Group A streptococcus (GAS) is a bacterial pathogen responsible for a wide array of disease pathologies in humans. GAS surface M protein plays multiple key roles in pathogenesis, and serves as a target for typing and vaccine development. In this review, we have compiled GAS epidemiological studies from several countries around the world to highlight the consequences on the theoretical efficacy of two different M protein-based vaccine strategies.

**Keywords:** group A streptococcus • M protein • M protein vaccine • Streptococcus pyogenes

Group A streptococcus (GAS) colonizes and infects the upper respiratory tract and skin of humans, resulting in a broad spectrum of diseases. As elegantly described by Carapetis et al., the burden of GAS-related diseases predominantly occurs in settings of poverty where systems for collection of accurate disease-burden data are usually absent [1]. The most recent estimates suggest that GAS is responsible for more than 500,000 deaths per year [1]. This places GAS among the major human pathogens, with attributable mortality by a single species of pathogen only exceeded by HIV, Mycobacterium tuberculosis, Plasmodium falciparum and Streptococcus pneumoniae [1]. The most common GAS disease following colonization of the respiratory tract is pharyngitis. More than 600 million cases are estimated to occur each year [1]. Similarly, it is estimated that more than 111 million children under the age of 15 years in developing countries have pyoderma of any etiology [1]. While pharyngeal and pyodermal GAS infections are noninvasive and usually self-limiting, they too require significant healthcare attention. The costs of antibiotic treatment and time away from school or work for children and parents are a formidable financial burden [2]. After disseminating from primary colonization sites to otherwise sterile compartments of the body, GAS can cause life-threatening invasive diseases such as necrotizing fasciitis and streptococcal toxic shock syndrome (TSS). Invasive GAS infections kill at least 163,000 people each year [3]. The incidence of invasive diseases has appeared to have increased over the last 20 years [3,4]. Most importantly, GAS has the ability to cause the nonsuppurative sequelae: acute poststreptococcal glomerulonephritis, rheumatic fever (RF) and rheumatic heart disease (RHD) [5]. While most industrialized countries now report a low incidence of RF and RHD [7], these diseases are endemic in developing countries and indigenous groups within developed countries [6,9,10]. Globally, it is estimated that RHD is responsible for 233,000–294,000 deaths per year [1]. In addition to the greater disease burden in developing countries, the spectrum of GAS diseases also differs between developed and developing countries. For example, GAS skin infections are more prevalent in tropical developing countries, whereas infections of the throat are common in industrialized countries [11].

The mechanism behind the different disease phenotypes and disease severity in different locations around the world is still undefined. It is believed that many factors, including the genetic susceptibility of the host, the environmental factors and ease of access to healthcare, play a role in disease progression and severity. However, there is no doubt that differences in the repertoire of virulence factors play a critical role in the pathogenesis of individual GAS
strains. In particular, the genetic diversity of the highly expressed M protein is believed to play an important role in the observed epidemiological differences.

The M protein
Discovery
The M protein is believed to be the major virulence determinant of GAS [12]. Discovered by Rebecca Lancefield in the early part of the 20th Century, the protein is found on the surface of all GAS isolates. M protein or related M-like proteins can also be found in the closely related Streptococcus dysgalactiae subsp. equisimilis (group G streptococcus [GGS]). Streptococcus equi spp. [13] and Streptococcus iniae [14], the latter two being pathogens of farm or domestic animals and fish. Within the GAS genome, the emm gene is found as part of the mga regulon, which also includes a transcriptional regulator (mga), the emm gene, emm-like genes and scpA [15,16]. In GGS, however, the genetic architecture is somewhat different [17]. These findings are consistent with evidence [18] suggesting that the emm gene was once part of a large mobile genetic element that was acquired by progenitors of GAS and GGS prior to sub specification, that is, the ancient parental emm gene evolved outside of GAS. Subsequent selection of distinct emm-types is probably driven by immune pressure targeting the N-terminus. Recombination between the emm and emm-like genes within a strain, as well as interstrain emm gene transfers [19,20], have all contributed to the diversity of modern emm genes we see today.

