(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(10) International Publication Number
WO 2015/169773 A2

(43) International Publication Date
12 November 2015 (12.11.2015)

(51) International Patent Classification: *A61K 38/00* (2006.01)

(21) International Application Number: PCT/EP2015/059772

(22) International Filing Date: 5 May 2015 (05.05.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 14167445.7 7 May 2014 (07.05.2014) EP

(71) Applicant: **GLAXOSMITHKLINE BIOLOGICALS S.A.** [BE/BE]; rue de l'Institut 89, B-1330 Rixensart (BE).

(72) Inventors: **D'AURIZIO, Romina**; GlaxoSmithKline, Via Fiorentina 1, I-53100 Siena (IT). **GALLOTTA, Marilena**; GlaxoSmithKline, Via Fiorentina 1, I-53100 Siena (IT). **GRANDI, Guido**; GlaxoSmithKline, Via Fiorentina 1, I-53100 Siena (IT). **MARGARIT Y ROS, Immaculada**; GlaxoSmithKline, Via Fiorentina 1, I-53100 Siena (IT). **MORA, Mariarosa**; GlaxoSmithKline, Via Fiorentina 1, I-53100 Siena (IT).

(74) Agent: **THORNLEY, Rachel M.**; GlaxoSmithKline, Global Patents (CN925.1), 980 Great West Road, Brentford Middlesex TW8 9GS (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:
- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
- *with sequence listing part of description (Rule 5.2(a))*

WO 2015/169773 A2

(54) **Title**: MUTANTS OF SPY0269

(57) **Abstract**: The invention relates to to mutated antigens derived from *Streptococcus pyogenes* spy0269 and their use in immunogenic compositions and vaccine compositions.

MUTANTS OF SPY0269

TECHNICAL FIELD

The invention relates to the fields of immunology and vaccinology. In particular, it relates to mutated antigens derived from *Streptococcus pyogenes* spy0269 and their use in immunogenic compositions and vaccine compositions.

BACKGROUND ART

Group A streptococcus ("GAS", *S. pyogenes*) is a gram positive, non-spore forming coccus shaped bacteria that typically occurs in chains or in pairs of cells. GAS is one of the most frequent pathogens of humans. It is estimated that between 5 and 15% of normal individuals harbour the bacterium, usually in the respiratory tract, without showing any signs of disease. When host defences are compromised, or when the organism is able to exert its virulence, or when it is introduced to vulnerable tissues or hosts, however, an acute infection occurs. Acute *Streptococcus pyogenes* infections may take the form of pharyngitis, scarlet fever (rash), impetigo, cellulitis, or erysipelas. Invasive infections can result in necrotizing fasciitis, myositis and streptococcal toxic shock syndrome. Patients may also develop immune-mediated sequelae such as acute rheumatic fever and acute glomerulonephritis (Patterson 1996). Although *S. pyogenes* may be treated using antibiotics, a prophylactic vaccine to prevent the onset of disease is desired. Efforts to develop such a vaccine have been ongoing for many decades. While various GAS vaccine approaches have been suggested and some approaches are currently in clinical trials, to date, there are no GAS vaccines available to the public.

GAS40, also known as "Spy0269" (M1), "SpyM3_0197" (M3), "SpyM18_0256" (M18) and "prgA" (Uniprot Accession No. Q9A1H3), is a putative surface exclusion protein that possesses a 130-amino-acid region showing similarity with the EzrA protein, which interacts with the cell division protein FtsZ. GAS40 consistently provides protection in the animal model of systemic immunization and challenge and induction of bactericidal antibodies. Hence, GAS40 is a potential candidate for the production of a vaccine against GAS. However, there have been some concerns in the past that GAS vaccines based on the M protein contain antigenic determinants that have been determined to cross-rea with heart tissue, potentially resulting in a reaction important in rheumatic heart disease.

Molecular mimicry is defined as the theoretical possibility that sequence similarities between foreign and self-peptides are sufficient to result in the cross-activation of autoreactive T or B cells by pathogen-derived peptides. Upon the activation of B or T cells, it is believed that these "peptide mimic" specific T or B cells can cross-react with self-epitopes, thus leading to tissue pathology (autoimmunity). Molecular mimicry is a phenomenon that has been recently discovered as one of

several ways in which autoimmunity can be evoked. Therefore, vaccine candidates must be selected that will not elicit the disease it is intended to prevent.

Whilst there is no evidence that protein antigens such as GAS40 comprise antigenic determinants that might be problematic, historical fears concerning molecular mimicry and GAS vaccines *per se* may hinder the uptake and use of GAS vaccines.

In order to reduce these fears and the possibility that GAS40 might produce autoimmune sequelae of GAS infection by molecular mimicry, the inventors have produced GAS40 polypeptides which are less similar to human proteins than wild type GAS40 proteins.

SUMMARY OF THE INVENTION

The Inventors have developed mutants of GAS40 polypeptides that elicit antibodies that are cross-reactive with wild-type GAS40 but which have reduced sequence identity to amino acid sequences present within human proteins. Particularly, the human proteins comprise coiled-coil domain containing 81 protein, UPF0492 protein C20orf94, janus kinase and microtubule-interacting protein 3, myosin or tropomyosin.

Thus, in a first aspect there is provided a purified recombinant streptococcal GAS40 antigen which comprises an epitope that elicits opsonic antibodies to Group A Streptococcus, the recombinant streptococcal GAS40 antigen comprising at least two substitutions in any of amino acids 235 to 243, and/or at least two substitutions in any of amino acids 684 to 692 and/or at least two substitutions in any of amino acids 699 to 706 and/or a deletion of Lysine at position 208 and/or at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 amino acid substitutions at positions selected from the group consisting of amino acid positions 38, 52, 66, 148, 158, 179, 182, 186, 196, 203, 218, 225, 544, 558, 562, 576, 586, 614, 618, 621, 665, 672, 683 and 693, wherein the amino acid positions are numbered according to SEQ ID NO: 2.

In some embodiments the purified recombinant streptococcal GAS40 antigen comprises substitutions at amino acid positions T237, A241, A688, S692, A700 and A703. For example, the purified recombinant streptococcal GAS40 antigen comprises the substitutions T237S, A241V, A688S, S692T, A700S and A703S. The recombinant streptococcal GAS40 polypeptide may comprise a deletion of Lysine at position 208. The recombinant streptococcal GAS40 polypeptide may comprise amino acid substitutions at amino acid positions A38, Q52, T66, A148, S158, A179, A182, T186, A196, Q203, Q218, K225, A544, T558, S562, K576, A586, T614, A618, A621, T665, A672, Q683 and T693. For example, the substitutions at amino acid positions may be A38L, Q52L, T66L, A148L, S158L, A179L, A182L, T186L, A196L, Q203L, Q218L, K225L, A544L, T558L, S562L, K576L, A586L, T614L, A618L, A621L, T665L, A672L, Q683L and T693L. The recombinant streptococcal GAS40 polypeptide may comprise a combination of these mutations, for example, the

substitutions T237S, A241V, A688S, S692T, A700S, A703S, A38L, Q52L, T66L, A148L, S158L, A179L, A182L, T186L, A196L, Q203L, Q218L, K225L, A544L, T558L, S562L, K576L, A586L, T614L, A618L, A621L, T665L, A672L, Q683L, T693L and a deletion of L208.

5 An aim of the present invention is to reduce any theoretical risk of serological cross-reactivity of GAS40 antigens with mammalian tissue antigens. Particularly, the mammal is a human. Particularly, the recombinant streptococcal GAS40 antigen does not contain 7-mers, 8-mers or 9-mers that are present in human proteins. Yet more particularly, the GAS40 antigen does not contain a sequence selected from the group consisting of QLTEELAAQ (SEQ ID NO: 8), KQDLAKTTS (SEQ ID NO: 10 9) or EALAALQA (SEQ ID NO: 10). Still yet more particularly, the recombinant streptococcal GAS40 polypeptide does not comprise an amino acid sequence having 100% sequence identity with a sequence selected from the group consisting of QLTEELAAQ (SEQ ID NO: 8), KQDLAKTTS (SEQ ID NO: 9) or EALAALQA (SEQ ID NO: 10). Yet still more particularly, the 7-, 8- or 9-mer amino acid sequences of the recombinant streptococcal GAS40 polypeptide from positions 235 to 15 243 and/or positions 684 to 692 and/or 699 to 707, numbered according to SEQ ID No: 2, have sequence identity of less than 100%, less than 95%, less than 90%, less than 85%, less than 80%, less than 79%, less than 78%, less than 77%, less than 76%, less than 75%, less than 74%, less than 73%, less than 72%, less than 71%, less than 70%, less than 65%, less than 60%, less than 55%, less than 50%, less than 45%, less than 40% or less than 35% with a sequence selected from the group 20 consisting of QLTEELAAQ (SEQ ID NO: 8), KQDLAKTTS (SEQ ID NO: 9) or EALAALQA (SEQ ID NO: 10).

Particularly, the recombinant streptococcal GAS40 polypeptide may share at least 92% identity with any one of SEQ ID NOs: 4, 6, 7, 16, 17, 18, 24, 25 or 26. More particularly, the recombinant streptococcal GAS40 polypeptide comprises a sequence selected from the group consisting of SEQ 25 ID NOs: 4, 6, 7, 16, 17, 18, 24, 25 and 26.

In some embodiments, there is provided a fragment of a recombinant streptococcal GAS40 polypeptide according to the first aspect. For example, a fragment that shares at least 97% identity with any one of SEQ ID NOs: 20, 21, 22, 28, 29, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 44, 45, 46, 48, 49, 50, 52, 53, 54, 56, 57, 58, 60, 61, 62, 64, 65, 66, 67, 68, 69, 70, 72, 73, 74, 76, 77, 78, 80, 81, 82, 30 84, 85, 86, 88, 89, 90, 92, 93, 94, 96, 97, 98, 100, 101, 102, 104, 105, 106, 108, 109, 110, 112, 113, 114, 116, 117 and 118. Other examples of fragments include polypeptide sequences comprising or consisting of any one of SEQ ID NOs: 4, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80, 84, 88, 92, 96, 100, 104, 108, 112, 116, 7, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114 and 118. The term "fragment" as used herein refers to 35 a polypeptide that has an amino-terminal and/or carboxy-terminal deletion as compared to the full length protein, but wherein the remaining amino acid sequence is identical to the corresponding

positions of the full length protein. Fragments typically are greater than 200 amino acids long, particularly no more than 850, 849, 848, 847, 846, 845 amino acids long, yet more particularly no more than 830, 827, 826, 825, 824, 823, 822, 822, 821, 820 amino acids long, still yet more particularly no more than 460, 459, 458, 457, 456, 455, 454, 453, 452, 451, 450 amino acids long, including ranges between these values. For example, from 200 to 847 amino acids long, from 200 to 823 amino acids long, from 200 to 452 amino acids long, from 450 to 850 amino acids long, from 450 to 825 amino acids long. For the avoidance of doubt, fragments of the present invention (numbered according to SEQ ID NO: 2) will comprise substitutions corresponding to those described above where those amino acid positions are present in the fragment.

In a second aspect of the invention, there is provided an immunogenic composition comprising one or more GAS40 polypeptides of the first aspect. Generally such immunogenic compositions will comprise one or more additional GAS polypeptides such as, by way of non-limiting example, GAS57, GAS25 or mutants thereof. Particularly immunogenic compositions of the invention will comprise an immunologically effective amount of the GAS40 polypeptide(s). Yet more particularly, immunogenic compositions of the invention are vaccines. Such compositions may comprise an adjuvant such as alum or MF59. Particular immunogenic compositions comprise an immunologically effective amount of (i) a GAS40 polypeptide of the invention, (ii) a GAS57 double mutant of SEQ ID NO:120, (iii) a GAS25 double mutant of SEQ ID NO:121 and optionally (iv) a conjugate comprising a carrier protein selected from the group consisting of tetanus toxoid, diphtheria toxoid and CRM_{197} conjugated to at least one Group A Carbohydrate.

