTSVAAINSKWVIQRLMTVKELNENRADLNHGKAIYSESVDFFKDIKDSL G
YDKSHQFAYVKESTADAGYNAQDVKGKIALIERDPNKTYDEMIALAKKHGALGVLIFFNNKPGQSNRSMRLTA
NGMGIPSAFISHEFGKAMSQNLNGTGSLEFDSVVS KAPSQKGNEMNHFSNWGLTSDGYLKPDIITAPGGDI
YSTYNDNHYGSQTGTSMASPOIAGASLLVKQYLEKTQPNLPKEKIADIVKNLLMSNAQIHVNPETKTTTSP
RQOGAGLLNIDGAVTSGLYVTGKDN YGSI SLGNITDTMTFDVTVHNL SNKDKTLRYDTELLTDHVDPPQGR
FTLTSLSLKYQGGEVTVPANGKVTVRVTMDVSQPTKELTKQMPNGYYLEGFVFRFRDSQDDQLNRVNI PFV
GFKGQFENLAVAEESIYRLKSQGKTGFYFDES GPKDDIYVGKHTGLVTLGSETNVSTKTI SDNGLHTLGT
FFKNADGKFILEKNAQGNPVLAI SPNGDNNDFAAFKGVFLRKYQGLKASVYHASDKEHKNPLWVSPESFKG
DRNFNSDIRFAKSTTLTGTA FSGKSLTGAELPDGHVHYVVSYPDVVGAQRQEMTFDMI LDRQKPVLSQAT
FDPETNRFKPEPLKDRGLAGVRKDSVFYLERKDNKPYT VTINDSYKYVSVEDNKT FVERQADGSFILPLDK
AKLGDFYVMVEDFAGNVAIAKLGDLHPQTLGKTPIKLKLTDGNYQTKETLKNLEMTQS D TGLVTNQAQLA
VVHRNQPSQLTKMNQDFFI SPNEDGNKDFVAFKGLKNNVYNDLTVNVYAKDDHQKQTP IWSSQAGASVSA
IESTAWYGITARGSKVM PGDYQYVV TYRDEHGKEHQKQYTI SVNDKKPMITQGRFDTINGVDHFTPDKTKA
LDSSGIVRBEV FYLAKKNGRKFVDT EGGKDGITVSDNKVYIPKNPDGSYTI SKRDGVTLS DYYLVEDRAGN
VSFATLRDLKAVGKDKAVVNFGLDL PVPEDKQIVNFTY LVRDADGKPIENLEYYNNSGNLSILPYGKYTVE
LLTYDTNAAKLES DKIVSFTLSADNNFQQVTFKITMLATSQITAHFDHLLPEGSRVSLKTAQDQLI PLEQS
LYVPKAYGKTVQEGTYEVVVS LPKGYRIEGNTKVNTLPNEVHEL SLRLVKVGDA SDSTGDHKVMSKNNSQA
LTASATPTKSTTSATAKA

SEQ ID NO: 123 amino acid sequence of a GAS M protein

MAKNNTNRHYSRLKLTGTASVAVALTVLGAGFANQTEVKANGDGNPREVIEDLAANNPATQNTIRLAYENK
DLKARLENAMEVAGRDFKRAEELEKAKQALEQQRKDLETKLKELQOQYDLAKESTSWDRQRLKEKELEEKKE
ALELAIDQASRDYHRATALEKELEEKKKALELAIDQASQDYNRANVLEKELETITREQEINRNLLGNAKLE
LDQLSSEKEQLTIEKAKLEEEKQISDASRQSLARDLDASREAKKQVEKDLANLTAELDKVKEDKQISDASR
QGLRRDLASREAKKQVEKDLANLTAELDKVKEEKQISDASRQGLRRDLASREAKKQVEKALEEANKLA
ALEKLNKELEESKKLTEKEKAELEAKLEAEKALKEQLAKQAEELAKLRAGKASDSQTFIDTKPGNKAVPGK
GQAPQAGTKPNQNKAPMKETKRQLPSTGETANFFFTAAALTVMATAGVAAVVKRKEEN

SEQ ID NO: 124 amino acid sequence of GAS SfbI

MSFDGFFLHHLTNELKENLLYGRIQKVNQPPERELVLTIRNHRKNYKLLLSAHPVFGRVQITQADFQNPQV
PNTFTMIRKYLQCAVIEQLEQIDNDRIIEIKVSNKNEIGDAIQATLIIEIMGKHSNITLVDRANKEIES
IKHVGFSSQNSYRITLPGSTYIEFPKTAAVNPFTITDVPLFELQTQELTVKSLQQHFQGLGRDTAKELAE
LTTDKLKRFRFFARPTQANLTTASFAPVLFSDSHATFETLSMDLDHFYQDKAERDRINQASDLIHRVQT
ELDKNRNKLKQEAELLATENAELEFRQKGEELLTYLSLVPNNQDSVILDNYITGEEKIETALDKALTPNQNA
QRYFKKYQKLKEAVKHLISGLIADTKQSTTYFESVDYNLSQASIDDIETIREELYQAGFLKSRQRDKRHKRK
KPEQYLASDGTITLMVGRNNLQNEELTFKMAKKGELWPHAKDIPGSHVITKDNLDPSDEVKTDAAELAAYY
SKARLSNLVQVDMIEAKKLHRPSGAKPGFVYTYTGQKTLRVTPDQAKILSMKLS

SEQ ID NO: 125 amino acid sequence of a GAS Shp protein

MTKVVIKQLLQVIVVFMISLSTWNTLVYADKGGQIYGCIIQRNYRHPIISGQIEDSGGEHSFDIGQGMVEGTV
YSDAMLEVSDAGIXVLTFRMSLADYSGNYQFWIQPGGTGSGQAVDYNITQKGTDTNGTTLDAISLPTVNS
IIRGSMFVEPMGREVVFFYLSASELIQKYSGNMLAQLVTETDNSQNQEVKDSQKPVDTKLGESQDESHTGAM
ITQNKPKANSSNNKSLSDKKILPSKMGLTTSLELKKEDKFRSKKDLSIMIYYFPTFFLMLGGFAVWVKRR
KKNDKTM

SEQ ID NO: 126 amino acids 10 to 30 of GAS protein Saga
FSIATGSGNSQGGSGSYTPGKC

SEQ ID NO: 127 polynucleotide sequence comprising fusion construct 117-40a-RR

ATGGCCTTTAACACAAGCCAGAGTGTCAAGTGCACAACTTTATAGCAATGAAGGGTATCACCAGCATTGAC
 TGATGAAAAATCACACCTGCAATATAGTAAAGACAACGCACAACCTTCAATTGAGAAATATCCTTGACGGCT
 ACCAAAATGACCTAGGGAGACACTACTCTAGCTATTATTACTACAACCTAAGAACCTTATGGGACTATCA
 AGTGAGCAAGACATTGAAAAACACTATGAAGAGCTTAAGAACAAAGTTACATGATATGTACAATCATTATG
 TAGCGGTGGCGGATCCATGAGTGTAGGCGTATCTCACCAGTCAAAGCAGATGATAGAGCCTCAGGAGAAA
 CGAAGCGGAGTAATACTCACGACGATAGTTTACCAAAACCAGAAACAAATTCAGAGGCAAGGCAACTATT
 GATGCACTTGAaaaaactctcagTCAACAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAAC
 TACTGCTGAAATCAACCACTTAAAAGAGCAGCAAGATAATGAACAAAAAGCTTTAACCTCTGCACAAGAAA
 TTTTACATAATACTCTTGCAAGTAGTGAGGAGACGCTATTTAGCCCAAGGAGCCGAACATCAAAGAGAGTTA
 ACAGCTACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAAGAGAGCTGCATTGTGCAACAA
 AAAAGCTAGCATTTCAGCAGAAACTACTCGAGCTCAAGATTTAGTGGAAACAGTCAAACAGCTCTGAACAAA
 ATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCATTAAAGCAGCTCAAACGGCTAATGAT
 AATACAAAAGCATTAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAAATCAAAAAGCTAAAGTTAA
 AAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAGAGGCGAGAACTTAGTCTGCT
 TTAATTCCTCAGCTCCGCTCTACTCAAGATAGCAATTGTGGGTAATAATACCATGAAAGCACCGCAAGCTAT
 CCTCTTGAAGAACCTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAAPTATTACAA
 AGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCAG
 CAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCAGAAAGTGCAAAATGAGCTAGCGCAGTTTSCA
 GCTCACATGATTAATAGTGTACGtCGtCAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAGA
 ATTTSCAAGATTACTTAGTACCAGCTATAAGAAAACCTCATGGTAATACAAGACCATTATTGTTCTACGGAC
 AGCCAGGGGTATCAGGGCATTATGTTGTTGGGCTCATGATAAACTATTTATTGAAGACTCTGCCGAGCG
 TCAGGGCTCATTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAAACGATGTGCATCTGTGAA
 TGGTATTAAACGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTTACACGGAATACAT
 ACGGCCATGCTATTAACTTTTTACGTGTAGATAAACATAACCTAATGCGCCTGTTTACCTTGGATTTTCA
 ACCAGCAATGTAGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTAACCATCAACG
 CTTTAAATAAGACCCCTATAAAAGCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCATGATCTG
 ATACTATTGACGCGATCAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCCGCTATTTCATCAAGAAGCT
 GATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGA
 CAGCTTAAATCTCCAAGTGAGACAAATAATGATACTAAAGGTTCTTTGAGAACAGAACTACTAGCAGCTA
 AAGCAAAACAAGCACAACCTCGAAGCTACTCGTGATCAATCATTAGCTAAAGCTAGCATCGTTGAAGCCGCA
 CTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGAGCCAGAGTGACAGCACTGGTGGCTAAAAAGCTCA
 TTTGCAATATCTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACGTGAGCGCATGATAATA
 CTAAGCAAGATTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCTTTAGCAGCCTTACAAGCT
 AAACAAAGCAGTCTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTTTGCTTAAACCTTAGCTAA
 CGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCTGATTTGCAAGTAGCTCCACCTCTTA

 CGGGCGTAAAACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTTCAGAAATGGTTAAAGAAACG
 AAACAACCTATTAGAAGCTTCAGCAAGATTAGCTGTGAAATAACAAGTCTTGTAGCAGAAGCGCTTGTGG
 CCAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGCTTAAATCACATCTTCGATTACTCAGCCCTCAT
 CTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATTTCGATGTTGATGAAAGTACTCAA
 cGtgcgccgcgactcgagCACCACCACCACCACCAC

SEQ ID NO: 128 amino acid sequence comprising fusion construct 117-40a-RR

MAFNTS QSVSAQVYSNEG YH QH L T D E K S H L Q Y S K D N
A Q L Q L R N I L D G Y Q N D L G R H Y S S Y Y Y Y N L R T V M G L S S
E Q D I E K H Y E E L K N K L H D M Y N H Y A S G G G S M S V G V S H Q
V K A D D R A S G E T K A S N T H D D S L P K P E T I Q E A K A T I D A
V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H L K E Q Q D N E
Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E H Q R E L T A T
E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E T T R A Q D L V
E Q V K T S E Q N I A K L N A M I S N P D A I T K A A Q T A N D N T K A
L S S E L E K A K A D L E N Q K A K V K K Q L T E E L A A Q K A A L A E
K E A E L S R L K S S A P S T Q D S I V G N N T M K A P Q G Y P L E E L
K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A K A S P G N Q L N
Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L A Q F A A H M I N S
V R R Q L G L P P V T V T A G S Q E F A R L L S T S Y K K T H G N T R P
S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A G A S G L I R N D
D N M Y E N I G A F N D V H T V N G I K R G I Y D S I K Y M L F T D H L
H G N T Y G H A I N F L R V D K H N P N A P V Y L G F S T S N V G S L N
E H F V M F P E S N I A N H Q R F N K T P I K A V G S T K D Y A Q R V G
T V S D T I A A I K G K V S S L E N R L S A I H Q E A D I M A A Q A K V
S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T K G S L R T E L L
A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L H Q T E A L A E Q
A A A R V T A L V A K K A H L Q Y L R D F K L N P N R L Q V I R E R I D
N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q S S L E A T I A T
T E H Q L T L L K T L A N E K E Y R H L D E D I A T V P D L Q V A P P L
T G V K P L S Y S K I D T T P L V Q E M V K E T K Q L L E A S A R L A A
E N T S L V A E A L V G Q T S E M V A S N A I V S K I T S S I T Q P S S
K T S Y G S G S S T T S N L I S D V D E S T Q R A A A L E H H H H H

SEQ ID NO: 129 amino acid sequence comprising a linker in the 117-40a-RR construct
YASGGGS

SEQ ID NO: 130 polynucleotide sequence comprising 40a-RR-117 fusion construct

ATGAGTGTAGGCGTATCTCACCAGTCAAAGCAGATGATAGAGCCTCAGGAGAAAACGAAGGCGAGTAATAC
TCACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAGGCAACTATTGATGCAGTTGAAAAA
CTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAACTACTGCTGAAATCAAC
CACTTAAAAGAGCAGCAAGATAATGAACAAAAGCTTTAACCTCTGCACAAGAAATTTACACTAATACTCT
TGCAAGTAGTGAAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCTACTGAAACAG
AGCTTCATAATGCTCAAGCAGATCAACATTCAAAGAGAGCTGCATTGTCAGAACAAAAGCTAGCATTTCAG
GCAGAACTACTCGAGCTCAAGATTTAGTGGAAACAGTCAAACGCTCTGAACAAAATATTGCTAAGCTCAA
TGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAATACAAAAGCATTAA
GCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAGCTAAAGTTAAAAAGCAATTGACTGAA
GAGTTGCGAGCTCAGAAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGCTCTTAAATCCTCAGCTCC
GTCTACTCAAGATAGCATTTGTGGGTAATAATACCATGAAAGCACCGCAAGGCTATCCTCTTGAAGAAGCTTA
AAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAAAGAGCATGCAGATCAA
ATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCCAGCAGATCGTAATCGCTT
TGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTTGCAGCTCACATGATTAAATA
GTGTACGtCGtCAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCACAAAGAATTTGCAAGATTACTT
AGTACCAGCTATAAGAAAACTCATGGTAATACAAGACCATCAATTGCTCTACGGACAGCCAGGGGTATCAGG
GCATTATGGTGTGGGCTCATGATAAACTATTATTGAA GACTCTGCCGGAGCGTCAGGGCTCATTCGAA
ATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATACTGTGAATGGTATTAAACGTGGT

ATTTATGACAGTATCAAGTATATGCTCTTTACA.GATCATTTACACGGAAATACATACGGCCATGCTATTAA
 CTTTTTACGCTAGATATAACATAACCCCTAATGCCGCTGTTTACCCTTGGATTTTCAACCAGCAATGTAGGAT
 CTTTGAATGAACACTTTTGAATGTTTCCAGAGTCTTAACATTGCTAACCATCAACGCTTTAATAAGACCCCT
 ATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGGCAGTGTATCTGATACTATTGCAGCGAT
 CAAAGGAAAAGTAAGCTCATTAGAAAATCGTTTGTCCGCTATTTCATCAAGAAAGCTGATATTATGGCAGCCC
 AAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACACTTAAGCAGTCAGACAGCTTAAATCTCCRA
 GTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGAATTACTAGCAGCTAAAGCAAAACAAGCACA
 ACTCGAAGCTACTCGTGATCAATCATTAGCTAA.GCTAGCATCGTTGAAAGCCGCACTGCACCAGACAGAAG
 CCTTAGCAGAGCAAGCCGAGCCAGAGTGACAGCACTGGTGGCTAAAAAAGCTCATTTGCAATATCTAAGG
 GACTTTAAATTGAATCCTAACCCGCTTCAAGTGATCGTGAGCGCATTGATAATACTAAGCAAGATTTGGC
 TAAAACCTACCTCATCTTTGTTAAATGCACAAAGA.GCTTTAGCAGCCTTACAAGCTAAACAAAGCAGCTTAG
 AAGCTACTATTGCTACCACAGAACACCAAGTTGAC.TTTTGCTTAAACCTTAGCTAACGAAAAGGAATATCGC
 CACTTAGACGAAGATATAGCTACTGTGCCTGATTTTGCAAGTAGCTCCACCTCTTACGGCGTAAACCCGCT
 ATCATATAGTAAGATAGATACTACTCCGCTTGT.TCAAGAAATGGTTAAAGAAACGAAACAACCTATTAGAAG
 CTTCAGCAAGATTAGCTGCTGAAAATACAAGTCTTTGTAGCAGAAGCGCTTGTGGCCAAACCTCTGAAATG
 GTAGCAAGTAAATGCCATGTGTCTAAAATCACA.TCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGG
 CTCAGGATCTTCTACAACGAGCAATCTCATTPC.TGATGTTGATGAAAGTACTCAACGCTGCTAGCGCTGCGC
 GATCCCATGGCCTTTAACACAAGCCAGAGTGTCA.GTGCAACAAGTTTATAGCAATGAAGGGTATCACCAGCAT
 TTGACTGTATGAAAAATCACACCTGCAATATAGTAAAGACAACGCACAACCTTCAATTGAGAAATATCCTTGA
 CGGCTACCAAAATGACCTAGGGAGACACTACTCT.TAGCTATTATTACTACAACCTAAGAACCCTTATGGGAC
 TATCAAGTGAGCAAGACATTGAAAAACACTATGAAAGAGCTTAAAGAAACAAGTTACATGATATGTACAATCAT
 TATGcgggcgcgactcgagCACCACCAACCACACAC