Typing
The M protein exists as a coiled-coil dimer [21] extending from the surface of the bacteria to beyond the peptidoglycan layer. The protein is divided into a number of specific regions that have unique sequence, conformational and functional attributes (Figure 1). The amino acid sequence of the N-terminal region of the mature M protein is hypervariable, and antibodies raised against this region of the M protein are serotype specific and provide protection against that strain only. This had been the basis of the typing scheme for GAS for decades [22,23]. More recently, serotyping has been replaced by nucleotide sequence-based typing methods that target the DNA encoding for the hypervariable regions. Using this molecular typing method, more than 200 GAS emm-types have been reported [24,25].

Why is the M protein so important?
The major function of the M protein is to prevent opsonization of GAS by phagocytes of the host. The M protein achieves this by inhibiting the deposition of active complement molecules on the bacterial surface. Intriguingly, the M protein has developed a number of independent strategies to achieve this goal. The redundancy in function and evolution of unique anticompment strategies by the different M proteins highlights the importance of this activity for the survival of GAS [26]. A common feature of all these mechanisms is the recruitment of host proteins to the bacterial surface that either actively inhibit complement activation or passively prevent complement deposition on the bacterial surface.

For example, FHI-1 and C4BP, host proteins that regulate complement activation, bind to regions of the M protein [26]. These molecules may also compete with serotype-specific antibodies for binding to the hypervariable region. A second strategy utilized by the M protein is 'passive' evasion of complement and other immune molecules. This is achieved by binding of host plasma proteins to the bacterial surface, thereby preventing the pathogen from being recognized for complement deposition. Perhaps the best examples for this are fibrinogen and albumin, which bind to the B- and C repeats of the M protein, respectively [27-31]. In the absence of antibody, fibrinogen–M protein complexes can reduce complement deposition by the classical pathway, possibly through inhibition of the convertase [32,33]. Both fibrinogen and albumin also block antibody binding to the conserved repeat region (CRR) [34], preventing antibody-mediated opsonization. The latter observation highlights the importance of the serotype-specific regions in promoting antibody-mediated opsonization in natural infections. Recent findings have also demonstrated a role of anti-M protein antibodies in disease pathogenesis by promoting a pathologic inflammatory response as well as platelet aggregation [35,36].

An additional role for the M protein in GAS virulence is in host cell adherence and cellular invasion. The adhesion of GAS to epithelial cells is the first stage of bacterial colonization of both the throat and/or the skin, and is an essential step to GAS infection. Primarily, it is the interaction between the M protein and the keratinocyte cell marker CD46 and fibronectin that promote GAS adherence [37]. Although GAS is considered an extracellular pathogen, recent evidence confirms that it can invade and reside within host cells [37]. The role of the M protein in this process has been confirmed by numerous researchers and is also believed to be mediated through the binding of fibronectin [38-40]. Whether this ability to invade host cells is a strategy to evade host defense mechanisms or a prerequisite for invasive disease is still to be established. In addition to understanding GAS pathogenesis, determining the mechanism the M protein employs to promote cell adherence and invasion may also provide information for vaccine design based on anti-attachment/dissemination strategies.