In a third aspect of the invention there is provided a nucleic acid encoding the GAS40 polypeptide of the first aspect.

In a fourth aspect of the invention, there is provided a cell comprising the nucleic acid of the third aspect.

In a fifth aspect, there is provided a method of reducing the amino acid sequence similarity of a wild type GAS40 polypeptide to one or more human proteins, wherein said method comprises the steps of (a) identifying similarities between the amino acid sequence of said wild type GAS40 and human proteins; and (b) introducing one or more mutations in the sequence of said wild type GAS40. Particularly said mutations are introduced in short amino-acid stretches that are common to wild type GAS40 and human proteins. Particularly the human proteins are one or more of coiled-coil domain containing 81, UPF0492 protein C20orf94, janus kinase and microtubule-interacting protein 3, myosin or tropomyosin. Yet more particularly the mutations are introduced in the coiled-coil regions of GAS40 wherein said mutations idealize the coiled-coil regions.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 shows a diagrammatic representation of the structure of wild type GAS40.

FIG. 2 shows a prediction of the secondary structure of SEQ ID NO: 2 (the wild-type GAS40 polypeptide). The prediction indicates that the integrity of the coil-coiled propensity profile and leucine zipper motif would remain unaltered.

FIG. 3 shows a prediction of the secondary structure of SEQ ID NO: 4 (the 8-9mer mutant GAS40 polypeptide). The prediction indicates that the integrity of the coil-coiled propensity profile and leucine zipper motif would remain unaltered.

FIG. 4 shows the identification of the coiled-coil regions in GAS40. These regions were highlighted with a color scale from red to blue where red represents hydrophobic residues.

FIG. 5 shows the identification of the coiled-coil regions in GAS40. These regions were highlighted with a color scale from red to blue where red represents hydrophobic residues.

FIG. 6 shows the identification of the coiled-coil regions in GAS40. These regions were highlighted with a color scale from red to blue where red represents hydrophobic residues.

DISCLOSURE OF THE INVENTION

Polypeptides of the invention

The invention provides a method of reducing the amino acid sequence similarity of a GAS40 polypeptide to one or more human proteins, which may be achieved by the introduction of mutations in the sequence of a wild type GAS40. The purpose of this reduction in the amino acid sequence similarity of GAS40 polypeptide to one or more human proteins is to reduce the theoretical risk of eliciting of antibodies that might be cross-reactive with human proteins. Such mutations may reduce or eliminate the risk of autoimmune sequelae of GAS infection. Said wild type GAS40 polypeptide may consist or comprise the sequence shown in SEQ ID NO: 2, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63, 67, 71, 75, 79, 83, 87, 91, 95, 99 or 103, or may have a sequence identity to SEQ ID NO: 2, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63, 67, 71, 75, 79, 83, 87, 91, 95, 99 or 103 which is greater than 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%. Generally such proteins avoid the elicitation of antibodies that might be cross-reactive with human proteins that comprise coiled-coil domain containing 81 (GI:47271449), UPF0492 protein C20orf94 (GI:61102723), janus kinase and microtubule-interacting protein 3 (GI:157502225), myosin or tropomyosin.

In some embodiments, said mutations include single amino acid insertions, single amino acid deletions and/or single amino acid substitutions. In some embodiments, a number of single amino acid mutations may be introduced into GAS40, for example: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 , 50 or more amino acid mutations.

In a preferred embodiment of the invention, said mutations may be introduced into short amino-acid stretches, *e.g.* 7-mers, 8-mers or 9-mers, which are common to GAS40 and human proteins. Examples of such short amino-acid stretches include QLTEELAAQ (SEQ ID NO:8), QLTEELAA (SEQ ID NO:128), KQDLAKTTS (SEQ ID NO: 9) and EALAALQA (SEQ ID NO: 5 10), which are found in the human proteins coiled-coil domain containing 81, UPF0492 protein C20orf94, and janus kinase and microtubule-interacting protein 3, respectively.

In some embodiments of the invention, the mutations that are introduced have no significant effect on the secondary structure of GAS40. In other words, the secondary structure of GAS40 polypeptides of the invention is preferably substantially the same as that of wild-type GAS40 10 polypeptides. The secondary structure of polypeptides, such as GAS40, can be assessed by *ab initio* computer prediction methods or by circular dichroism, analytical ultracentrifugation, sedimentation equilibrium analyses, X-ray crystallography, NMR spectroscopy, *etc*. As shown in Figure 1, wild-type GAS40 comprises a leader peptide at its N-terminus, two coiled-coil regions (between amino acid residues 58 and 261 and between amino acid residues 556 and 733 in SEQ ID NO: 2), a leucine 15 zipper (located within the C-terminal coiled-coil region and between amino acid residues 673 and 701 in SEQ ID NO: 2) and a transmembrane domain (located between amino acid residues 849 and 866 in SEQ ID NO: 2). The locations of these domains and regions in any GAS40 polypeptide can be predicted based on pairwise alignment of a given sequence to SEQ ID NO: 2, for example by aligning the amino acid sequence of a GAS40 polypeptide of interest to SEQ ID NO: 2 and 20 identifying the sequences that align to the respective domains/regions of SEQ ID NO:2.

Preferably, any mutations that are introduced have no significant effect on the two alpha-helical coiled-coil regions and the leucine zipper. For example, two substitutions may be introduced into each of the SEQ ID NO: 8, SEQ ID NO: 9 and/or SEQ ID NO: 10 sequences. Preferably, said mutations include T237S, A241V, A688S, S692T, A700S and/or A703S with respect to the amino 25 acid sequence of SEQ ID NO: 2. The locations of these substitutions in a GAS40 polypeptide of a different length and/or sequence to SEQ ID NO: 2 are identified based on pairwise alignment of a given sequence to SEQ ID NO: 2, for example by aligning the amino acid sequence of a GAS40 polypeptide of interest to SEQ ID NO: 2 and identifying the amino acids that correspond to T237, A241, A688, S692, A700 and A703. SEQ ID NO: 4 is an example of a GAS40 polypeptide which 30 includes all six of these substitutions (T237S, A241V, A688S, S692T, A700S and A703S).

The invention also provides GAS40 polypeptides which have been produced by a method of the invention. In some embodiments, the invention provides a GAS40 polypeptide which contains a reduced number of 9-mers, 8-mers and/or 7-mers that are present in any human polypeptides. For example, the invention also provides a GAS40 polypeptide which lacks one or more of the 35 sequences: SEQ ID NO: 8, SEQ ID NO: 9 and/or SEQ ID NO: 10. In some embodiments, the GAS40 polypeptide has a secondary structure that is substantially the same as wild-type GAS40.

Preferably the secondary structure of the two alpha-helical coiled-coil regions and the leucine zipper that is located within the C-terminal coiled-coil region is substantially the same as wild-type GAS40. Examples of GAS40 polypeptide which lack one or more of the sequences: SEQ ID NO: 8, SEQ ID NO: 9 and SEQ ID NO: 10 include SEQ ID NOs: 4, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80, 84, 88, 92, 96, 100, 104, 108, 112 and 116. In some embodiments, the invention provides a mutant protein which shares at least 95%, 96%, 96.5%, 97%, 97.2%, 97.4%, 97.6%, 97.8%, 98%, 98.2%, 98.4%, 98.6%, 98.8%, 99%, 99.2%, 99.4%, 99.6%, 99.8% or 100% identity with any one of SEQ ID NO: 4, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80, 84, 88, 92, 96, 100, 104, 108, 112 or 116.

Coiled-Coil Irregularities were recently demonstrated in M1 protein (Science 319(5868):1405-8). Similar structural irregularities occur in myosin and tropomyosin, providing a possible explanation of the M protein associated patterns of cross-reactivity in autoimmune sequelae of GAS infection. When substitutions were made within the M1 coiled-coil region resulting in core residues optimal for the formation and stability of dimeric, parallel coiled-coils ("sequence idealization"), fibrinogen binding, proinflammatory effects, and antibody cross-reactivity were diminished. A similar "idealization" of coiled-coils has been developed for GAS40.

A coiled coil is a structural motif in proteins, in which 2-7 alpha-helices are coiled together like the strands of a rope (dimers and trimers are the most common types). Coiled coils usually contain a repeated pattern, hxxhcxc (SEQ ID NO: 14), of hydrophobic (h) and charged (c) amino-acid residues, referred to as a heptad repeat. The positions in the heptad repeat are usually labeled abcdefg, where a and d are the hydrophobic positions, often being occupied by isoleucine, leucine or valine. However, the coiled coils of tropomyosin and myosin deviate from the canonical coiled coil structure in subtle ways, including insertions within heptads and charged residues and alanine residues at a and d heptad positions. Such coiled-coil destabilizing sequences alter the local conformation or energetics without interrupting the coiled coil. It is thought that cross-reactivity of GAS M proteins attributable to molecular mimicry may result from the structural similarities between the coiled coils of those proteins and those of myosin and tropomyosin. GAS40 has been found to contain similar coiled-coil destabilizing sequences.

The invention provides a GAS40 polypeptide which contains mutations in the coiled-coil regions of GAS40, preferably in order to reduce the structural irregularities in the coiled coil regions, thereby 'idealizing' said coiled-coil regions and making the mutant protein potentially less structurally similar to myosin or tropomyosin. Preferably, the hydrophilic residues in position 1 and 4 of the heptads forming the coiled coils are substituted with a hydrophobic residue such as leucine, and/or lysine 208 is deleted in order to correct a heptad frame-shift. SEQ ID NOs: 6, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61, 65, 69, 73, 77, 81, 85, 89, 93, 97, 101, 105, 109, 113, 117 are examples of GAS40 polypeptides which include these features. In some embodiments, the invention provides

a mutant protein which shares at least 90%, 91%, 92%, 92.5%, 93%, 93.5%, 94%, 94.5%, 95%, 95.5%, 96%, 96.5%, 97%, 97.5%, 98%, 98.5%, 99%, 99.5% or 100% identity with any one of SEQ ID NOs: 6, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61, 65, 69, 73, 77, 81, 85, 89, 93, 97, 101, 105, 109, 113 or 117.

The invention also provides a GAS40 polypeptide with 9 or more copies of SEQ ID NO: 14, wherein optionally said polypeptide comprises less than 32 imperfect heptad repeats. Preferably said imperfect heptad repeats contain one or two mutations in SEQ ID NO: 14 and/or share between 71 and 86% identity with SEQ ID NO: 14. In one embodiment, the invention provides a GAS40 polypeptide, wherein said polypeptide has n or more copies of SEQ ID NO: 14 and shares at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO: 2.

In some embodiments, the GAS40 polypeptide contains any combination of the mutations outlined above. For example, the invention provides a GAS40 polypeptide containing both:

- a reduced number of 9-mers, 8-mers and/or 7-mers that are present in any human polypeptides (*e.g.* lacking one or more of the sequences: SEQ ID NO: 8, SEQ ID NO: 9 and/or SEQ ID NO: 10); and
- one or more idealized coiled coil regions wherein the hydrophilic residues in position 1 and 4 of the heptads forming the coiled coils are substituted with a hydrophobic residue such as leucine, and/or lysine 208 is deleted in order to correct a heptad frame-shift.

Examples of such GAS40 polypeptides (*i.e.* those containing both: a reduced number of 9-mers, 8-mers and/or 7-mers that are present in any human polypeptides <u>and</u> one or more idealized coiled coil regions) include: SEQ ID NOs: 7, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114, 118. In one embodiment, the invention provides a GAS40 polypeptide, wherein said polypeptide shares at least 80%, 85%, 90%, 91%, 92%, 93%, 93.5%, 94%, 94.5%, 95%, 95.5%, 96%, 96.5%, 97%, 97.5%, 98%, 98.5%, 99%, 99.5% or 100% with any one of SEQ ID NOs: 7, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114, 118.