SEQ ID NO: 131 amino acid sequence comprising the 40a-RR-117 fusion construct

M S V G V S H Q V K A D D R A S G E T K A S N T H D D S L P K P E T I Q
 E A K A T I D A V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H
 L K E Q Q D N E Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E
 H Q R E L T A T E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E
 T T R A Q D L V E Q V K T S E Q N I A K L N A M I S N P D A I T K A A Q
 T A N D N T K A L S S E L E K A K A D L E N Q K A K V K K Q L T E E L A
 A Q K A A L A E K E A E L S R L K S S A P S T Q D S I V G N N T M K A P
 Q G Y P L E E L K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A K
 A S P G N Q L N Q Y Q D I P A D R N R F V D P D N L T P E V Q N E L A Q
 F A A H M I N S V R R Q L G L P P V T V T A G S Q E F A R L L S T S Y K
 K T H G N T R P S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A G
 A S G L I R N D D N M Y E N I G A F N D V H T V N G I K R G I Y D S I K
 Y M L F T D H L H G N T Y G H A I N F L R V D K H N P N A F V Y L G F S
 T S N V G S L N E H F V M P P E S N I A N H Q R F N K T P I K A V G S T
 K D Y A Q R V G T V S D T I A A I K G K V S S L E N R L S A I H Q E A D
 I M A A Q A K V S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T K
 G S L R T E L L A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L H
 Q T E A L A E Q A A A R V T A L V A K K A H L Q Y L R D F K L N P N R L
 Q V I R E R I D N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q S
 S L E A T I A T T E H Q L T L L K T L A N E K E Y R H L D E D I A T V P
 D L Q V A P P L T G V K P L S Y S K I D T T P L V Q E M V K E T K Q L L
 E A S A R L A A E N T S L V A E A L V G Q T S E M V A S N A I V S K I T
 S S I T Q P S S K T S Y G S G S S T T S N L I S D V D E S T Q R A S G G
 G S M A F N T S Q S V S A Q V Y S N E G Y H Q H L T D E K S H L Q Y S K
 D N A Q L Q L R N I L D G Y Q N D L G R H Y S S Y Y Y N L R T V M G L
 S S E Q D I E K H Y E E L K N K L H D M Y N H Y A A A L E H H H H H

SEQ ID NO: 132 polynucleotide sequence comprising fusion construct GAS 117 - 40a

ATGGCCTTTAACACAAGCCAGAGTGTCAAGTGCAC.AAGTTTATAGCAATGAAGGGTATCACCAGCATTTGAC
 TGATGAAAAATCACACCTGCAATATAGTAAAGACAACGCACAACCTTCAATTGAGAAATATCCTTGACGGCT
 ACCAAAATGACCTAGGGAGACACTACTCTAGCTA.TTATTACTACAACCTAAGAAC.GTTATGGGACTATCA
 AGTGAGCAAGACATTGAAAAACACTATGAAGAGCTTAAAGAAACAAGTTACATGATATGTACAATCATTTATGC

cgtggtggcgatccatgagtgtagggctatctcaccagtcagagcagatgatagagcctcaggagaaa
 cgaaggcgagtaatactcacgacgatagtttaccaaaaccagaaacaattcaagaggcgaaggcactatt
 gatgcagttgaaaaaactctcagctcaacaaaaaccagaactgcagagcttgctaccgctctgcacaaaaac
 tactgctgaaatcaaccac[**T**aaaaagagcagcagataatgaacaaaagctttaacctctgcacaaagaaa
 tttaactaatactcttgcaagtagtgaggagacgctatttagcccaaggagccgaacatcaaaagagagtta
 acagctactgaaacagagcttcataatgctcaagcagatcaacattcaaaagagactgcattgtcagaaca
 aaaagctagcattttcagcagaaaactactcagagctcaagatttagtggaacaagtcaaaacgctcgaacaaa
 atattgctaagctcaatgctatgattagcaatcctgatgctatcactaaagcagctcaaacggctaattgat
 aatacaaaagcatttaagctcagaatttgagagaaggctaaagctgacttagaaaaatcaaaaagctaaagttaa
 aaagcaattgactgaagagttggcagctcagaaagctgctctagcagaaaaagaggcagaacttagtcgctc
 tttaaattcctcagctccgctactcaagatagcattgtgggtaataataccatgaaagcaccgcaaggctat
 cctcttgaagaaacttaaaaaattagaagctagtggttatatttgatcagctagtataataattattacaa
 agagcatgcagatcaaatatttgccaaagctagtcaggtaatacaataaataccaaagatattccag
 cagatcctaattcgcttggttgatccccgataatttgacaccagaagtgcaaaatg[**G**ctagccgagtttgca
 gctcacatgatataatagtgtaagaagacaatttaggtctaccaccagttactgttacagcaggatcacaaga
 atttgcaagattacttagtacagctataagaaaactcatggttaatacaagaccatcatttgctacggac
 agccaggggtatcagggcatttatgggtgtgggctcatgataaaaactattattgaagacttgccggagcg
 tcagggtcatttcgaatgatgataacatgtacgagaatatcggtgcttttaacgatgtgcatactgtgaa
 tgggtattaaacgtggtattttatgacagtatcaagtatatgctctttacagatcattttacacggaaatcacat
 acggccatgctatttaactttttacgtgtagataaacataaccctaattgcgctgtttaccttggaattttca
 accagcaattgtaggactcttgaatgaacactttgtaattgtttccagagctcaacattgcttaaccatcaacc
 ctttaataagacccctataaaaagccgttggaagtacaaaagattatgcccaagagtaggcaactgtatctg
 atactattgcagcgatcaaaaggaagagtaagctcattagaaaatcggtttgctggctattcatcaagaagct
 gatattatggcagcccaagctaaagtaagctcaacttcaaggtaaattagcaagcacacttaagcagtcaga
 cagcttaaatctccaagtgagacaattaaatgatactaaaggttcttttgagaacagaattactagcagcta
 aagcaaaacaagcacaaactcgaagctactcgtgatcaatcattagctaagctagcatcgttgaaagccgca
 ctgcaccagacagaagccttagcagagcaagccgagccagagtgacagcactgggtggctaaaaagctca
 tttgcaatatcttaagggacttttaaattgaatcctaaccgcttcaagtgatagctgagcgcatttgataata
 ctaagcaagatttggtcaaaactacctcatctttgttaaatgcacagaagcttttagcagccttacaagct
 aaacaaagcagctctagaagctactatttgctaccacagaacaccagttgactttgcttaaaaccttagctaa
 cgaaaaggaatatcgccacttagacgaagatatagctactgtgcctgatttgcaagtagctccacctcttca
 cggcgctaaaaccgctatcatatagtaagatagatactactccgcttggttcaagaaatgggttaaagaaacg
 aaacaaactattagaagcttcagcaagatttagctgtaaaatataagctcttgtagcagaagccttggttg
 ccaaacctctgaaatggtagcaagtaattgccattgtgtctaaatcacatcttgcattactcagccctcat
 ctaagacatcttatggctcaggatcttctacaacgagcaatctcattttctgatgttgatgaaagtaactcaa
 cgtgcccgcgactcgagcaccaccaccaccac

SEQ ID NO: 133 amino acid sequence comprising fusion construct GAS 117-40a

M A F N T S Q S V S A Q V Y S N E G Y H Q H L T D E K S H L Q Y S K D N
 A Q L Q L R N I L D G Y Q N D L G R H Y S S Y Y Y Y N L R T V M G L S S
 E Q D I E K H Y E E L K N K L H D M Y N H Y A S G G S M S V G V S H Q
 V K A D D R A S G E T K A S N T H D D S L P K P E T I Q E A K A T I D A
 V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H L K E Q Q D N E
 Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E H Q R E L T A T
 E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E T T R A Q D L V
 E Q V K T S E Q N I A K L N A M I S N P D A I T K A A Q T A N D N T K A
 L S S R L E K A K A D L E N Q K A K V K K Q L T E E L A A Q K A A L A E
 K E A E L S R L K S S A P S T Q D S I V G N N T M K A P Q G Y P L E E L
 K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A K A S P G N Q L N
 Q Y Q D I P A D R N R F V D P D N L T P E V Q N G L A Q F A A H M I N S
 V R R Q L G L P P V T V T A G S Q E F A R L L S T S Y K K T H G N T R P
 S F V Y G Q P G V S G H Y G V G P H D K T I I E D S A G A S G L I R N D
 D N M Y E N I G A F N D V H T V N G I K R G I Y D S I K Y M L F T D H L
 H G N T Y G H A I N F L R V D K H N P N A P V Y L G F S T S N V G S L N
 E H F V M F P E S N I A N H Q R F N K T P I K A V G S T K D Y A Q R V G
 T V S D T I A A I K G K V S S L E N R L S A I H Q E A D I M A A Q A K V

S Q L Q G K L A S T L K Q S D S L N L Q V R Q L N D T K G S L R T E L L
 A A K A K Q A Q L E A T R D Q S L A K L A S L K A A L H Q T E A L A E Q
 A A A R V T A L V A K K A H L Q Y L R D F K L N P N R L Q V I R E R I D
 N T K Q D L A K T T S S L L N A Q E A L A A L Q A K Q S S L E A T I A T
 T E H Q L T L L K T L A N E K E Y R H L D E D I A T V P D L Q V A P P L
 T G V K P L S Y S K I D T T P L V Q E M V K E T K Q L L E A S A R L A A
 E N T S L V A E A L V G Q T S E M V A S N A I V S K I T S S I T Q P S S
 K T S Y G S G S S T T S N L I S D V D E S T Q R A A A L E H H H H H