Roles of M protein in disease
In addition to its role in preventing opsonophagocytosis, the M protein is thought to be directly involved in a number of streptococcal diseases. Chiefly, the M protein is thought to play an integral role in RF and RHD. RF and RHD are autoimmune diseases in which antibodies and immune cells raised against GAS are redirected to attack host proteins, cells and tissues. In the case of RHD, these immune molecules attack heart tissue [41]. These diseases occur several weeks after a streptococcal infection. The postinfectious nature of primary cases of RF/RHD makes antibiotic treatment of the disease ineffective. However, ongoing antibiotic prophylaxis is recommended as a measure to prevent the future re-occurrence of RF/RHD. Molecular mimicry between the M protein and several host proteins, including myosin, keratin and vimentin, has been postulated as a leading cause of RHD [42,43]. Similar to the M protein, these proteins
also have a coiled-coil structure [51,44]. The epitopes that contribute to RF/RHF have been mapped to the central region of the M protein [45]. Thus, anti-GAS vaccines should not include this part of the protein. In fact, there is some evidence from an early vaccine study that administration of a whole M protein vaccine contributed to the development of RF in volunteers [46]. The results from this study obviously had a major impact on later GAS vaccine design.

Historically, certain GAS types have been classed as 'rheumatogenic' due to their association with RF (e.g., emm-types 1, 3, 5, 6, 18, 19 or 24) [10,12]. These strains are predominantly found in the throat rather than the skin. Thus, RF/RHD is normally considered to follow throat infections. However, more recent epidemiological studies in geographic regions where RF/RHD and streptococcal infections are endemic have failed to find significant numbers of these rheumatogenic strains [47,48]. In fact, the data emerging from the epidemiological studies in tropical regions challenge the accepted link between 'rheumatogenic GAS', pharyngitis and RF/RHD. Epidemiological studies in these regions tend to highlight the diversity of circulating emm-types. Rheumatogenic emm-types appear to be absent or rare. In these locations, the incidence of GAS impetigo is high and possibly involved directly or indirectly in the development of RF in some communities [48-51]. Other studies describe the recovery rate of GGS rather than GAS in the throat in these settings [52]. At present, a link between GGS and RF/RHD is tenuous but has been suggested in a number of studies [52,53].

A preferential association between certain emm-types and the development of invasive diseases has also been described. As an example, emm 1, as well as emm 3, have been predominately associated with invasive infections [54,55]. Moreover, a recent study demonstrated that the M1 protein itself is a superantigen [56] hence possibly explaining the particular predominance of emm-type 1 among TSS cases. Moreover, multiple horizontal gene transfer events might also have contributed to the emergence of successful GAS clones [57,58]. However, strains expressing major virulence genes are also commonly found in

**Figure 1. M proteins.** Individual M proteins form coiled coils, which extend from the bacterial surface. Two coils wind around each other, forming a dimer that is stabilized by intermolecular interactions between amino acids in both strands. The hypervariable N-terminal region, lacking helical structure, is followed by the A, B, C and D repeat regions. Different regions of the protein also bind a number of different host molecules, assisting *Streptococcus pyogenes* to avoid opsonophagocytosis. FHL: Factor H-like protein.
GAS isolated from benign infections and carriage [54, 59, 60]. The prevalence of these enm-types in invasive diseases might simply reflect the prevalence of these types in the circulating GAS population [61], rather than increased virulence of these organisms. This diverse distribution of enm-types among strains from a broad spectrum of clinical sources does not enable simple causal links to be established between enm-types and specific pathologies.