Preferably, the GAS40 polypeptides of the invention retain the immunogenicity of wild-type GAS40 polypeptides, for example the GAS40 polypeptides of the invention will preferably elicit antibodies which are able to bind wild-type GAS40 polypeptides. The GAS40 polypeptides of the invention preferably retain substantially the same level of immunogenicity as wild-type GAS40 polypeptides. For example, the geometric means of the antibody titers resulting from the GAS40 polypeptides of the invention are preferably not significantly different from those resulting from wild-type GAS40 polypeptides when assessed by two sample t tests (wherein differences are considered significant if the P value obtained by one-tailed analysis is <0.05, <0.01, or <0.001) when

other variables (such as dose, delivery route, GAS40 strain, population group, *etc.*) are controlled. Methods of determining the immunogenicity of bacterial polypeptides are well known in the art.

The invention also provides fragments of the GAS40 polypeptides of the invention, wherein said fragments are at least: 177, 203, 250, 300, 400, 450, 460, 465, 470, 471, 500, 550, 600, 650, 700, 750 or 800 amino acids in length. Examples of such fragments include those recited in SEQ ID NOs: 107, 111 and 115.

The invention provides a method of reducing the risk of auto-immunogenicity of a GAS40 polypeptide, wherein said method comprises the steps of: a) identifying short (7-mer, 8-mer or 9-mer) amino acid stretches that are common to human proteins and GAS40; and b) introducing mutations into said short amino acid stretches that are common to human proteins and GAS40.

The invention also provides a method of reducing the risk of auto-immunogenicity of a GAS40 polypeptide, wherein said method comprises the steps of: a) identifying coiled-coil regions in GAS40; and b) introducing mutations to idealize the coiled-coil structure.

Step b) may involve in the introduction of one or more of the following mutations: substitution of one or more of the hydrophilic residues in position 1 and 4 of the coiled coil heptads with a hydrophobic residue, preferably leucine; and deletion of an amino acid residue, preferably lysine 208, in order to correct a heptad frame-shift.

The invention also provides a method of reducing the risk of auto-immunogenicity of a GAS40 polypeptide, wherein said method comprises the steps of: a) identifying short amino acid stretches that are common to human proteins and GAS40; b) introducing mutations into said short amino acid stretches that are common to human proteins and GAS40; c) identifying coiled-coil regions in GAS40; and d) introducing mutations to idealize the coiled-coil structure.

Step d) may involve in the introduction of one or more of the following mutations: substitution of one or more of the hydrophilic residues in position 1 and 4 of the coiled coil heptads with a hydrophobic residue, preferably leucine; and deletion of an amino acid residue, preferably lysine 208, in order to correct a heptad frame-shift.

Preferably, GAS40 polypeptides of the invention do not cause molecular mimicry and/or do not cause cross-activation of autoreactive T or B cells.

The invention provides a GAS40 polypeptide which can elicit antibodies that are cross-reactive with wild-type GAS40 polypeptides, but which does not elicit antibodies that are cross-reactive with human polypeptides. The invention also provides a GAS40 polypeptide which can elicit antibodies that are cross-reactive with wild-type GAS40 polypeptides, but which does not contain 7-mers, 8-mers or 9-mers that are present in human proteins. The invention also provides a GAS40 polypeptide which can elicit antibodies that are cross-reactive with wild-type GAS40 polypeptides and contains 9, 10, 15, 20, 25, 30, 31, 32 or more copies of SEQ ID NO: 14.

Compositions

The invention provides compositions comprising the GAS40 polypeptides of the invention. Such compositions may comprise additional polypeptides, such as other polypeptides from GAS (*e.g.* GAS57, GAS25 and/or GAS polysaccharide antigen). These compositions may be for use as medicaments (e.g., as immunogenic compositions or vaccines). Compositions of the invention are useful for preventing *S. pyogenes* infection, reducing the risk of *S. pyogenes* infection, and/or treating disease caused as a result of *S. pyogenes* infection, such as bacteremia, meningitis, puerperal fever, scarlet fever, erysipelas, pharyngitis, impetigo, necrotizing fasciitis, myositis or toxic shock syndrome.

Compositions containing GAS antigens are preferably immunogenic compositions, and are more preferably vaccine compositions. The pH of such compositions preferably is between 6 and 8, preferably about 7. The pH can be maintained by the use of a buffer. The composition can be sterile and/or pyrogen free. The composition can be isotonic with respect to humans.

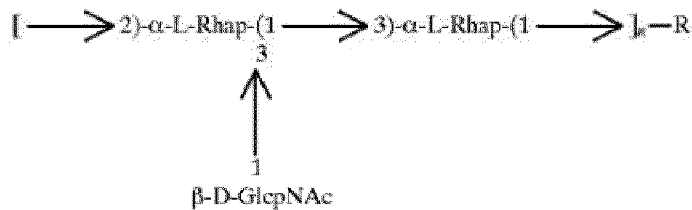
Vaccines according to the invention may be used either prophylactically or therapeutically, but will typically be prophylactic. Accordingly, the invention includes a method for the therapeutic or prophylactic treatment of a *Streptococcus pyogenes* infection. The animal is preferably a mammal, most preferably a human. The methods involve administering to the animal a therapeutic or prophylactic amount of the immunogenic compositions of the invention. The invention also provides the immunogenic compositions of the invention for treating, reducing the risk or, and/or preventing a *S. pyogenes* infection.

Some compositions comprise GAS40 polypeptides of the invention and at least one or at least two different GAS antigens, as described below. Other compositions of the invention comprise at least one nucleic acid molecule which encodes the two different antigens. See, e.g., Robinson & Torres (1997) Seminars in Immunology 9:271-283; Donnelly et al. (1997) Ann. Rev Immunol 15:617-648; Scott-Taylor & Dalgleish (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. Typically the nucleic acid molecule is a DNA molecule, e.g., in the form of a plasmid.

GAS57 is also referred to as 'Spy0416 (M1)', 'SpyM3_0298' (M3), 'SpyM18_0464' (M18), and 'prtS'. Spy0416 has been identified as a putative cell envelope proteinase. See WO02/34771 and US 2006/0258849. GAS57 mutants useful in the invention include those with at an amino acid alteration (i.e., a substitution, deletion, or insertion) at one or more of amino acids D151, H279, particularly an amino acid sequence of SEQ ID NO:120.

GAS25 (Spy0167, streptolysin, SLO) is a potent pore-forming toxin which induces host cell lysis and is described, inter alia, in WO 02/34771. Mutant forms of GAS25 have at least 50% less hemolytic activity than wild-type Spy0167 (e.g., 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99, or 100%) relative to wild-type Spy0167 as determined by a hemolysis assay but are immunogenic and preferably confer protection against GAS lethal challenge in a mouse model. Spy0167 mutants for use in compositions of the invention include those with an amino acid alteration (i.e., a substitution, deletion, or insertion) at one or more of amino acids P427 and W535 as exemplified in SEQ ID NO: 121.

GAS polysaccharide (PS) is a cell-wall polysaccharide present in all GAS strains. Antibody titers to PS correlate inversely with disease and colonization in children. In some embodiments compositions of the invention comprise a GAS polysaccharide antigen. *S. pyogenes* GAS carbohydrate typically features a branched structure with an L-rhamnopyranose (Rhap) backbone consisting of alternating alpha-(1→2) and alpha-(1→3) links and D-N-acetylglucosamine (GlcpNAc) residues beta-(1→3)-connected to alternating rhamnose rings (Kreis et al., Int. J. Biol. Macromol. 17, 117-30, 1995). GAS polysaccharide antigens useful in compositions of the invention have the formula:



wherein R is a terminal reducing L-Rhamnose or D-GlcpNAc and n is a number from about 3 to about 30. The GAS polysaccharide antigen used according to the invention may be a substantially full-length GAS carbohydrate, as found in nature, or it may be shorter than the natural length. Full-length polysaccharides may be depolymerized to give shorter fragments for use with the invention, e.g., by hydrolysis in mild acid, by heating, by sizing chromatography, etc. However, it is preferred to use saccharides of substantially full-length. In particular, it is preferred to use saccharides with a molecular weight of about 10 kDa. Molecular masses can be measured by gel filtration relative to dextran standards. The saccharide may be chemically modified relative to the GAS carbohydrate as found in nature. For example, the saccharide may be de N acetylated (partially or fully), N propionated (partially or fully), etc. The effect of de-acetylation etc., for example on immunogenicity, can be assessed by routine assays. In some embodiments the GAS polysaccharide antigen is conjugated to a carrier, such as the mutated diphtheria toxin CRM197.

In one embodiment, the invention provides compositions comprising GAS57 and one or more GAS40 polypeptides of the invention. Sequences of exemplary GAS57 polypeptides are provided in SEQ ID NOs: 119 and 120. Such compositions may also comprise other GAS antigens,

particularly GAS25 having SEQ ID NO: 121. In some embodiments, the GAS40 polypeptides of the invention (and optionally GAS57) can be incorporated into an immunogenic composition, including a vaccine composition. Such compositions can be used to raise antibodies in a mammal (*e.g.* a human). In these compositions, the GAS40 polypeptides of the invention (and optionally GAS57, if present) may act as immunogens and/or as antigens. The invention provides pharmaceutical compositions comprising a GAS40 polypeptide of the invention. Such pharmaceutical compositions may also comprise additional polypeptides such as GAS57.

The invention also provides processes for making a pharmaceutical composition involving combining a GAS40 polypeptide of the invention with a pharmaceutically acceptable carrier. Such processes may also comprise the step of adding additional polypeptides, such as GAS57.

In certain embodiments the vaccine composition will comprise one or more pharmaceutically acceptable carriers, diluents and/or adjuvants. Adjuvants which may be used in compositions of the invention include, but are not limited to:

• mineral salts, such as aluminium salts and calcium salts, including hydroxides (e.g. oxyhydroxides), phosphates (e.g. hydroxyphosphates, orthophosphates) and sulphates, etc.;

• oil-in-water emulsions, such as squalene-water emulsions, including MF59 (5% Squalene, 0.5% Tween 80, and 0.5% Span 85, formulated into submicron particles using a microfluidizer), complete Freund's adjuvant (CFA) and incomplete Freund's adjuvant (IFA);

• saponin formulations such as QS21 and ISCOMs;

• virosomes and virus like particles (VLPs);

• bacterial or microbial derivatives, such as non toxic derivatives of enterobacterial lipopolysaccharide (LPS), Lipid A derivatives, immunostimulatory oligonucleotides, such as IC 31™;

• human immunomodulators, including cytokines, such as interleukins (e.g. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, interferons (e.g. interferon γ), macrophage colony stimulating factor, and tumor necrosis factor;

• bioadhesives and mucoadhesives, such as chitosan and derivatives thereof, esterified hyaluronic acid microspheres or mucoadhesives, such as cross linked derivatives of poly(acrylic acid), polyvinyl alcohol, polyvinyl pyrollidone, polysaccharides and carboxymethylcellulos;

• microparticles (i.e. a particle of ~100nm to ~150μm in diameter, more preferably ~200nm to ~30μm in diameter, and most preferably ~500nm to ~10μm in diameter) formed from materials that are biodegradable and non toxic (e.g. a poly(α-hydroxy acid), a polyhydroxybutyric acid, a polyorthoester, a polyanhydride, a polycaprolactone, etc.);

• liposomes;

• polyoxyethylene ethers and polyoxyethylene esters;

• PCPP formulations;

• muramyl peptides, including N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), and N acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine MTP-PE); and

• imidazoquinolone compounds, including Imiquamod and its homologues (e.g. “Resiquimod 3M”).

The invention may also comprise combinations of one or more of the adjuvants identified above. For example, the following adjuvant compositions may be used in the invention: (1) a saponin and an oil-in-water emulsion; (2) a saponin (e.g. QS21) + a non-toxic LPS derivative (e.g. 3dMPL); (3) a saponin (e.g. QS21) + a non-toxic LPS derivative (e.g. 3dMPL) + a cholesterol; (4) a saponin (e.g. QS21) + 3dMPL + IL 12 (optionally + a sterol); (5) combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions; (6) SAF, containing 10% squalane, 0.4% Tween 80™, 5% pluronic-block polymer L121, and thr-MDP, either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion. (7) RibiTM adjuvant system (RAS), (Ribi Immunochem) containing 2% squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); and (8) one or more mineral salts (such as an aluminum salt) + a non-toxic derivative of LPS (such as 3dMPL).