SEQ ID NO: 134 polynucleotide sequence comprising fusion construct GAS 117-40N

ATGGCCTTTAACACAAGCCAGAGTGTCTAGTGCACAAGTTTATAGCAATGAAGGGTATCACCAGCATTGAC
 TGATGAAAAATCACACCTGCAATATAGTAAAGACAACGCACAACCTCAATTGAGAAATATCCTTGACGGCT
 ACCAAAATGACCTAGCGAGACACTACTCTAGCTATTATTACTACAACCTAAGAACCGTTATGGGACTATCA
 AGTGAACAAGACATTGAAAAACACTATGAAGAGCTTAAGAACAAGTTACATGATATGTACAATCATTATGG
 TGGCGGTGGCGGATGCTAGAGTGTAGGCGTATCTCACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAA
 CGAAGGCGAGTAATACCTACGACGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATT
 GATGCAAGTTGAAAAAATCTCTCAGTCAACAAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAAAC
 TACTGCTGAATCAACCACTTAAAGAGCAGCAAGATAATGAACAAAAAGCTTTACCTCTGCACAAGAAA
 TTTACACTAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTA
 ACAGCTACTGAACACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAGAGAGCTGCATTGTCAGAAACA
 AAAAGCTAGCATTTTCAGCAGAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAAACGCTCTGAACAAA
 ATATTGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGAT
 AATACAAAAGCATTAAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAAATCAAAAAGCTAAAGTTAA
 AAAGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGCCAGAACTTAGTCGCT
 TTAATCCTCAGCTCCGCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAGCACCAGGCAAGGCTAT
 CCTCTTGAAGAAGCTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACAA
 AGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGcggcgcac
 tccgagCACCACCACCAACCACCAC

SEQ ID NO: 135

M A F N T S Q S V S A Q V Y S N E G Y H Q H L T D E K S H L Q Y S K D N
 A Q L Q L R N I L D G Y Q N D L G R H Y S S Y Y Y Y N L R T V M G L S S
 E Q D I E K H Y E E L K N K L H D M Y N H Y A S G G G S M S V G V S H Q
 V K A D D R A S G E T K A S N T H D D S L P K P E T I Q E A K A T I D A
 V E K T L S Q Q K A E L T E L A T A L T K T T A E I N H L K E Q Q D N E
 Q K A L T S A Q E I Y T N T L A S S E E T L L A Q G A E H Q R E L T A T
 E T E L H N A Q A D Q H S K E T A L S E Q K A S I S A E T T R A Q D L V
 E Q V K T S E Q N I A K L N A M I S N P D A I T K A A Q T A N D N T K A
 L S S E L E K A K A D L E N Q K A K V K K Q L T E E L A A Q K A A L A E
 K E A E L S R L K S S A P S T Q D S I V G N N T M K A P Q G Y P L E E L
 K K L E A S G Y I G S A S Y N N Y Y K E H A D Q I I A K A S P G N Q L N
 Q Y Q A A A L E H H H H H H

SEQ ID NO: 136

AGTTGGTA

REIVINDICAÇÕES

1. Composição imunogénica compreendendo uma combinação de antigénios GAS, consistindo a referida combinação de 2-10 antigénios GAS, na qual a referida combinação inclui GAS40, sendo o dito "GAS40" uma proteína que compreende uma sequência de aminoácidos (a) que tem 95% ou mais de identidade para SEQ ID N°: 1 e/ou (b) que é um fragmento de pelo menos 200 aminoácidos consecutivos da SEQ ID N°: 1 e em que a referida combinação inclui GAS57, sendo o dito "GAS57" uma proteína que compreende uma sequência de aminoácidos (a) tendo 95% ou mais de identidade com a SEQ ID N°: 116 e/ou (b) que é um fragmento de pelo menos 200 aminoácidos consecutivos da SEQ ID N°: 116.
2. Composição da reivindicação 1, em que o referido antigénio GAS40 compreende uma sequência de aminoácidos compreendendo uma primeira região em espiral enrolada, e uma segunda região em espiral enrolada.
3. Composição de acordo com a reivindicação 1, em que o antigénio GAS40 compreende uma sequência de aminoácidos que compreende uma primeira região de espiral enrolada.
4. Composição da reivindicação 3, em que a referida primeira região de espiral enrolada compreende uma sequência de aminoácidos que compreende a SEQ ID N°: 12.
5. Composição da reivindicação 2, em que a referida segunda região em espiral enrolada inclui uma região de fecho de leucina.
6. Composição da reivindicação 5, em que a segunda região em espiral enrolada, compreende uma sequência de aminoácidos que compreende a SEQ ID N°: 13.

7. Composição de qualquer das reivindicações anteriores para utilização em terapia.
8. Processo de fabrico de qualquer uma das composições imunogénicas das reivindicações 1 a 6.
9. Kit compreendendo um primeiro componente que compreende qualquer uma das composições imunogénicas das reivindicações 1 a 6.

FIGURE 1: Annotation of GAS 40

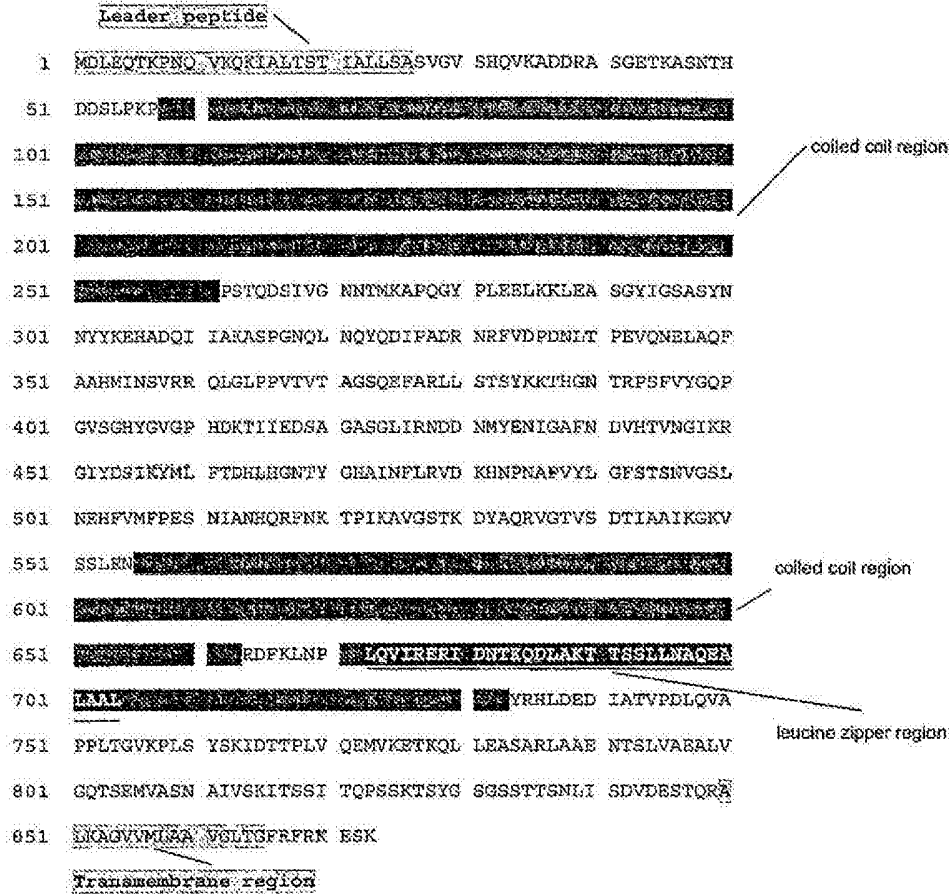


FIGURE 2 : Schematic of GAS40: putative surface exclusion protein prgA (873aa)