M protein molecular epidemiology

The aim of the present section is to describe our current knowledge of the geographic distribution of circulating GAS enm-types. Unfortunately, epidemiological information is unavailable or incomplete for many countries. The parameters of specific epidemiological surveys (e.g., invasive infections only, pharyngitis only, carriage only or all clinical manifestations) can skew apparent enm-type distribution data in specific locations. Other criteria, such as the methods used to identify cases, the characteristics of the population studied (in terms of age, setting, size and inclusion criteria) and the length of the study, could also affect the apparent enm-type distribution in individual studies. Changes in epidemiology over time are not available for many countries. In order to overcome these problems and still allow strain comparison at a global level, we have developed and applied guidelines for the inclusion of epidemiological studies in our assessment. As highlighted in the previous section, the link between a defined enm-type and pathology is not universal. We therefore assume that the concept of the circulating enm-types, although clearly imperfect, is probably the best strategy for understanding the global picture. Hence, we have included all epidemiological studies to obtain a representation of the circulating enm-types, irrespective of their associated pathology. As the focus of this review is to emphasize the GAS M-protein epidemiology relative to vaccine development, we believe that the compilation of studies on a national level is mandatory as vaccination recommendations are most often also developed on this level. However, for reasons of clarity, we have combined data from 11 European countries involved in the StrepFam project [62] with three other Western European countries (Norway, Spain and Austria). For this review, we only considered epidemiological studies in which GAS isolates were recovered after 1 January 1988, assuming that this time window (1988–2007) probably represents contemporary GAS M-protein epidemiology. Studies that relied solely on serological typing methods or on hybridization diagnostics without a complementary sequencing step were excluded as they probably underestimate the diversity of circulating enm-types. Some studies could not be included owing to missing information. We have used the Simpson’s reciprocal index (SRI) to measure GAS strain diversity. The SRI takes into account both species richness and evenness. In this context, richness is a measure of the number of enm-types present in a defined area, while evenness represents the relative abundance of these enm-types. A SRI value of 1 corresponds to a theoretical situation where only one enm-type has been recovered, representing the minimal diversity possible. A value of 250 would correspond to a situation where all known enm-types would be recovered at a similar proportion, illustrating the maximal diversity of enm-types possible.

To examine whether the diversity of enm-types differed depending on the clinical manifestation being assessed, we compared the diversity of enm-types recovered from either pharyngitis, cutaneous or invasive infections in Fiji, a country where a high diversity of enm-type has been described [63]. Table 1 shows that, irrespective of the associated clinical manifestations, the SRI were similar [23–29]. This was also observed when comparing the distribution of enm-types from the American GAS pharyngitis and invasive infections surveillances [55, 64], with both clinical pathologies displaying a low diversity of enm-types with a clear predominance of a few types. Moreover, seven of the ten most prevalent enm-types were identical, irrespective of clinical manifestations [55, 64]. These comparisons suggest that the enm-type diversity observed within a clinical manifestation probably represents the overall enm-type diversity in that geographic location.

Using the SRI, we observed that the diversity in GAS strains around the world range from eight to 50 and presented a bimodal distribution (Figure 2). For the purpose of this review, we have therefore defined a low diversity of enm-types as corresponding to a SRI ranging from 8 to 13 and a high diversity of enm-types with a SRI of 27–50.

Figure 3 clearly shows that our knowledge of GAS epidemiology of some areas of the world where streptococcal diseases are endemic is sadly lacking. This is particularly true of South America, Africa and some locations in Asia (Figure 3). The predominance of GAS

Table 1. Comparison of enm-type diversity according to the associated pathology in Fiji.

<table>
<thead>
<tr>
<th>Associated pathology</th>
<th>Isolates (n)</th>
<th>enm-types (n)</th>
<th>Simpson’s reciprocal index (95% CI)</th>
<th>Cumulative frequency of the ten most prevalent enm-types (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharyngitis</td>
<td>61</td>
<td>36</td>
<td>25.3 (19.6–35.4)</td>
<td>52% (39–66)</td>
</tr>
<tr>
<td>Impetigo</td>
<td>379</td>
<td>52</td>
<td>23.2 (20.3–27.1)</td>
<td>55% (42–68)</td>
</tr>
<tr>
<td>Invasive infections</td>
<td>55</td>
<td>37</td>
<td>27.7 (21.4–39.3)</td>
<td>49% (36–62)</td>
</tr>
<tr>
<td>Total</td>
<td>495</td>
<td>66</td>
<td>28.8 (25.5–33)</td>
<td>49% (44–53)</td>
</tr>
</tbody>
</table>