The use of an aluminium hydroxide and/or aluminium phosphate adjuvant is useful, particularly in children, and antigens are generally adsorbed to these salts. Squalene-in-water emulsions are also preferred, particularly in the elderly. Useful adjuvant combinations include combinations of Th1 and Th2 adjuvants such as CpG & alum or resiquimod & alum. A combination of aluminium phosphate and 3dMPL may be used.

Vaccines of the invention may be prophylactic (*i.e.* to prevent disease) or therapeutic (*i.e.* to reduce or eliminate the symptoms of a disease).

Preferably, when said compositions of the invention are administered to a large number of patients, immunity against GAS is produced in some or all of the patients and a reduced percentage of patients develop autoimmune sequelae of GAS infection, such as rheumatic fever or acute glomerulonephritis, compared to the percentage of patients that develop said autoimmune sequelae when wild type GAS40 is administered.

Nucleic acids

The invention also provides nucleic acids encoding the GAS40 polypeptides of the invention. For example, the nucleic acid sequences shown in SEQ ID NOs: 1, 3 and 5 encode the GAS40 polypeptides depicted in SEQ ID NOs: 2, 4 and 6.

5 The invention also provides nucleic acid comprising nucleotide sequences having sequence identity to such nucleotide sequences. Identity between sequences is preferably determined by the Smith-Waterman homology search algorithm. Such nucleic acids include those using alternative codons to encode the same amino acid.

The invention includes nucleic acid comprising sequences complementary to these sequences 10 (*e.g.* for antisense or probing, or for use as primers).

Nucleic acids of the invention can be used in hybridization reactions (*e.g.* Northern or Southern blots, or in nucleic acid microarrays or 'gene chips') and amplification reactions (*e.g.* PCR, SDA, SSSR, LCR, TMA, NASBA, *etc.*) and other nucleic acid techniques.

Nucleic acid according to the invention can take various forms (*e.g.* single-stranded, 15 double-stranded, vectors, labeled *etc.*). Nucleic acids of the invention may be circular or branched, but will generally be linear. Unless otherwise specified or required, any embodiment of the invention that utilizes a nucleic acid may utilize both the double-stranded form and each of two complementary single-stranded forms which make up the double-stranded form.

Nucleic acids of the invention are preferably provided in purified or substantially purified form 20 *i.e.* substantially free from other nucleic acids (*e.g.* free from naturally-occurring nucleic acids), particularly from other *E. coli* or host cell nucleic acids, generally being at least about 50% pure (by weight), and usually at least about 90% pure.

Nucleic acids of the invention may be prepared in many ways *e.g.* by chemical synthesis (*e.g.* phosphoramidite synthesis of DNA) in whole or in part, by digesting longer nucleic acids using 25 nucleases (*e.g.* restriction enzymes), by joining shorter nucleic acids or nucleotides (*e.g.* using ligases or polymerases), from genomic or cDNA libraries, *etc.*

Nucleic acid of the invention may be attached to a solid support (*e.g.* a bead, plate, filter, film, slide, microarray support, resin, *etc.*). Nucleic acid of the invention may be labeled *e.g.* with a radioactive or fluorescent label, or a biotin label. This is particularly useful where the nucleic acid is 30 to be used in detection techniques *e.g.* where the nucleic acid is a primer or as a probe.

The term "nucleic acid" includes in general means a polymeric form of nucleotides of any length, which contain deoxyribonucleotides, ribonucleotides, and/or their analogs. It includes DNA, RNA, DNA/RNA hybrids. It also includes DNA or RNA analogs, such as those containing modified backbones (*e.g.* peptide nucleic acids (PNAs) or phosphorothioates) or modified bases. Thus the

invention includes mRNA, tRNA, rRNA, ribozymes, DNA, cDNA, recombinant nucleic acids, branched nucleic acids, plasmids, vectors, probes, primers, *etc.*. Where nucleic acid of the invention takes the form of RNA, it may or may not have a 5' cap.

Nucleic acids of the invention may be part of a vector *i.e.* part of a nucleic acid construct designed for transduction/transfection of one or more cell types. Vectors may be, for example, “cloning vectors” which are designed for isolation, propagation and replication of inserted nucleotides, “expression vectors” which are designed for expression of a nucleotide sequence in a host cell, “viral vectors” which is designed to result in the production of a recombinant virus or virus-like particle, or “shuttle vectors”, which comprise the attributes of more than one type of vector. Preferred vectors are plasmids, as mentioned above. A “host cell” includes an individual cell or cell culture which can be or has been a recipient of exogenous nucleic acid. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation and/or change. Host cells include cells transfected or infected *in vivo* or *in vitro* with nucleic acid of the invention.

Where a nucleic acid is DNA, it will be appreciated that “U” in a RNA sequence will be replaced by “T” in the DNA. Similarly, where a nucleic acid is RNA, it will be appreciated that “T” in a DNA sequence will be replaced by “U” in the RNA.

The term “complement” or “complementary” when used in relation to nucleic acids refers to Watson-Crick base pairing. Thus the complement of C is G, the complement of G is C, the complement of A is T (or U), and the complement of T (or U) is A. It is also possible to use bases such as I (the purine inosine) *e.g.* to complement pyrimidines (C or T).

Nucleic acids of the invention can be used, for example: to produce polypeptides; as hybridization probes for the detection of nucleic acid in biological samples; to generate additional copies of the nucleic acids; to generate ribozymes or antisense oligonucleotides; as single-stranded DNA primers or probes; or as triple-strand forming oligonucleotides.

The invention provides a process for producing nucleic acid of the invention, wherein the nucleic acid is synthesized in part or in whole using chemical means.

The invention provides vectors comprising nucleotide sequences of the invention (*e.g.* cloning or expression vectors) and host cells transformed with such vectors.

For certain embodiments of the invention, nucleic acids are preferably at least 1340 nucleotides in length (*e.g.* 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, or 2400 nucleotides or longer).

Cells of the invention

The invention also provides a cell comprising a nucleic acid of the invention. Said cell may be stably transformed with said nucleic acid. The invention also provides a cell that expresses a GAS40 polypeptide of the invention.

5 ***Production of GAS protein antigens***

The redundancy of the genetic code is well-known. Thus, any nucleic acid molecule (polynucleotide) which encodes one of the GAS antigens described herein can be used to produce that protein recombinantly. Nucleic acid molecules encoding wild-type GAS antigens also can be isolated from the appropriate *S. pyogenes* bacterium using standard nucleic acid purification
10 techniques or can be synthesized using an amplification technique, such as the polymerase chain reaction (PCR), or by using an automatic synthesizer. See Caruthers et al., Nucl. Acids Res. Symp. Ser. 215, 223, 1980; Horn et al., Nucl. Acids Res. Symp. Ser. 225, 232, 1980; Hunkapiller et al., Nature 310, 105-11, 1984; Grantham et al., Nucleic Acids Res. 9, r43-r74, 1981. cDNA molecules can be made with standard molecular biology techniques, using mRNA as a template. cDNA
15 molecules can thereafter be replicated using molecular biology techniques well known in the art. An amplification technique, such as PCR, can be used to obtain additional copies of polynucleotides of the invention, using either genomic DNA or cDNA as a template.

If desired, polynucleotides can be engineered using methods generally known in the art to alter antigen-encoding sequences for a variety of reasons, including but not limited to, alterations which
20 modify the cloning, processing, and/or expression of the polypeptide or mRNA product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides can be used to engineer the nucleotide sequences. For example, site directed mutagenesis can be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

25 Sequence modifications, such as the addition of a purification tag sequence or codon optimization, can be used to facilitate expression. For example, the N-terminal leader sequence may be replaced with a sequence encoding for a tag protein such as polyhistidine ("HIS") or glutathione S-transferase ("GST"). Such tag proteins may be used to facilitate purification, detection, and stability of the expressed protein. Codons preferred by a particular prokaryotic or eukaryotic host can
30 be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half life which is longer than that of a transcript generated from the naturally occurring sequence. These methods are well known in the art and are further described in WO05/032582.

A nucleic acid molecule which encodes a GAS antigen for use in the invention can be inserted
35 into an expression vector which contains the necessary elements for the transcription and translation

of the inserted coding sequence. Methods which are well known to those skilled in the art can be used to construct expression vectors containing coding sequences and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination.

5 Host cells for producing GAS antigens can be prokaryotic or eukaryotic. *E. coli* is a preferred host cell, but other suitable hosts include *Lactococcus lactis, Lactococcus cremoris, Bacillus subtilis, Vibrio cholerae, Salmonella typhi, Salmonella typhimurium, Neisseria lactamica, Neisseria cinerea, Mycobacteria* (e.g., *M. tuberculosis*), yeasts, baculovirus, mammalian cells, etc.

A host cell strain can be chosen for its ability to modulate the expression of the inserted 10 sequences or to process the expressed polypeptide in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post translational processing which cleaves a "prepro" form of the polypeptide also can be used to facilitate correct insertion, folding and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post 15 translational activities are available from the American Type Culture Collection (ATCC; 10801 University Boulevard, Manassas, VA 20110-2209) and can be chosen to ensure the correct modification and processing of a foreign protein. See WO 01/98340.

Expression constructs can be introduced into host cells using well-established techniques which include, but are not limited to, transferrin-polycation-mediated DNA transfer, transfection 20 with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, "gene gun" methods, and DEAE- or calcium phosphate-mediated transfection.

Host cells transformed with expression vectors can be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed 25 cell can be secreted or contained intracellularly depending on the nucleotide sequence and/or the expression vector used. Those of skill in the art understand that expression vectors can be designed to contain signal sequences which direct secretion of soluble antigens through a prokaryotic or eukaryotic cell membrane.

Signal export sequences can be included in a recombinantly produced GAS antigen so that the 30 antigen can be purified from cell culture medium using known methods. Alternatively, recombinantly produced GAS antigens can be isolated from engineered host cells and separated from other components in the cell, such as proteins, carbohydrates, or lipids, using methods well-known in the art. Such methods include, but are not limited to, size exclusion chromatography, ammonium sulfate fractionation, ion exchange chromatography, affinity chromatography, and preparative gel 35 electrophoresis. A preparation of purified GAS antigens is at least 80% pure; preferably, the preparations are 90%, 95%, or 99% pure. Purity of the preparations can be assessed by any means

known in the art, such as SDS-polyacrylamide gel electrophoresis or RP-HPLC analysis. Where appropriate, mutant Spy0167 proteins can be solubilized, for example, with urea.

GAS antigens can be synthesized, for example, using solid phase techniques. See, e.g., Merrifield, J. Am. Chem. Soc. 85, 2149 54, 1963; Roberge et al., Science 269, 202 04, 1995. Protein
5 synthesis can be performed using manual techniques or by automation. Automated synthesis can be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Optionally, fragments of GAS antigens can be separately synthesized and combined using chemical methods to produce a full-length molecule.

General

10 The term “comprising” encompasses “including” as well as “consisting” *e.g.* a composition “comprising” X may consist exclusively of X or may include something additional *e.g.* X + Y.

The word “substantially” does not exclude “completely” *e.g.* a composition which is “substantially free” from Y may be completely free from Y. Where necessary, the word “substantially” may be omitted from the definition of the invention. The term “consisting essentially
15 of” means that the composition, method or structure may include additional ingredients, steps and/or parts, but only if the additional ingredients, steps and/or parts do no materially alter the basic and novel characteristics of the claimed composition, method or structure. The term “consisting of” is generally taken to mean that the invention as claimed is limited to those elements specifically recited in the claim (and may include their equivalents, insofar as the doctrine of equivalents is applicable).

20 The term “about” as used herein when referring to a measurable value such as an amount, a temporal duration, and the like, is meant to encompass variations of ±20% or ±10%, more preferably ±5%, even more preferably ±1%, and still more preferably ±0.1% from the specified value, as such variations are appropriate to perform the disclosed methods.