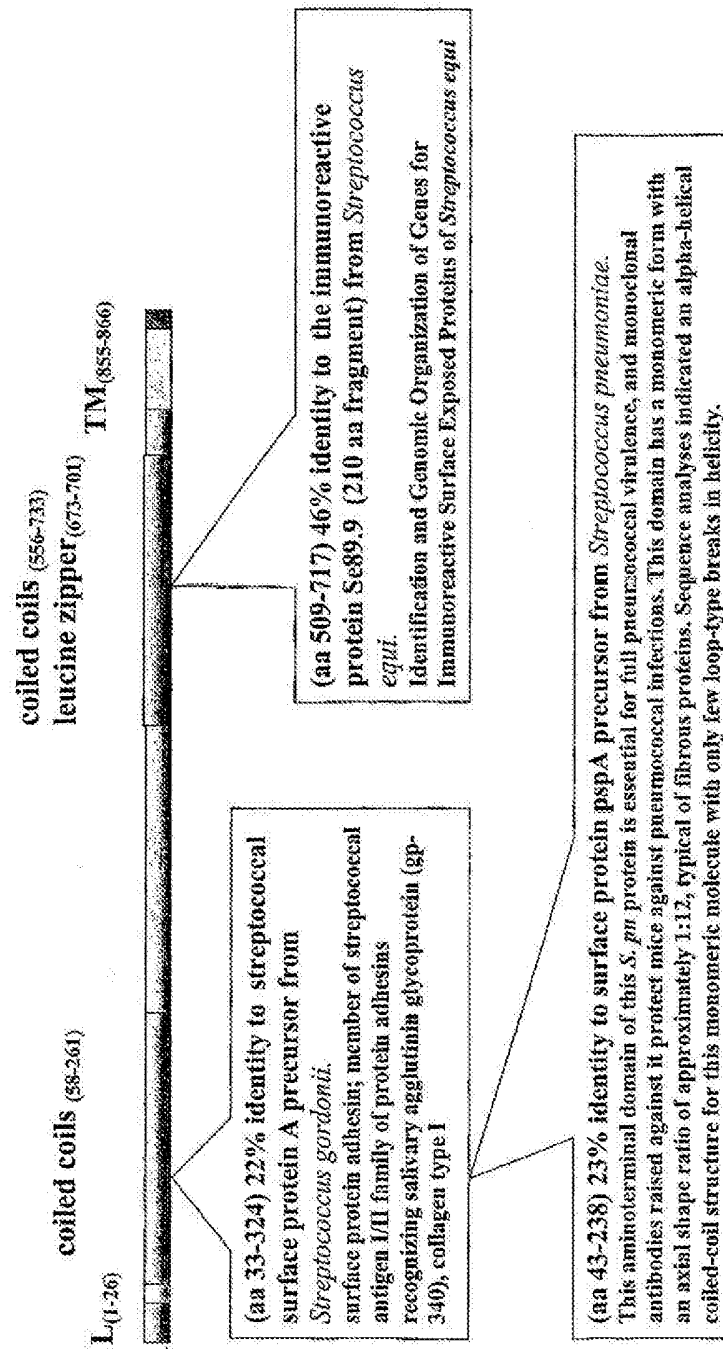


FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other *Streptococcus* bacteria

3(a) BLAST alignment of amino acid sequence of GAS 40 including the first coiled-coil region with SpA precursor of *Streptococcus gordonii*

```
>gi|25990270|gb|AAC44101.3| streptococcal surface protein A precursor
[Streptococcus gordonii]
    Length = 1575
>ref|NP_268623.1| putative surface exclusion protein [Streptococcus pyogenes]
    Length = 873

Score = 63.2 bits (152), Expect = 5e-11
Identities = 65/293 (22%), Positives = 124/293 (42%), Gaps = 13/293
(4%)

Query: 112 QDQTSDEKGTATTAAENAQKQAEIKSDYAKQA---EEIKKTTEAYKKEVEARQAETDKIN
167
      Q + D + + T A N   + K + ++A   + ++KT   K E+       K
Sbjct: 33 QVKADDRASGETKASNTTHDDSLPKPETIQEAKATIDAVEKTLSSQKAEITELATALTTKTT
92

Query: 168 AENKAAEDKYQEDLKAHQAEVEKINTANATAKAEEYAKLAQYQKDLAAVQKANEDSQLDY
227
      AE   +++   + KA + E       A+++   A+ A++Q++L A +   ++Q D
Sbjct: 93 AEINHLKEQQDNEQKALTSAQRIYTNTLASSEETLLAQGAHQRELTATETELHNAQADQ
152

Query: 228 QNKLSAYQAEELARVQKANAEAKEAYE--KAVKENTAKNAALQAENEAIKQRNETAKANYD
285
      +K +A   + A +       A++ E K ++N AK A+ + +AI + +TA N
Sbjct: 153 HSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDALTAAQTANDNTK
212

Query: 286 AAMKQYEADLAAIKKAKEDNDADYQAKLAAYQAEELARVQKANADAKAAYEKAVEENTAKN
345
      A   + E   A ++ K       +LAA +A LA +   + K++       +   N
Sbjct: 213 ALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLLKSSAPSTQDSIVGNN
272

Query: 346 TAIQAEN---EAIKQRNAA---AKATYERALKQYEADLAAAKKANEDSDADYQ 392
      T   +   E +K+ A+       A+Y   K++ AD AK + +   YQ
Sbjct: 273 TMKAPQGYPLEELKKLEASGYIGSASYNYYKEH-ADQIIAKASPGNQLNQYQ 324
```

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other *Streptococcus* bacteria

3(b) BLAST alignment of amino acid sequence of GAS 40 including the first coiled-coil region with SpB precursor of *Streptococcus gordonii*

```
>gi|25055226|gb|AAC44102.3| streptococcal surface protein B precursor
[Streptococcus gordonii]
      Length = 1499
>ref|NP_268623.1| putative surface exclusion protein [Streptococcus
pyogenes]
      Length = 873
```

```
Score = 54.3 bits (129), Expect = 2e-08
Identities = 53/226 (23%), Positives = 98/226 (43%), Gaps = 13/226
(5%)
```

```
Query: 111 QDQTSDEKGTATTAAENAAQKQAEIKSDYAKQA---EEIKKTTEAYKKEVEAHQAETDKIN
166
      Q + D+ + T A N + K + ++A + ++KT K E+ K
Sbjct: 33 QVKADDRASGETKASNTDSDLPKPEITQEAATIDAVEKTLSSQKAEITELATALTCTT
92
```

```
Query: 167 AENKAAEDKYQEDLKAHQAEVEKINTANATAKAIEYEAQLAQYQKDLAAVQKANEDSQLDY
226
      AE +++ + KA + E A+++ A+ A++Q++L A + ++Q D
Sbjct: 93 AEINHLKEQQDNEQKALTSAGEIYTNLTASSEETLLAQGAHQRELATETELHNAQADQ
152
```

```
Query: 227 QNKL SAYQAELARV--QXXXXXXXXXXXXXXXXXNTAKNAALQAENEAIKQNETAKANYD
284
      +K +A + A + + N AK A+ + +AI + +TA N
Sbjct: 153 HSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTK
212
```

```
Query: 285 AAMKQYE---ADL----AAIKKAKEDNDADYQAKLAAYQAELARVQ 323
      A + E ADL A +KK + A +A LA +AEL+R++
Sbjct: 213 ALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLEK 258
```

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other Streptococcus bacteria

3(c) BLAST alignment of amino acid sequence of GAS 40 including the first coiled-coil region with Surface Protein PspA precursor of *Streptococcus pneumoniae*

```
>gi|282335|pir||A41971 surface protein pspA precursor - Streptococcus
pneumoniae
>ref|NP_268623.1| putative surface exclusion protein [Streptococcus
pyogenes]
      Length = 873

Score = 48.1 bits (113), Expect = 6e-07
Identities = 46/200 (23%), Positives = 89/200 (44%), Gaps = 23/200
(11%)

Query: 139 KTKFNTVRAMVVPEPEQLAETK-----KKSEEAKQKAPELTKKLEEAkakLEE-AEKK
190
      +TK +      +P+PE + E K      K + K + EL L + A++ E++
Sbjct: 43 ETKASNTHDDSLPKPETIQEAKATIDAVEKTLsqQKAELTELATALTKTTAEINHLKBQQ
102

Query: 191 ATEAKQKVDAAEEVAPQAKIAELENQVHRLEQELKEIDeseSEDYAKEGFRAPLQSKLDAK
250
      E K A+E+      + E + + + +E+ +E+E + + + ++ L +
Sbjct: 103 DNEQKALtsAQEIYNTLASSEETLLAQGAHQRELtATETELHNAQADQHskETALSEQ
162

Query: 251 KAKLS----KLEELSDKIDELDAEIAKLEDQL-----KAAEENNNVEDYfKEGLEKTI
299
      KA +S + ++L +++ + IAKL + KAA+ N+ LEK
Sbjct: 163 KASISAEtTRAQDLVEQVKTSEQNIakLNAMISNFDaitKAAQTANDNTKALSSELEKA-
221

Query: 300 AAKKAELEKTEADLKKAVNE 319
      KA+LE +A +KK + E
Sbjct: 222 ---KADLENQKAKVKKQLTE 238
```

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other Streptococcus bacteria