The Simpson’s index (D) has been calculated with the formula: D = Σ (n/N)^2, where n is the total number of isolates from one enm-type and N is the total number of isolates of all the enm-types recovered in an area (198). Simpson’s reciprocal index (1/D) has been used in this review because it better illustrates our data than the classical Simpson’s index of diversity (1-D). While the value of the Simpson’s index of diversity only varies from 0 to 1, the value of the Simpson’s reciprocal index varies in our case from 1 to 250. Confidence intervals (in brackets) have been calculated as previously described [120]. Based on [63] and [Steer A, Peas Comm].
research in Western countries clearly shows that the epidemiological information is not in proportion to the burden of disease, which mainly lies in developing nations. Table 2 shows that some emm-types were frequently recovered from nearly all the countries involved. emm-type 1 and 12, for example, were among the ten most prevalent emm-types in most countries.

When considering both the GAS M-protein diversity and the relatedness of circulating emm types within the different countries, as seen in both Table 2 and Figure 3, it becomes possible to categorize them into three different groups. The first group includes Canada, the USA, Mexico, Western Europe, Korea and Japan. The emm-type diversity observed in these countries is relatively low (SRI 8–13) with only a few predominant emm-types. Moreover, the predominant emm-types are quite consistent between each of the countries, although some geographical differences can be noted. For example, emm-types 87, 83 or 81 are within the top ten most frequent emm-types recovered in Europe; however, they are less frequently (emm 87) or rarely (emm 83 or emm 81) recovered in North America.

The second group, which includes countries from Eastern Europe such as Serbia and Poland, but also interestingly China, also display a relatively low diversity of emm-types (SRI 10–12) but the predominant circulating emm-types are different to those found in the countries of group 1. Chinese data remain to be confirmed with larger epidemiological collections, but the diversity of circulating emm-types is apparently as low as in Europe. However, emm-types 60, 63, 58, 8 and 95 seem to be predominant in this geographical region, while they are not frequently recovered in the other regions.

The third group of countries includes Brazil, Ethiopia, Israel, India, Nepal, Australia and Fiji. The GAS strains recovered from these countries display a high diversity of emm-types (SRI 27–50). In comparison with groups 1 and 2, no real predominant emm-types can be observed, as shown by the low cumulative frequency of the ten most prevalent emm-types (34–49%). In addition, the emm-types circulating in these countries are quite different to the ones present in group 1 and 2 but also between each of the countries within group 3. For example, only a third of the emm-types recovered in Ethiopia are present in Fiji.

**Figure 2. Bimodal distribution of Simpson's index among 16 countries.** The value of the Simpson's reciprocal index (SRI) is shown per country, indicating the presence of two groups of countries with 'low' and 'high' SRI, respectively. Confidence intervals have been calculated as previously described [16].
The selective pressures driving these differences in terms of the diversity of circulating emm-types remain to be found. Factors such as climate, primary care settings, antibiotic treatment, schooling, socioeconomic status, population immunity and host genetic susceptibilities might certainly influence the circulating emm-types. However, when looking at the data from all three of the defined groups, none of these factors seems sufficient to explain the observed differences.

M protein-based vaccines
As the major surface-expressed virulence factor and the determinant of naturally acquired immunity against GAS, the M protein has been the major focus of vaccine design [65,66]. Murine studies in the early 1900s highlighted the vaccine potential of the M protein [67], which was later confirmed by whole-protein vaccinations of both monkeys and then humans [68–70]. However, as the understanding of M protein-induced immunity progressed, it soon become evident that vaccine development would not be without hurdles. Primarily, the variability found at the N-termini of the M protein, responsible for type-specific immunity, would hinder the development of a vaccine targeting all circulating serotypes [24,71,72]. Second, the presence of potentially human cross-reactive epitopes within the M protein prevented the development of the whole M-protein vaccine [73–75]. This initiated a large research effort focused on the identification of non-M-protein vaccine candidates [76,77]. However, it also resulted in the establishment of two main foci of M-protein vaccine design, each addressing safety and efficacy concerns. One strategy was directed solely towards the variable highly immunogenic N-terminus of the M protein, while the other focused on the conserved CRR of the M protein.