The word “substantially” does not exclude “completely” e.g. a composition which is
25 “substantially free” from Y may be completely free from Y. For example, “substantially free” from Y can be understood as a composition containing not more than 5% Y, not more than 4% Y, not more than 3% Y, not more than 2% Y, not more than 1% Y, or not more than 0.1% Y. Where necessary, the word “substantially” may be omitted from the definition of the invention.

The term “mutant” refers to a gene or gene product that displays modifications in sequence
30 and/or functional properties (i.e., altered characteristics) when compared to the wild-type gene or gene product. Thus, wild-type sequences and fragments of wild-type sequences, which do not include substitutions of the present invention, are excluded. For example, sequences not including at least two or more positions selected from the group consisting of T237, A241, A688, S692, A700, A703, A38, Q52, T66, A148, S158, A179, A182, T186, A196, Q203, Q218, K225, A544, T558,
35 S562, K576, A586, T614, A618, A621, T665, A672, Q683, T693 may be excluded.

All GenBank Accession numbers provided herein are incorporated by reference.

Unless specifically stated, a process comprising a step of mixing two or more components does not require any specific order of mixing. Thus components can be mixed in any order. Where there are three components then two components can be combined with each other, and then the 5 combination may be combined with the third component, *etc.*

Where animal (and particularly bovine) materials are used in the culture of cells, they should be obtained from sources that are free from transmissible spongiform encaphalopathies (TSEs), and in particular free from bovine spongiform encephalopathy (BSE). Overall, it is preferred to culture cells in the total absence of animal-derived materials.

10 Where a compound is administered to the body as part of a composition then that compound may alternatively be replaced by a suitable prodrug.

Sequence identity between polypeptide sequences is preferably determined by pairwise alignment algorithm using the Needleman-Wunsch global alignment algorithm (Needleman& Wunsch (1970) *J. Mol. Biol.* 48, 443-453), using default parameters (*e.g.* with Gap opening penalty = 15 10.0, and with Gap extension penalty = 0.5, using the EBLOSUM62 scoring matrix). This algorithm is conveniently implemented in the *needle* tool in the EMBOSS package (Rice *et al.* (2000) *Trends Genet* 16:276-277). Sequence identity should be calculated over the entire length of the polypeptide sequence of the invention.

EXAMPLES

20 ***Example 1***

Bioinformatics analysis to identify short amino acid stretches common to human proteins potentially responsible of linear epitope mimicry, leading to autoimmune diseases

In order to detect short amino acid stretches that are common to both the GAS40 antigen and one or more human proteins and potentially responsible of linear epitope mimicry, the amino acid 25 sequences of GAS40 was subdivided into overlapping 7-mers, 8-mers, 9-mers and 10-mers, with a sliding window of 1 amino acid residue. Each of these fragments was compared to a collection of human proteins available from ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/protein/. The same analysis was also performed on the sequence of tetanus toxin (GI:135624), the safety of which is widely accepted, and M1 and M5 protein of *Streptococcus pyogenes* (Streptococcus pyogenes M1 GAS 30 GI:15675799, Streptococcus pyogenes serotype M5 GI:126669), that have also been reported to contain epitopes that cross-react with human tissues (Baird, et al. 1991). The results of this analysis are summarized in Table 1.

Table 1

	Tetanus toxin	*M5*	*M1*	*GAS40*

Stretch size	*Number of hits*	*Number of hits*	*Number of hits*	*Number of hits*
7-mer	32	84	102	94
8-mer	0	15	16	1
9-mer	0	0	0	2
10-mer	0	0	0	0

Two 9-mer hits and one 8-mer hit were detected for GAS40. The first 9-mer has the sequence shown in QLTEELAAQ (SEQ ID NO: 8), and it is located at position 235-243 of GAS40 which is in the first coiled-coil domain. This 9-mer is also found in a 562-amino acid human protein known as "*coiled-coil domain containing 81*" (GI:47271449). The second 9-mer has the sequence
5 shown in SEQ ID NO: 9, and it is located at position 684-692 of GAS40, which is in the leucine zipper domain. This 9-mer is also found in a 408-amino acid human protein known as *"UPF0492 protein C20orf94"* (GI:61102723). The 8-mer has the sequence shown in SEQ ID NO: 10 and it is located at position 699-706, which bridges the C-terminal boundary of the leucine zipper domain. This 8-mer is also found in a 844-amino acid human protein known as *"janus kinase and*
10 *microtubule-interacting protein 3"* (GI:157502225).

In order to reduce the sequence identity of the above-mentioned 9-mers and 8-mer to the corresponding human protein sequences, multiple mutations were introduced into these 9-mers and 8-mer. In order to avoid a reduction in the immunological efficacy of GAS40, six amino acid substitutions (two in each of the 8/9-mer hits) were introduced which were designed to have no effect
15 on the overall structure of the protein (alpha-helical coiled-coil/leucine zipper). The two substitutions that were made to the first 9-mer (QLTEELAAQ) were T237S and A241V, resulting in the mutated 9-mer sequence: QLsEELvAQ (SEQ ID NO: 11). The two substitutions that were made to SEQ ID NO: 9 were A688S and S692T, resulting in the mutated 9-mer sequence: KQDLsKTTt (SEQ ID NO: 12). The two substitutions that were made to SEQ ID NO: 10 were A700S and
20 A703S, resulting in the mutated 8-mer sequence: EsLAsLQA (SEQ ID NO: 13).

Figure 1 shows a prediction of the secondary structure of GAS40 wild-type and mutant proteins. As shown, the prediction indicates that the integrity of the coil-coiled propensity profile and leucine zipper motif would remain unaltered by the above-mentioned six mutations.

Example 2

25 *Bioinformatics analysis to identify possible coiled-coil irregularities similar to human myosin and tropomyosin in GAS40*

Coiled-coil irregularities were recently demonstrated in M1 protein (McNamara, et al. 2008). Similar structural irregularities occur in myosin and tropomyosin, providing a possible explanation on the M protein associated patterns of cross-reactivity in autoimmune sequelae of
30 GAS infection. When substitutions were made within the M1 coiled-coil region resulting in core

residues optimal for the formation and stability of dimeric, parallel coiled-coils ("sequence idealization"), fibrinogen binding, proinflammatory effects, and antibody cross-reactivity were diminished. A similar "idealization" of coiled-coil approach was followed in the case of GAS40 to mitigate any potential cross-reactivity.

5 Using the Jalview package (http://www.jalview.org/), the GAS40 coiled-coil regions were identified and highlighted with a colour scale from red to blue where red represents hydrophobic residues and blue represents hydrophilic residues. Hydrophilic residues in position 1 and 4 of the heptads forming the coiled coils were substituted with the hydrophobic residue leucine to idealize the coiled-coil structure, and lysine 208 was deleted to correct a heptad frame-shift. A list of the 10 mutations that were introduced into GAS40 is shown in Table 2.

Table 2

pos. 38 A → L	pos. 182 A → L	pos. 225 K → L	pos. 614 T → L
pos. 52 Q → L	pos. 186 T → L	pos. 544 A → L	pos. 618 A → L
pos. 66 T → L	pos. 196 A → L	pos. 558 T → L	pos. 621 A → L
pos. 148 A → L	pos. 203 Q → L	pos. 562 S → L	pos. 665 T → L
pos. 158 S → L	pos. 208 K → -	pos. 576 K → L	pos. 672 A → L
pos. 179 A → L	pos. 218 Q → L	pos. 586 A → L	pos. 683 Q → L
			pos. 693 T → L

Example 3

Cloning & General DNA techniques

3.1 GAS40 without 8 and 9-mer homologies

15 Three rare AGA codons encoding arginine (Arg 359, Arg 360 and Arg 849) were mutagenized to the common arginine CGT codon in order to optimize expression of GAS40. The nucleotide sequence of GAS40 containing the AGA to CGT substitutions of arginine codons, but excluding the leader peptide and transmembrane domain sequences, is shown in SEQ ID NO:1 and the corresponding amino acid sequence is shown in SEQ ID NO:2.

20 A pET21b+ vector containing the codon-optimised GAS40 gene was used as template to create the GAS40 8-9-mers mutant.

The introduction of the 6 mutations was performed in 3 steps, each incorporating 2 mutations in one of the human-homologous stretches, resulting in two intermediate constructs and one final construct incorporating all three mutations.

In each step, the Polymerase Incomplete Primer Extension (PIPE) Cloning Method (Klock and Lesley 2009) was used to amplify the whole plasmid using two primers designed to create two mutations (substitutions) and to be complementary to each other, so that the linearized PCR product could self-anneal to recreate a viable, mutant plasmid. The primers used in each of the three steps are shown below:

Table 3

Mutations T237S and A241V	
T237S_A241V_F1:	AAGCAATTGTCTGAAGAGTTGGTAGCTCAGAAAGCTGCTCTAGCAG
T237S_A241V_R1	TCTGAGCTACCAACTCTTCAGACAATTGCTTTTTAACTTTAGCTTTTTGATTTTCT
Mutations A688S and S692T	
A688S_S692T_F1	AAGCAAGATTTGTCTAAAACTACCACATCTTTGTTAAATGCACAAGAAGCT
A688S_S692T_R1	TAACAAAGATGTGGTAGTTTTAGACAAATCTTGCTTAGTATTATCAATGCGC
Mutations A700S and A703S	
A700S_A703S_F1	TGCACAAGAATCTTTAGCATCCTTACAAGCTAAACAAAGCAGTCTAGAA
A700S_A703S_R1	GCTTGTAAGGATGCTAAAGATTCTTGTGCATTTAACAAAGATGTGGTAGT
(Shading indicates the nucleotide substitutions introduced)	

The nucleotide and amino acid sequences of the GAS40 8-9-mers mutant, with leader peptide and transmembrane domain sequences deleted, are shown in SEQ ID NOs: 3 and 4, respectively.

3.2 GAS40 with idealized coiled-coils

A codon-optimised GAS40 'multi-mutant' protein containing idealized coiled-coils was produced by chemical synthesis. In order to optimise expression, GC-content was adjusted to prolong mRNA half-life. Codon usage was also adapted to the bias of *Escherichia coli*. The nucleotide and amino acid sequences of GAS40 with idealized coiled-coils (multi-mutant), with the leader peptide and transmembrane domain sequences deleted, are shown in SEQ ID NOs: 5 and 6.

Example 4

Expression & Purification of recombinant GAS40 and its derivatives

Recombinant GAS40 his-tagged mutant derivatives were expressed in BL21(DE3) *E. coli* cells and purified by affinity chromatography on immobilized metal ion affinity chromatography columns.

The cells were first incubated at 30°C with agitation at 180 rpm until they reached a culture density of OD_{600nm} = 0.4 - 0.6, at which point the cells were induced 1 mM IPTG, and incubated for a further 3 hours at 25°C with agitation at 180 rpm. The cells were then recovered by centrifugation and frozen at -20°C.

The frozen *E. coli* pellets were suspended in lysis buffer (10 ml B-PER™ (Pierce), 0.1 mM $MgCl_2$, 100 units DNase I (Sigma), and 1 mg/ml lysozyme (Sigma)) and mixed for 30-40 minutes at room temperature. The resulting lysates were centrifuged at 30000-40000 x *g* for 20-25 minutes, and the supernatants were then loaded onto wash buffer A (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0) equilibrated columns (Poly-Prep with 1 ml of Ni-Activated Chelating Sepharose Fast Flow resin). The loaded resin was washed three times with wash buffer A and three times with wash buffer B (70 mM imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0). Proteins were eluted with elution buffer (250 mM imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0) in Eppendorf tubes containing 1 mM DTT. Total elution proteins were quantified with Bradford reagent, extensively dialyzed against PBS and then analysed by SDS-polyacrylamide gel electrophoresis (see Figure 7) using 4-12% Criterion™ pre-cast gels (Bio-Rad).