3(c) BLAST alignment of amino acid sequence of GAS 40 including the first coiled-coil region with Surface Protein PspA precursor of *Streptococcus pneumoniae*

```
>gi|282335|pir|A41971 surface protein pspA precursor - Streptococcus
pneumoniae
>ref|NP_268623.1| putative surface exclusion protein [Streptococcus
pyogenes]
      Length = 873

Score = 48.1 bits (113), Expect = 6e-07
Identities = 46/200 (23%), Positives = 89/200 (44%), Gaps = 23/200
(11%)

Query: 139 KTKFNTVRAMVVPPEBQLAETK-----KKSEEAKQKAPELTKKLEEAkakLEE-AEKK
190
      +TK +      +P+PE + E K      K + K + EL L + A++ E++
Sbjct: 43 ETKASNTHDDSLPKPETIQEAKATIDAVEKTLsqQKAELTELATALTKTTAEINHLKBQQ
102

Query: 191 ATEAKQKVDAAEEVAPQAKIAELENQVHRLEQELKEIDEESEEDYAKEGFRAPLQSKLDAK
250
      E K A+E+ + E + + + +E+ +E+E + + + ++ L +
Sbjct: 103 DNEQKALTSAQEIYNTLASSEETLLAQGAHQRELtATETELHNAQADQHskETALSEQ
162

Query: 251 KAKLS----KLEELSDKIDELDAEIAKLEDQL-----KAAEENNNVEDYfKEGLEKTI
299
      KA +S + ++L +++ + IAKL + KAA+ N+ LEK
Sbjct: 163 KASISAETTRAQDLVEQVKTSEQNIakLNAMISNFDaitKAAQTANDNTKALSSELEKA-
221

Query: 300 AAKKAELEKTEADLKKAVNE 319
      KA+LE +A +KK + E
Sbjct: 222 ---KADLENQKAKVKKQLTE 238
```

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other *Streptococcus* bacteria

3(d) BLAST alignment of amino acid sequence of GAS 40 including the second coiled-coil region with SpB precursor of *Streptococcus gordonii*

```
>gi|23380384|gb|AAN18299.1| immunoreactive protein Se89.9 (fragment)
[Streptococcus equi]
      Length = 210
>ref|NP_268623.1| putative surface exclusion protein [Streptococcus
pyogenes]
      Length = 873

Score = 173 bits (438), Expect = 4e-45
Identities = 98/209 (46%), Positives = 144/209 (68%)

Query: 1   ESDIVDTRFSTTEIPKSGQVIDRSASIQALTNDIASIKGKIASLESRLADPSSEAEVTA
60
      ES+I + RF+ T I   G   D + + +++ IA+IKGK++SLE+RL+   EA++ A
Sbjct: 509 ESNIANHQRFNKTPIKAVGSTKDYAQEVGTVSDTIAAIKGKVSSLENRLSAIHQEADIMA
568

Query: 61   AQAKISQLQHLEAAQAKSHKLDQQVEQLANTKDSLRTQLLAAKEEQAQLKANLDEKALAL
120
      AQAK+SQLQ +L +   +S L+ QV QL +TK SLRT+LLAAK +QAQL+A D++LA
Sbjct: 569 AQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAK
628

Query: 121  LASSKATLHKLEAAMEEAKARVAGLASQKAQLEDLLAFKKNPNRIELAQEKVAAAKKALA
180
      LAS KA LH+ EA E+A ARV L ++KA L+ L F+ NPNR+++ +E++ K+ LA
Sbjct: 629 LASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLA
688

Query: 181  DTEDKLLAAQASLSDLQAQRARLQLSIAT 209
      T LL AQ +L+ LQA+++ L+ +IAT
Sbjct: 689 KTTSSLLNAQEALAAALQAKQSSLEATIAT 717
```

Figure 4: Secondary Structure Prediction of GAS 40

Figure 4(a) Secondary Structure prediction alignment with GAS 40 amino acid sequence

```

      10      20      30      40      50      60      70
MDLEQTKPNQVKOKIALTSTIALLSASVGVSHQVKADDRASGETKASNTHTDSSLPKPETIQEAKATIDAV
CCCCCCCCcchhhhhhhhhhhhhhhhhhhhhcceeEEEEcCcCCCCCcCCCCCCCCCCCCcchhhhhhhhhhh
EKTLSQQAELTELATALTKTAEINHLKEQQDNEQKALTSAQEIYTNLTASSEETLLAQGAHQRELTATA
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
TETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEONIAKLNAMISNPDAITKAAQTANDN
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
TKALSSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRKSSAPSTQDSIVGNNTMKAPQGY
cHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
PLEELKKLEASGYIGSASYNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPDNLTPVQNELAQF
CHHHHHHHHHcCcccccchHHHHHHHHHHHHHHHHHHHHcCchhhhhhhcCcccccCCCCCCCCcchHHHHHHHH
AAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTGNTRPSPFVYGQPGVSGHYGVGPHDKTIIEDSA
HHHHHHHHHHHHcCCCCcchHHHHHHHHHHHHHHHHHHHHcCCCCcEEECcCCCCcEEEEcC
GASGLIRNDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLGNTYGHAINFLRVDKHNPAPVYL
CCCCcchHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
GFSTSNVGSLSLNEHFVMPESNIAHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGVSSLENRLSAI
EEEEcCcCCCCcchHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
HQEADIMAAQAKVSQLQGLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLA
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
SLKAAHQTEALAEQAARVLTALVAKKANLQYLRFKLPNRLQVIRERIDNTKQDLAKTTSSLLNAQEA
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
LAALQAKQSSLEATIATTEHQTLTKTLANEKEYRHLDEDIATVPDLQVAPPLTGKPLSYSKIDTTPLV
HHHHHHHHcCcccccchHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
QEMVKETQQLLEASARLAAENTSLVAEALVGQTSEMVASNAIVSKITSSITQPSKTSYSGSGSSTSNLI
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
SDVDESTQRLKAGVVMIAAVGLTGFRFRKESK
cCChHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

```

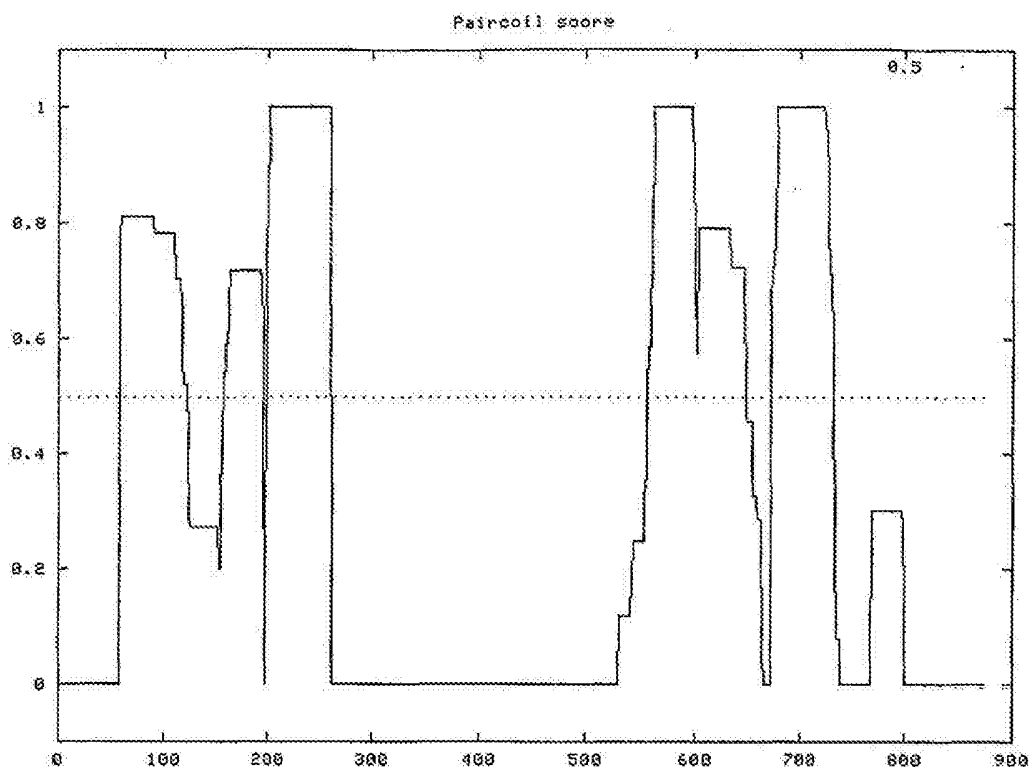
Sequence length : 873

PHD :

Alpha helix	{Hh} :	525 is	60.14%
β_{10} helix	{Gg} :	0 is	0.00%
Pi helix	{Ii} :	0 is	0.00%
Beta bridge	{Bh} :	0 is	0.00%
Extended strand	{Ee} :	63 is	7.22%
Beta turn	{Tt} :	0 is	0.00%
Bend region	{Ss} :	0 is	0.00%
Random coil	{Cc} :	285 is	32.65%
Ambiguous states (?)	:	0 is	0.00%



Figure 4(b): Secondary Structure prediction based on PairCoil Score



Coiled coils found:

positions	58- 121	initial register	'f'	probability	0.811
positions	156- 196	initial register	'e'	probability	0.720
positions	198- 246	initial register	'e'	probability	1.000
positions	247- 261	initial register	'a'	probability	1.000
positions	556- 646	initial register	'e'	probability	1.000
positions	671- 733	initial register	'f'	probability	1.000