N-terminal vaccine design
The N-terminal approach to M protein-based vaccine design is arguably the most progressed GAS vaccine strategy to date, and is currently being assessed in human clinical trials [78,79]. Led by Dale, this strategy aims to create a multivalent vaccine incorporating the N-terminal subunits of multiple GAS serotypes in a single preparation. The initial concept was established with a tetravalent vaccine in which the N-terminal subunits from four different M serotypes were expressed recombinantly as a single-protein antigen [80]. Immunogenicity of each of the included subtypes was further optimized in a hexavalent construct in which the
M-type present at the C-terminus of the recombinant protein was duplicated at the N-terminus to increase its immunogenicity [81]. Both the tetra- and hexavalent vaccines successfully induced antibodies against each of the included M types, which resulted in opsonophagocytosis of the corresponding strains [80–82].

The N-terminal multivalent candidate that is currently in Phase II clinical trials is the 26-valent vaccine [83]. The 26-valent vaccine is comprised of four recombinant proteins, each containing the N-terminal domains of either six or seven different GAS serotypes [84]. A protective epitope from the GAS surface antigen, SPA, has also been included. This antigen also induces opsonic antibodies and was included in the vaccine to enhance cross-protection. Cross-reactivity within the subtypes of vaccine emm-type is usually the rule [85], although some exceptions have been noted [86,87]. Vaccine-induced cross-reactivity with nonvaccine emm-types may possibly occur [Batzios M, Jr., Comm.]. The efficacy of the 26-valent vaccine has been tested in rabbits with the human-comparable adjacent alum. Type-specific antibodies were induced to each of the included M-protein antigens. Combining data from two independent functionality assays, 24 out of the 26 serotypes tested were opsonized by the 26-valent specific sera, suggesting that the vaccine-specific response was functional [84]. Phase I clinical trials assessing the safety and immunogenicity of the 26-valent vaccine in 30 healthy volunteers produced promising results [79]. Each of the incorporated N-terminal peptides was immunogenic and, in contrast to the studies in rabbits, GAS strains representing all of the included M types were opsonized by the 76-valent-specific human sera in in vitro assays. In addition, no human tissue-reactive antibodies were detected. These results supported its progression into Phase II assessment.

The rationale behind the choice of emm-types included in the 26-valent vaccine was the reduction of disease and economic burden associated with invasive disease and uncomplicated pharyngitis in Northern America and Western Europe [84]. In addition, serotypes historically associated with RF, such as M19 and M24, were also included as a precaution for the emergence of such strains under the selective pressure of a multivalent vaccine [80]. At the time of design, the 26-valent vaccine was predicted to provide protection against 85–90% of M types associated with pharyngitis cases and invasive diseases in the USA and Europe [89].

When assessed on a global scale, however, the level of M-type coverage provided by the current 26-valent vaccine is much lower than predicted in the USA. As described in the previous section and illustrated in Figure 1, there is a difference observed in both GAS strain diversity and prevalence in different countries. These differences would most probably impact vaccine efficacy. The level of coverage offered by the 26-valent vaccine is expected to be highest in group 1 countries with a low SR1: 68–86% of M types may be covered. The lower level of coverage predicted for Western Europe, Korea and Japan (~70%) is attributed to geographical differences in the circulating emm-types. As an example, although emm-types included in the multivalent vaccine, such as 1, 3 and 12, are highly prevalent in Western Europe, other emm-types, such as 87, 83 and 81, which are not included in the vaccine, are also prevalent.

Contrasting with group 1, the group 2 and 3 countries are not predicted to be significantly covered by the 26-valent vaccine. Group 2 countries, including Eastern Europe and China, mimic the limited strain diversity seen in group 1. However, the expected lack of vaccine coverage in these countries is attributed to the presence of different circulating emm-types. Hence, while the current 26-valent vaccine may not be suitable for group 2 countries, their relatively low emm type diversity may support the development of their own group or country-specific multivalent N-terminal vaccine.