Example 5

Immunization & Infection

To test whether the GAS mutants are able to induce protection against lethal challenge in mice, five-week old mice were immunized intraperitoneally three times (day 0, day 21 and day 35) with either his-tagged wild-type GAS40, GAS40 8-9-mers mutant or GAS40 multi-mutant, which were each combined with an alum adjuvant (20 μg protein in 2 mg/ml aluminium hydroxide). Mice immunized with adjuvant alone were used as negative controls. Two weeks after the third immunization, blood samples were taken. A few days afterwards, immunized mice were challenged intranasally with 10^8 cfu (50 μl) of the M1 3348 GAS strains. Mice survival was monitored for a 10-14 day period.

As shown in the table below, both the multi-mutant and '9-mer mutant' increased the survival rate of immunised mice by 4.2- to 4.3-times the survival rate of the negative control. This indicates that the mutant proteins produce high protection levels. Indeed, the mutant proteins were even more effective than wild-type GAS40.

Table 4

	Intranasal infection with M1 3348 strain - Alum adjuvant					
Antigen	***Form***	***Dose (μg/mouse)***	***N° mice Tested***	***Dead mice***	***% Survival***	***Fisher p-val (*)***
Mutant 9-mer	his	20	63	18	71	< 0,0001
Multi-mutant	his	20	48	13	73	< 0,0001
Wild Type GAS40	his	20	80	27	66	< 0,0001
Negative control	-	-	64	53	17	-
	(*) Fisher p-vals are calculated against negative control group					

Example 6

Cross-reactivity of mouse antisera to human tissues

The development of autoimmune disease is associated with persistent levels of high titer, high affinity antibodies. In order to assess the potential for the GAS40 to elicit antibodies recognizing specific human tissue components, a pilot cross-reactivity study with mouse anti-sera raised against GAS40 was performed. Sera were obtained from groups of 8 outbred CD1 mice immunized as shown in Table 10.

Table 10

Group	**Treatment**	**No. Animals**	**Dose (ug)**	**Dose Volume**
1.	PBS (Control)	8	0	200ul ($Al(OH)_3$ at 2mg/mL
2.	M5 Protein	8	20	200ul ($Al(OH)_3$ at 2mg/mL
3.	M6 Protein	8	20	200ul ($Al(OH)_3$ at 2mg/mL
4.	M12 Protein	8	20	200ul ($Al(OH)_3$ at 2mg/mL
5.	GAS40	8	20	200ul ($Al(OH)_3$ at 2mg/mL

The non-GLP human tissue cross-reactivity study was optimized and run at Charles River Laboratories/ PAI in Frederick, MD. A panel of selected human tissues was screened because of their association with specific post-GAS infection pathology: cartilage, brain (cerebrum), heart, heart valve, kidney and skeletal muscle. Integrity of the tissue was confirmed by staining with beta 2 microglobulin. Pooled sera from mice with the highest titers were optimized for use at dilutions of 1:10,000 and 1:40,000. Spot control slides of the individual antigens were incubated with the pooled anti-sera from the respective group of animals in order to confirm antigen-specific responses. No test article-specific cell membrane staining was observed in any group suggesting that antibodies to GAS40 did not recognize membrane elements which could influence biologically processes. Specific

staining was observed in the cytoplasm with anti-sera from mice treated with M6, M12 and GAS40 and was seen with more frequency in heart valve (M6 and M12) and heart (M6, M12 and GAS40). The cytoplasmic staining was considered to be of little toxicological relevance as the cytoplasmic compartment is generally thought to be of limited access to antibodies in vivo. It is the nature of this
5 assay that the cytoplasmic compartment is exposed and available for binding by elements in the mouse antisera. The results of this preliminary screen showed that (1) the immunohistochemical assay was reproducible between runs, (2) there were slight differences between staining of tissues from different donors, (3) staining decreased with increasing dilution of antisera, and (4) the addition of high stringency washes to reduce nonspecific binding of the unpurified mouse antibodies
10 decreased or eliminated staining. This latter observation indirectly confirmed that high affinity antibodies to human tissue elements were not elicited in mice. GAS40 mutants may be tested utilising similar protocols.

While certain embodiments of the present invention have been described and specifically
15 exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention as set forth in the following claims.

TABLE OF SEQUENCE ID NUMBERS USED:

SEQ ID NO	Description
SEQ ID NO: 1	>GAS40 Wild Type nucleic acid sequence (Streptococcus pyogenes)
SEQ ID NO: 2	>GAS40 Wild Type amino acid sequence (Streptococcus pyogenes)
SEQ ID NO: 3	>GAS40 8-9-mers mutant sequence
SEQ ID NO: 4	>GAS40 8- 9-mers mutant pep
SEQ ID NO: 5	>GAS40 multi-mutant sequence
SEQ ID NO: 6	>GAS40 multi-mutant pep
SEQ ID NO: 7	>GAS40 multi-mutant pep plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 8	>coiled-coil domain containing 81 (Homo sapiens)
SEQ ID NO: 9	>UPF0492 protein C20orf94 (Homo sapiens)
SEQ ID NO: 10	>Janus kinase and microtubule-interacting protein 3 (Homo sapiens)
SEQ ID NO: 11	>Mutant version of SEQ ID NO: 8 (Artificial)
SEQ ID NO: 12	>Mutant version of SEQ ID NO: 9(Artificial)
SEQ ID NO: 13	>Mutant version of SEQ ID NO: 10(Artificial)
SEQ ID NO: 14	>heptad repeat of coiled coils (hxxhcxc (SEQ ID NO: 14))(Artificial)
SEQ ID NO: 15	>GAS40 Wild Type Variant (Streptococcus pyogenes)
SEQ ID NO: 16	>8-9 mer mutant of GAS40 Variant
SEQ ID NO: 17	>idealized coiled coil of GAS40 Variant
SEQ ID NO: 18	>idealized coiled coil of GAS40 Variant plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 19	> GAS40 fragment
SEQ ID NO: 20	>8-9 mer mutant of GAS40 fragment
SEQ ID NO: 21	>idealized coiled coil of GAS40 fragment
SEQ ID NO: 22	>idealized coiled coil of GAS40 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 23	>GAS40 Variant 2
SEQ ID NO: 24	>8-9 mer mutant of GAS40 Variant 2
SEQ ID NO: 25	>idealized coiled coil of GAS40 Variant 2
SEQ ID NO: 26	>idealized coiled coil of GAS40 Variant 2 plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 27	> GAS40 Variant 2 fragment
SEQ ID NO: 28	>8-9 mer mutant of GAS40 Variant 2 fragment
SEQ ID NO: 29	>idealized coiled coil of GAS40 Variant 2 fragment
SEQ ID NO: 30	>idealized coiled coil of GAS40 Variant 2 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 31	> GAS40 Variant 3 fragment
SEQ ID NO: 32	>8-9 mer mutant of GAS40 Variant 3 fragment
SEQ ID NO: 33	>idealized coiled coil of GAS40 Variant 3 fragment
SEQ ID NO: 34	>idealized coiled coil of GAS40 Variant 3 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 35	> GAS40 Variant 4 fragment
SEQ ID NO: 36	>8-9 mer mutant of GAS40 Variant 4 fragment
SEQ ID NO: 37	>idealized coiled coil of GAS40 Variant 4 fragment
SEQ ID NO: 38	>idealized coiled coil of GAS40 Variant 4 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 39	> GAS40 Variant 5 fragment
SEQ ID NO: 40	>8-9 mer mutant of GAS40 Variant 5 fragment
SEQ ID NO: 41	>idealized coiled coil of GAS40 Variant 5 fragment
SEQ ID NO: 42	>idealized coiled coil of GAS40 Variant 5 fragment plus modified QLTEELAA (SEQ

	ID NO: 128)
SEQ ID NO: 43	> GAS40 Variant 6 fragment
SEQ ID NO: 44	>8-9 mer mutant of GAS40 Variant 6 fragment
SEQ ID NO: 45	>idealized coiled coil of GAS40 Variant 6 fragment
SEQ ID NO: 46	>idealized coiled coil of GAS40 Variant 6 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 47	> GAS40 Variant 7 fragment
SEQ ID NO: 48	>8-9 mer mutant of GAS40 Variant 7 fragment
SEQ ID NO: 49	>idealized coiled coil of GAS40 Variant 7 fragment
SEQ ID NO: 50	>idealized coiled coil of GAS40 Variant 7 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 51	> GAS40 Variant 8 fragment
SEQ ID NO: 52	>8-9 mer mutant of GAS40 Variant 8 fragment
SEQ ID NO: 53	>idealized coiled coil of GAS40 Variant 8 fragment
SEQ ID NO: 54	>idealized coiled coil of GAS40 Variant 8 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 55	> GAS40 Variant 9 fragment
SEQ ID NO: 56	>8-9 mer mutant of GAS40 Variant 9 fragment
SEQ ID NO: 57	>idealized coiled coil of GAS40 Variant 9 fragment
SEQ ID NO: 58	>idealized coiled coil of GAS40 Variant 9 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 59	> GAS40 Variant 10 fragment
SEQ ID NO: 60	>8-9 mer mutant of GAS40 Variant 10 fragment
SEQ ID NO: 61	>idealized coiled coil of GAS40 Variant 10 fragment
SEQ ID NO: 62	>idealized coiled coil of GAS40 Variant 10 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 63	> GAS40 Variant 11 fragment
SEQ ID NO: 64	>8-9 mer mutant of GAS40 Variant 11 fragment
SEQ ID NO: 65	>idealized coiled coil of GAS40 Variant 11 fragment
SEQ ID NO: 66	>idealized coiled coil of GAS40 Variant 11 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 67	> GAS40 Variant 12 fragment
SEQ ID NO: 68	>8-9 mer mutant of GAS40 Variant 12 fragment
SEQ ID NO: 69	>idealized coiled coil of GAS40 Variant 12 fragment
SEQ ID NO: 70	>idealized coiled coil of GAS40 Variant 12 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 71	> GAS40 Variant 13 fragment
SEQ ID NO: 72	>8-9 mer mutant of GAS40 Variant 13 fragment
SEQ ID NO: 73	>idealized coiled coil of GAS40 Variant 13 fragment
SEQ ID NO: 74	>idealized coiled coil of GAS40 Variant 13 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 75	> GAS40 Variant 14 fragment
SEQ ID NO: 76	>8-9 mer mutant of GAS40 Variant 14 fragment
SEQ ID NO: 77	>idealized coiled coil of GAS40 Variant 14 fragment
SEQ ID NO: 78	>idealized coiled coil of GAS40 Variant 14 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 79	> GAS40 Variant 15 fragment
SEQ ID NO:80	>8-9 mer mutant of GAS40 Variant 15 fragment
SEQ ID NO: 81	>idealized coiled coil of GAS40 Variant 15 fragment
SEQ ID NO: 82	>idealized coiled coil of GAS40 Variant 15 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 83	> GAS40 Variant 16 fragment