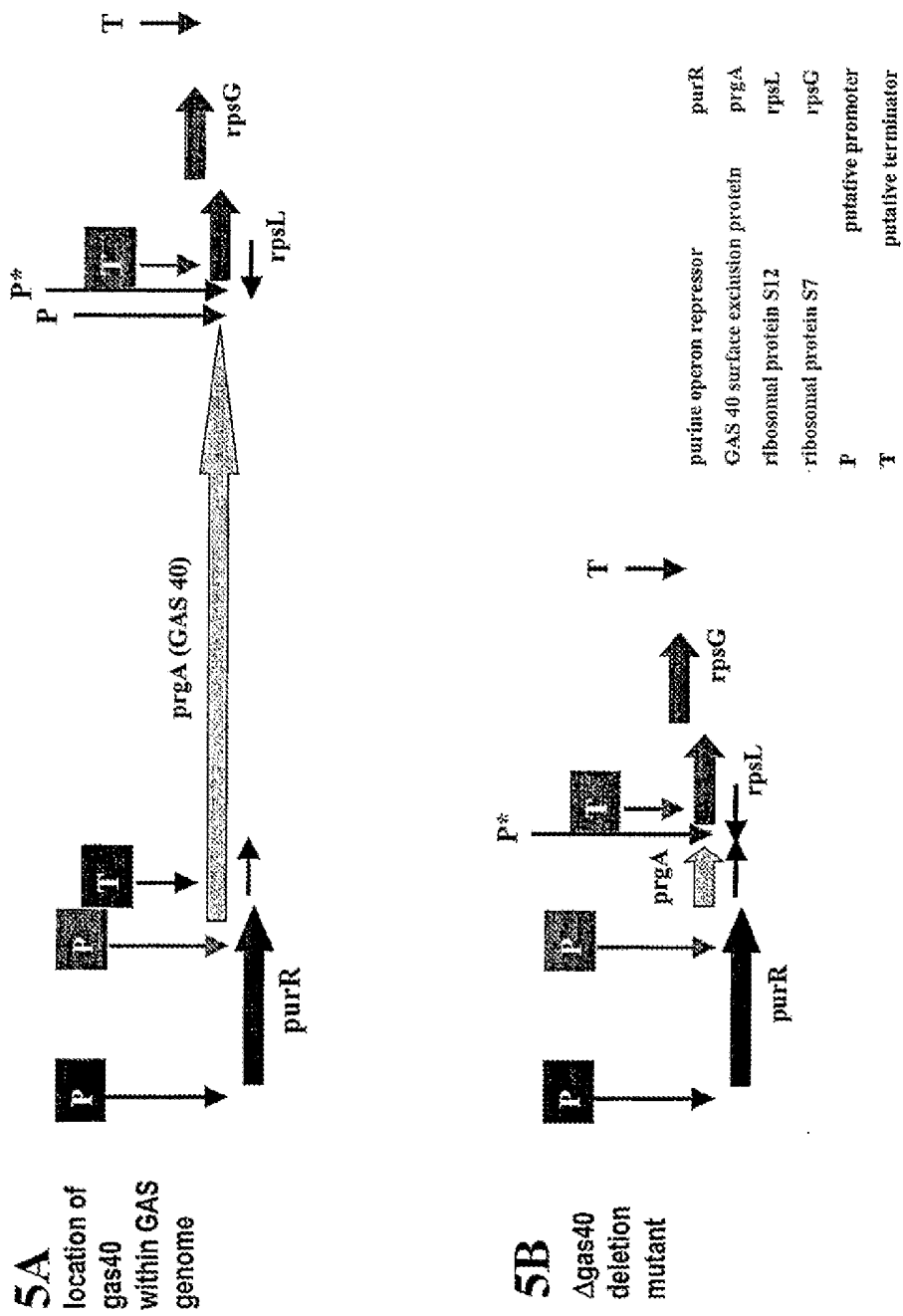
Figure 4(c): Secondary Structure prediction of Leucine Zipper within coiled coil.

```

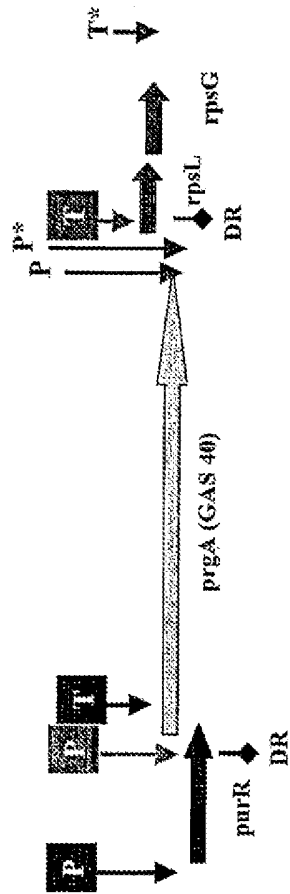
673                               701
QYLRFKLNPNRLQVIRERIDNTKQDLAKTTSSLNAQEALALQAKQSSLEATTATTEH
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
L-----I-----L-----L-----L
OOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOO

```

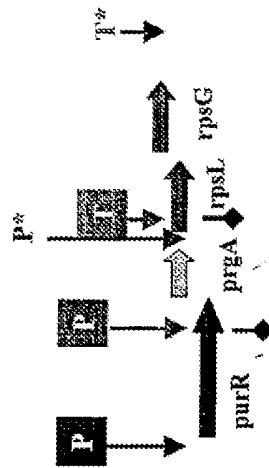
Figure 5



5C
direct repeats
surrounding GAS 40
within wild type genome



5D
location of direct
repeats within Δ gas40



5E
corresponding
genomic region
within GBS

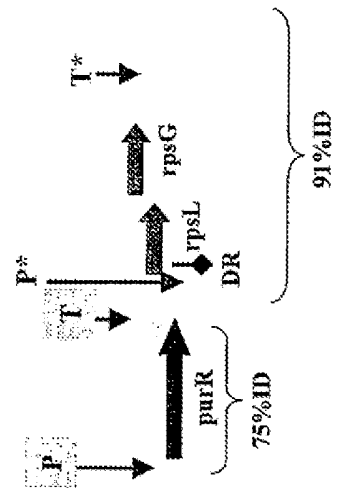


FIGURE 6: FACS Comparison of GAS 40 in wild type GAS and GAS 40 deletion mutant

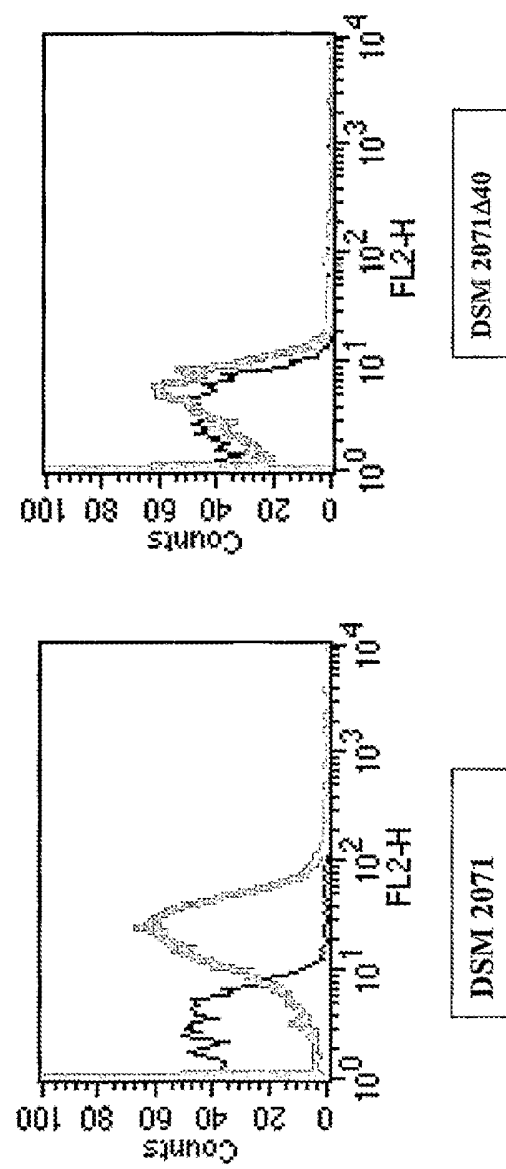
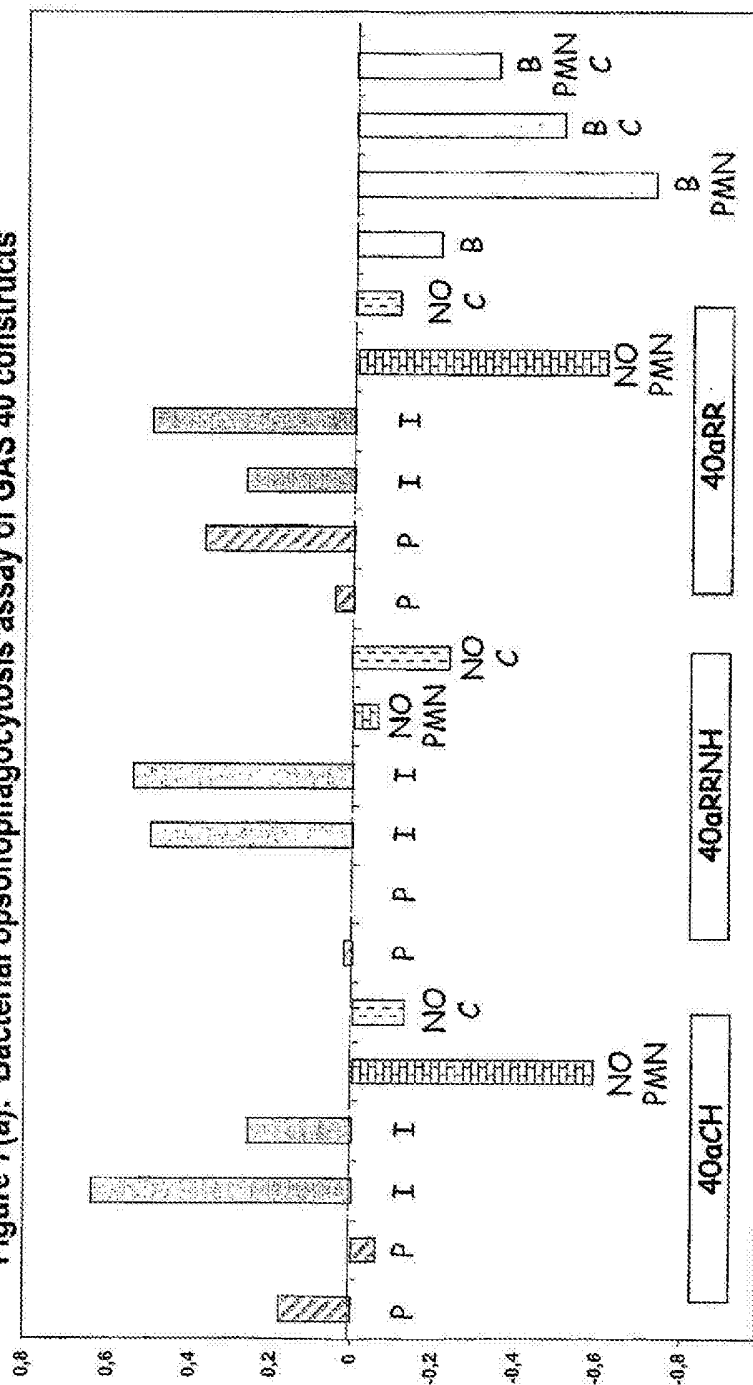