SEQ ID NO: 84	>8-9 mer mutant of GAS40 Variant 16 fragment
SEQ ID NO: 85	>idealized coiled coil of GAS40 Variant 16 fragment
SEQ ID NO: 86	>idealized coiled coil of GAS40 Variant 16 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 87	> GAS40 Variant 17 fragment
SEQ ID NO: 88	>8-9 mer mutant of GAS40 Variant 17 fragment
SEQ ID NO: 89	>idealized coiled coil of GAS40 Variant 17 fragment
SEQ ID NO: 90	>idealized coiled coil of GAS40 Variant 17 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 91	> GAS40 Variant 18 fragment
SEQ ID NO: 92	>8-9 mer mutant of GAS40 Variant 18 fragment
SEQ ID NO: 93	>idealized coiled coil of GAS40 Variant 18 fragment
SEQ ID NO: 94	>idealized coiled coil of GAS40 Variant 18 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 95	> GAS40 Variant 19 fragment
SEQ ID NO: 96	>8-9 mer mutant of GAS40 Variant 19 fragment
SEQ ID NO: 97	>idealized coiled coil of GAS40 Variant 19 fragment
SEQ ID NO: 98	>idealized coiled coil of GAS40 Variant 19 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 99	> GAS40 Variant 20 fragment
SEQ ID NO: 100	>8-9 mer mutant of GAS40 Variant 20 fragment
SEQ ID NO: 101	>idealized coiled coil of GAS40 Variant 20 fragment
SEQ ID NO: 102	>idealized coiled coil of GAS40 Variant 20 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 103	> GAS40 Variant 21 fragment
SEQ ID NO: 104	>8-9 mer mutant of GAS40 Variant 21 fragment
SEQ ID NO: 105	>idealized coiled coil of GAS40 Variant 21 fragment
SEQ ID NO: 106	>idealized coiled coil of GAS40 Variant 21 fragment plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 107	>GAS40 with N-terminal leader sequence removed fragment
SEQ ID NO: 108	>8-9 mer mutant of 40withNterminalleadersequenceremoved fragment
SEQ ID NO: 109	>idealized coiled coil of 40withNterminalleadersequenceremoved fragment
SEQ ID NO: 110	>idealized coiled coil of 40withNterminalleadersequenceremoved plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 111	>firstcoiledcoilregion
SEQ ID NO: 112	> 8-9 mer mutant of firstcoiledcoilregion
SEQ ID NO: 113	>idealized coiled coil of firstcoiledcoilregion
SEQ ID NO: 114	>idealized coiled coil of firstcoiledcoilregion plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 115	>secondcoiledcoilregion
SEQ ID NO: 116	> 8-9 mer mutant of secondcoiledcoilregion
SEQ ID NO: 117	>idealized coiled coil of secondcoiledcoilregion
SEQ ID NO: 118	>idealized coiled coil of 40 and 8-9 mer plus modified QLTEELAA (SEQ ID NO: 128)
SEQ ID NO: 119	>wild-type GAS57(Streptococcus pyogenes)
SEQ ID NO: 120	>GAS57 double mutant(D151A, S617A)
SEQ ID NO: 121	>GAS 25 W535F-P427L
SEQ ID NO: 122	T237S_A241V_F1:
SEQ ID NO: 123	T237S_A241V_R1

SEQ ID NO: 124	A688S_S692T_F1
SEQ ID NO: 125	A688S_S692T_R1
SEQ ID NO: 126	A700S_A703S_F1
SEQ ID NO: 127	A700S_A703S_R1

REFERENCES

Baird, RW, MS, Kraus, W Bronze, HR Hill, Veasey LG, and JB Dale. "Epitopes of group A streptococcal M protein shared with antigens of articular cartilage and synovium." *The Journal of Immunology* 146 (1991): 3132-3137.

McNamara, C, AS Zinkernagel, P Macheboeuf, MW Cunningham, V Nizet, and P Ghosh. "Coiled-coil irregularities and instabilities in group A Streptococcus M1 are required for virulence." *Science* 319 (2008): 1405-1408.

O'Hagan, Derek T., ed. *Vaccine Adjuvants: Preparation Methods and Research Protocols (Volume 42 of Methods in Molecular Medicine series).* Humana Press, 2000.

Patterson, MJ. "Streptococcus." In *Medical Microbiology*, edited by S Baron. Galveston (TX): University of Texas Medical Branch at Galveston, 1996.

Powell, Michael F, and Mark J Newman, . *Vaccine Design: The Subunit and Adjuvant Approach.* Plenum Press, 1995.

Remington, Joseph Price. *Remington: The Science and Practice of Pharmacy.* 20th. Edited by Alfonso R. Gennaro. Lippincott Williams and Wilkins, 2000.

CLAIMS:

1. A recombinant streptococcal GAS40 polypeptide which comprises an epitope that elicits opsonic antibodies to Group A Streptococcus and further comprises:

5 (i) at least two substitutions in any of amino acids 235 to 243, at least two substitutions in any of amino acids 684 to 692 and at least two substitutions in any of amino acids 699 to 706; and/or

(ii) a deletion of Lysine at position 208 and amino acid substitutions at amino acid positions 38, 52, 66, 148, 158, 179, 182, 186, 196, 203, 218, 225, 544, 558, 562, 576, 586, 614, 618,

10 621, 665, 672, 683 and 693,

wherein the amino acid positions are numbered according to SEQ ID NO: 2.

2. The recombinant streptococcal GAS40 polypeptide of claim 1 which comprises substitutions at amino acid positions T237, A241, A688, S692, A700 and A703.

15

3. The recombinant streptococcal GAS40 polypeptide of claim 2 which comprises the substitutions T237S, A241V, A688S, S692T, A700S and A703S.

4. The recombinant streptococcal GAS40 polypeptide of any one of claims 1 to 3 which comprises a

20 deletion of Lysine at position 208 and amino acid substitutions A38L, Q52L, T66L, A148L, S158L, A179L, A182L, T186L, A196L, Q203L, Q218L, K225L, A544L, T558L, S562L, K576L, A586L, T614L, A618L, A621L, T665L, A672L, Q683L and T693L

5. The recombinant streptococcal GAS40 polypeptide of claim 1 wherein said polypeptide shares at least 92% identity with any one of SEQ ID NOs: 4, 6, 7, 16, 17, 18, 24, 25 or 26.

25 6. The recombinant streptococcal GAS40 polypeptide of claim 5 wherein said polypeptide comprises a sequence selected from the group consisting of SEQ ID NOs: 4, 6, 7, 16, 17, 18, 24, 25 and 26.

7. A fragment of a recombinant streptococcal GAS40 polypeptide according to any preceding claim wherein the fragment shares at least 97% identity with any one of SEQ ID NOs: 20, 21, 22, 28,

30 29, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 44, 45, 46, 48, 49, 50, 52, 53, 54, 56, 57, 58, 60, 61, 62, 64, 65, 66, 67, 68, 69, 70, 72, 73, 74, 76, 77, 78, 80, 81, 82, 84, 85, 86, 88, 89, 90, 92, 93, 94, 96, 97, 98, 100, 101, 102, 104, 105, 106, 108, 109, 110, 112, 113, 114, 116, 117 and 118.

8. The fragment of claim 7 wherein the polypeptide comprises or consists of any one of SEQ ID NOs: 4, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80, 84, 88, 92, 96, 100, 104, 108,

112, 116, 7, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114 and 118.

9. A polypeptide comprising the recombinant streptococcal GAS40 polypeptide of claims 1 to 6 or the fragment of a recombinant streptococcal GAS40 polypeptide polypeptide of claims 7 to 8.

5 10. An immunogenic composition comprising the recombinant streptococcal GAS40 polypeptide of claims 1 to 6 or the fragment of a recombinant streptococcal GAS40 polypeptide polypeptide of claims 7 to 8.

11. The immunogenic composition of claim 10, wherein said composition comprises one or more additional GAS polypeptides.

10 12. The immunogenic composition of claim 11, wherein said composition comprises a GAS57 polypeptide of SEQ ID NO:120 and a GAS25 polypeptide of SEQ ID NO:121 and optionally a GAS polysaccharide.

13. The immunogenic composition of any one of claims 10 to 12, wherein said composition is a vaccine.

15 14. The immunogenic composition of any one of claims 10 to 13, which comprises an adjuvant.

15. A nucleic acid encoding the recombinant streptococcal GAS40 polypeptide of claims 1 to 6 or the fragment of a recombinant streptococcal GAS40 polypeptide polypeptide of claims 7 to 8.

16. A cell comprising the nucleic acid of claim 15.

20

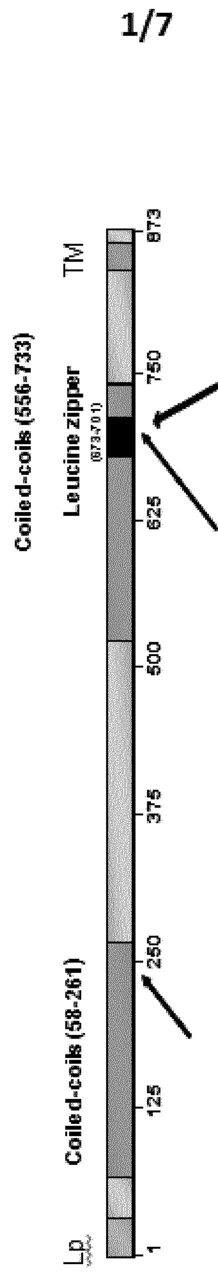


Figure 1

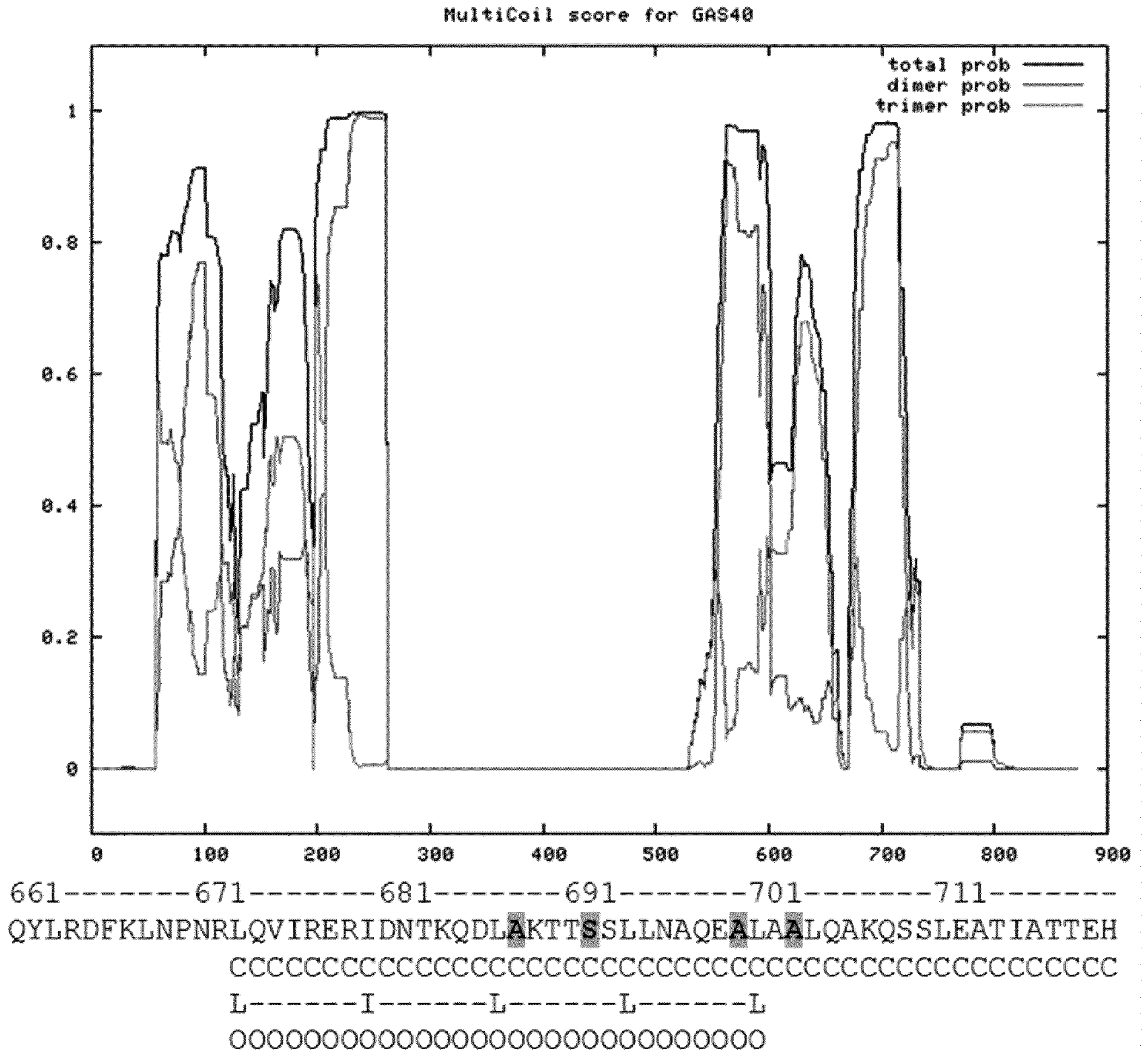


Figure 2

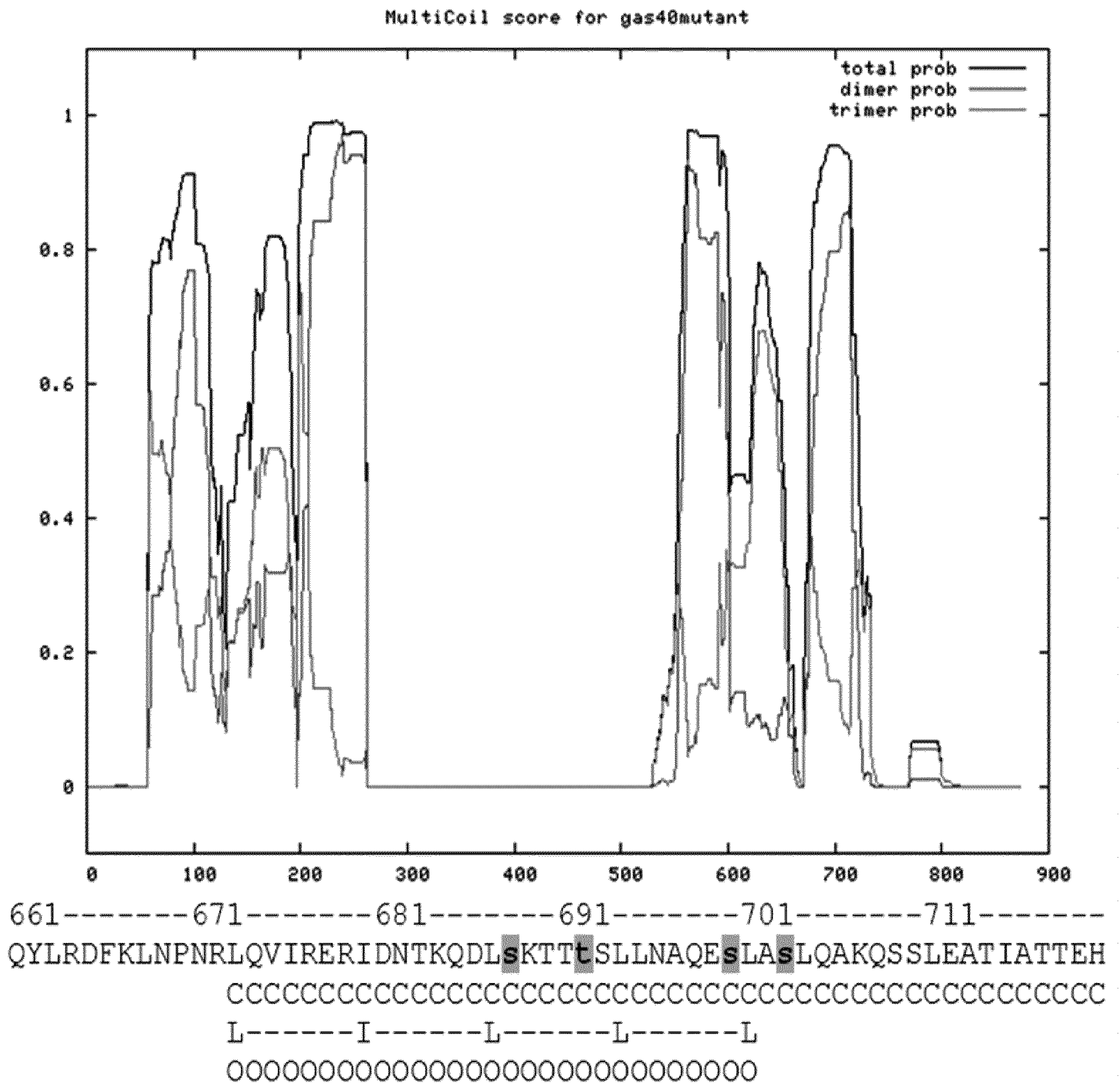


Figure 3

Figure 4

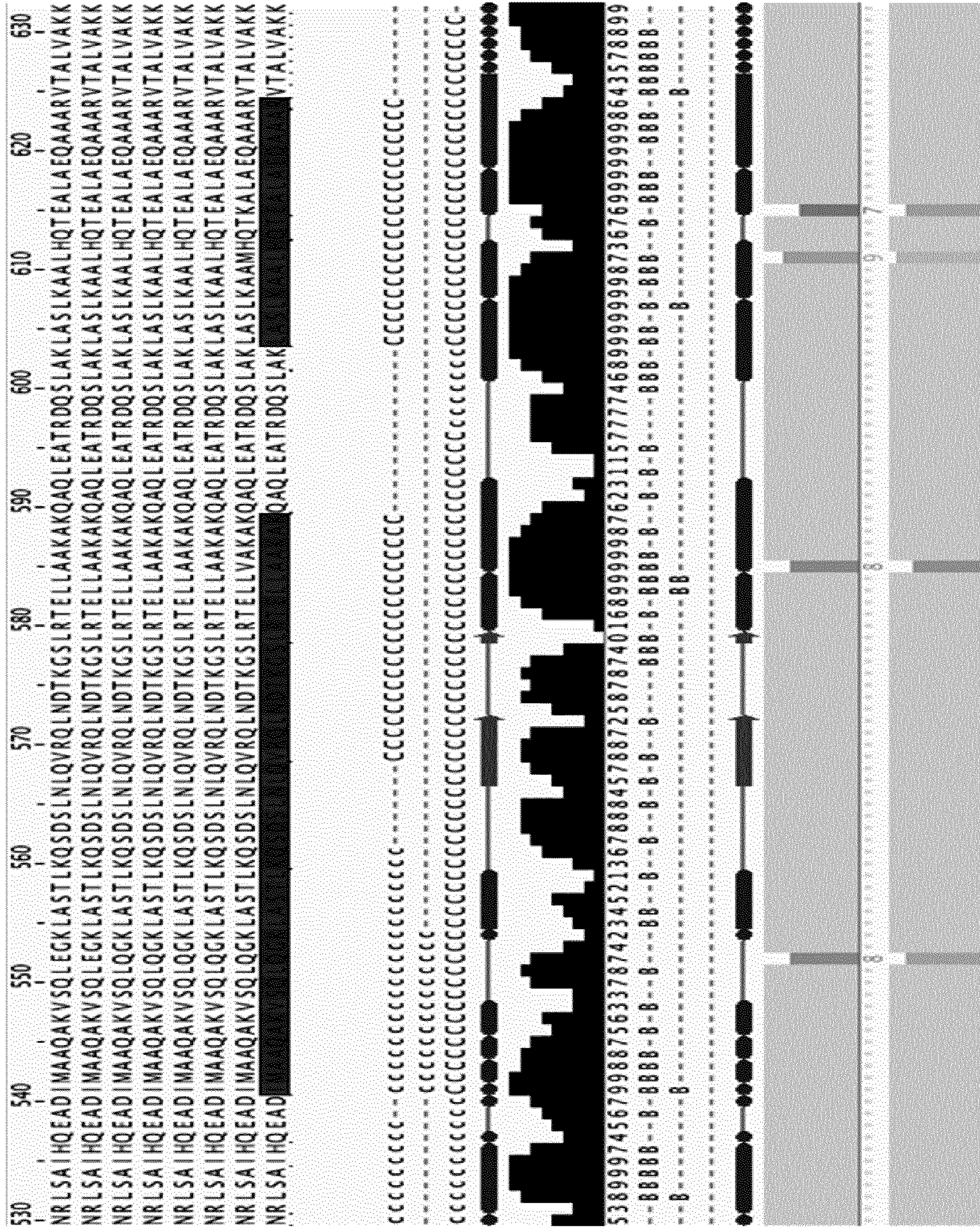


Figure 5

```
650       660       670       680       690       700       71
VIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYR
VIRERIDNTKQDLAKTTSSLLNAQEALAVLQAKQSSLEATIATTEHQLTLLKTLANEKEYR
VIRERIDNTKQDLAKTTSSLLNAQETLAALQAKQSSLEATIATTEHQLTLLKTLANEKEYR
VIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYR
VIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYR
VIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYR
VIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYR
VIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSSLEATIATTEHQLTLLKTLANEKEYR

------------CCCCCCCCCCCCCCCCCCCCCCCCCCCccccccc-------------
-------------------CCCCCCCCCCCCCC--------------------------
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-------

9999973567899873466999999999853788855888547753478888743777764
BB---B--B---BB---BBBB-B--BBBBB-B----B-B-BB-B-B-BBBB-BBB----B-
-B----------------------B--------------------B-BB---------
-------------------------------------------------------------
                         7  8
```

Figure 6

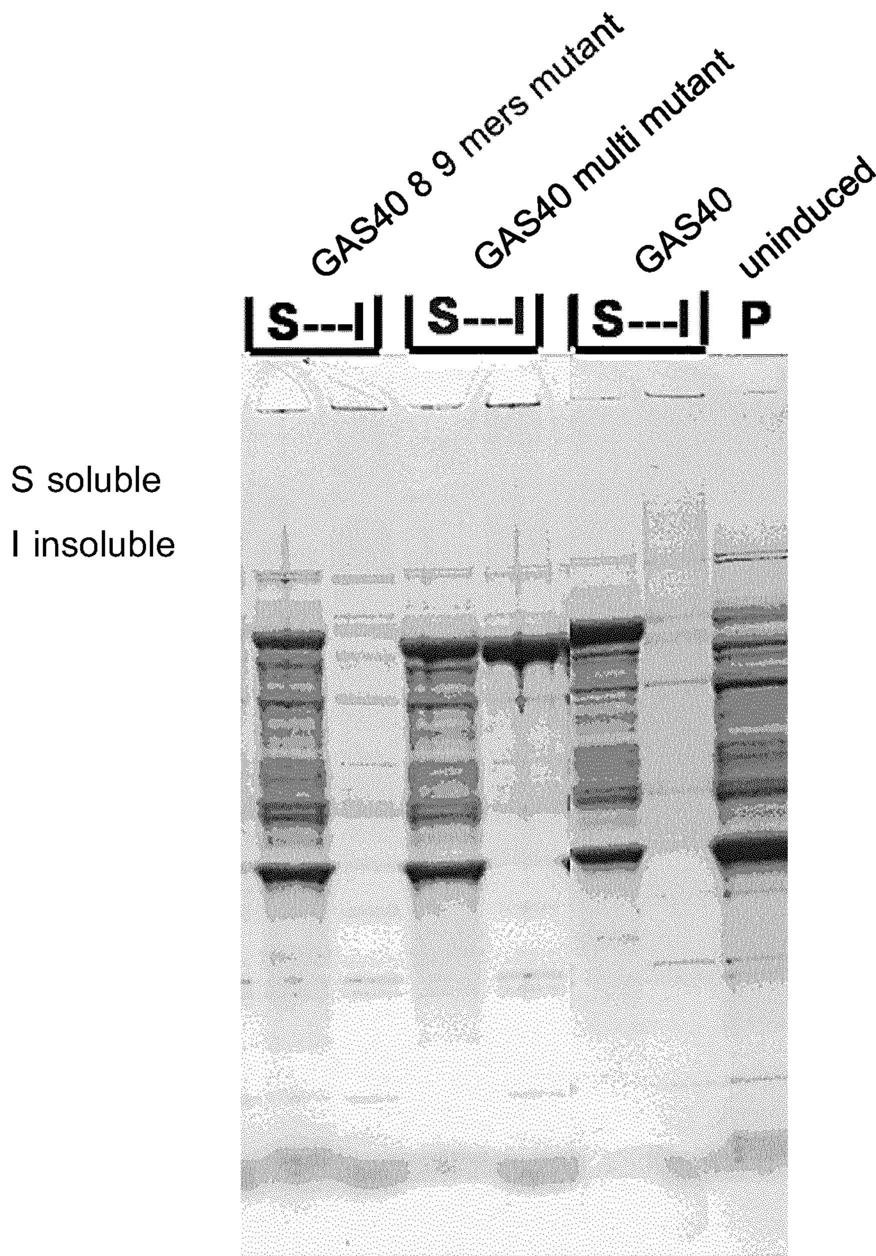


Figure 7