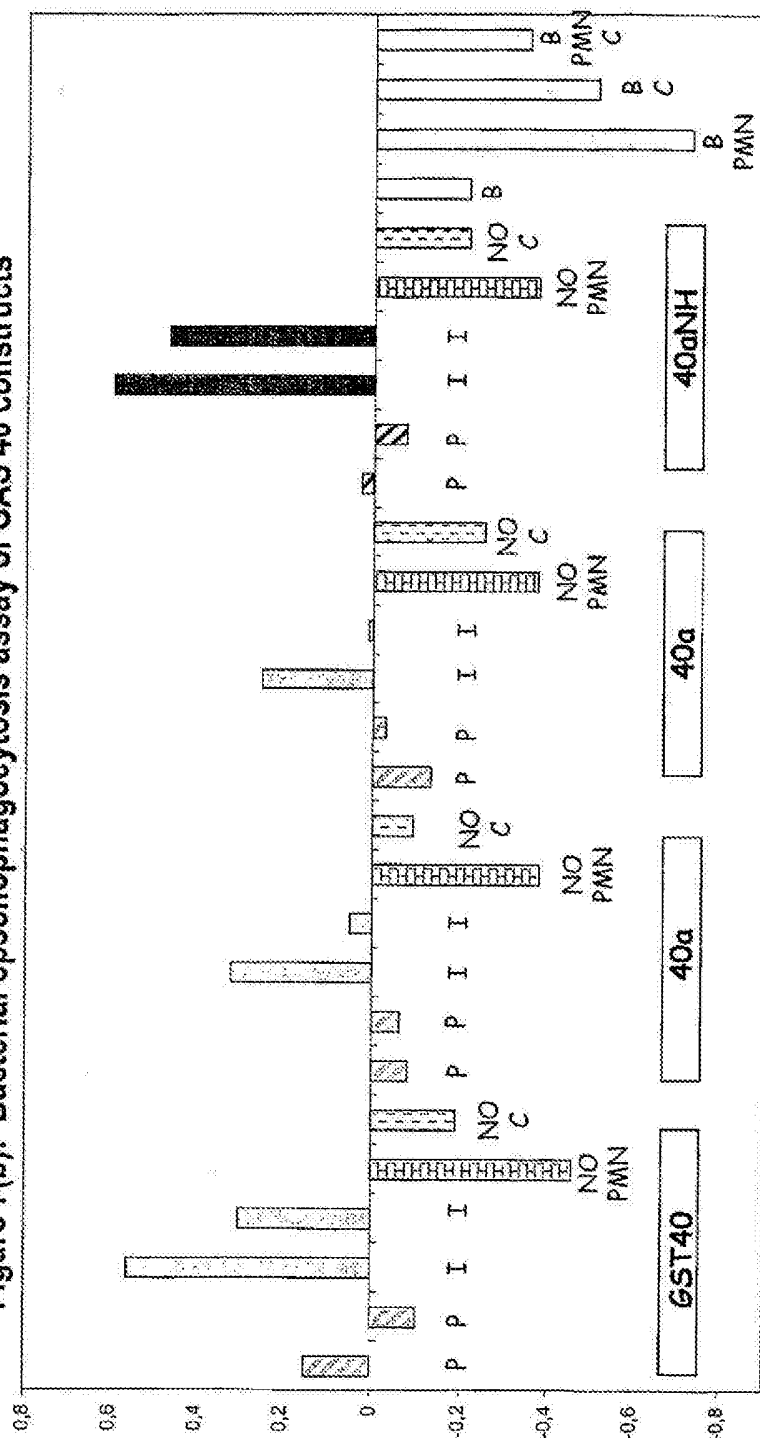
Figure 7(a): Bacterial opsonophagocytosis assay of GAS 40 constructs



B - bacteria
P - pre-immune serum
I - immune serum
No PMN - without polymorphonucleates
No C - without complement

Histogram bars represent the difference between logarithm at T 0 (initial time) and T60 (bacterial CFU counted after 60 minutes of incubation)

Figure 7(b): Bacterial opsonophagocytosis assay of GAS 40 constructs



B - bacteria
P - preimmune serum
I - immune serum
NO - without polymorphonucleates
C - without complement

Histogram bars represent the difference between logarithm at T0 (initial time) and T60 (bacterial CFU counted after 60 minutes of incubation)

Figure 8: Immunization in Murine Mouse Model

GAS antigen	Survival/Tested mice			Protection	pValue	Protein Purity
	alive	dead	tested			
gst 40	67	63	130	51	0.00012	%
253	14	36	50	28	0.008	18
253-urea	2	8	10	20		23
263-gst	2	8	10	20		36
39	9	31	40	22.5	0.09	20
38a	13	37	50	26	0.016	16
39a	10	30	40	25	0.039	
38a	12	28	40	30	0.0046	
urea 366	21	78	99	21.2	0.046	65
117	19	51	70	27	0.0036	18
117-urea	1	9	10	10		89
117-urea-2M	7	23	30	23.3	0.1	86
117-urea-2M (prep 117)	8	32	40	20	0.2	
urea 504	9	31	40	22.5	0.09	50
504	14	26	40	35	0.0003	40
504	7	33	40	17.5	0.4	80
urea 589	7	23	30	23	0.1	36
533	14	36	70	20	6.12	80
new 533	4	16	20	20	6.34	30
gst 57	12	48	60	20	8.14	60
57a	0	20	20	0		50
294	17	73	90	18.8	6.14	80
130	15	65	80	18.7	0.17	40
130	7	23	30	23.3	0.1	40
84	8	32	40	20	0.2	70
urea 159	7	33	40	17.5	0.4	5
159a	2	8	10	20		63
527	10	40	50	20	6.17	50
527	3	17	20	15		86
217	7	33	40	17.5	0.4	50
511	13	27	40	16.2	0.41	80
277	8	42	50	16	0.52	5
277a	2	28	30	6.5		50
gst 202	3	17	20	10	0.73	5
202a	5	25	30	16.6	0.53	5
45	5	25	30	16.6	0.53	80
urea 309	5	25	30	20	0.53	8
290	6	34	40	15	0.67	50
529	6	34	40	15	0.67	5
gst 59	10	60	70	14.2	0.71	30
384	7	43	50	14	0.78	80
384RR	1	19	20	5		80
urea 509	7	53	60	11.6	0.84	50
509-NH	2	8	10			75
509-CH	0	10	10			75
163	7	53	60	11.6	0.84	65
urea 372	4	26	30	13.7	0.85	20
gst 42	4	26	30	13.3	0.9	50
95	5	35	40	12.5	1	55
urea 235	5	35	40	12.5	1	80
new 236	2	8	10	20		70
137	5	35	40	12.5	1	75
His-Stop	29	201	230	12.06		