Group 3 countries clearly display a higher diversity of circulating emm-types with very little evidence of predominant strains [19,63,90,91]. Consequently, the coverage of a multivalent vaccine is predicted to be too low (between 21 and 57%) to have an efficient impact on the disease burden in these regions. There is also very little crossover between the prevalent emm-types in group 3 countries and that of the other groups. Furthermore, there is little consistency between the countries within group 3. One exception, however, is Israel, which, although having a high level of strain diversity (Simpson’s index [51]: 32), is predicted to have a higher level of coverage by the 26-valent vaccine (57%) compared with the other group 3 countries (21–36%). However, the overall diversity of emm-types in Israel is still high compared with that of groups 2 and 1. It seems that an N-terminal-based multivalent vaccine approach would probably not be an effective option.

In addition to high strain diversity observed in the group 3 countries, an association between certain emm-types and particular disease propensities is unclear. As previously discussed, no obvious rheumatogenic or invasive types that may be specific targets for a multivalent vaccine are apparent [92]. This has been further complicated by observations of high pharyngeal carriage of human GGS compared with GAS in communities where RF and RHD are endemic [52]. Although GGS is generally thought of as a commensal organism, it has been suggested that these streptococci contribute significantly to disease burden, and possibly even to RF/RHD [93–95]. GGS also expresses the M protein; however, their N-terminal sequences are different again to that displayed by GAS, hence would not be covered by current N-terminal vaccines [96]. This evidence clearly suggests that an N-terminal M protein approach to GAS vaccine design would require tailoring the vaccine for different geographical locations. While this may be a feasible strategy for group 1 and 2 countries, it would be quite difficult for the group 3 countries; however, the exact cost:benefit ratio for tailoring the vaccine to the specific groups would need to be determined by independent economic analysis.

An additional concern of the multivalent approach to GAS vaccination is the threat of serotype replacement. If the 26-valent vaccine was to be released in the communities it was designed for, it could provide an immunological environment in the host that allows for the emergence of new virulent emm-types or an increased prevalence of serotypes that were not included in the multivalent vaccine. This situation is analogous to the release of the pneumococcal seven-valent vaccine. Experiences with this vaccine may inform us as to what we may expect upon release of a multivalent GAS vaccine [97,99]. The multivalent pneumococcal vaccine combines seven serotype antigens that cover 70–88% of invasive pneumococcal
<table>
<thead>
<tr>
<th>Country</th>
<th>Associated pathology</th>
<th>Isolates (n)</th>
<th>Year of isolation</th>
<th>emm-types (n)</th>
<th>Simpson’s reciprocal index (95% CI)</th>
<th>Cumulative frequency of the ten most prevalent emm-types (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>Pharyngitis</td>
<td>1434</td>
<td>2000–2007</td>
<td>33</td>
<td>9.2 (8.7–9.8)</td>
<td>88% (86–90)</td>
</tr>
<tr>
<td>USA</td>
<td>Pharyngitis</td>
<td>7040</td>
<td>2000–2007</td>
<td>56</td>
<td>9.5 (9.3–9.8)</td>
<td>89% (88–89)</td>
</tr>
<tr>
<td>Mexico</td>
<td>Pharyngitis, invasive Others’</td>
<td>473</td>
<td>1991–2000</td>
<td>31</td>
<td>10.5 (9.3–12.1)</td>
<td>87% (79–86)</td>
</tr>
<tr>
<td>Brazil</td>
<td>Pharyngitis</td>
<td>128</td>
<td>2004</td>
<td>48</td>
<td>27.6 (22.3–36.1)</td>
<td>49% (41–58)</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Pharyngitis, impetigo invasive Sequelae Throat carrier</td>
<td>299</td>
<td>1990–2005</td>
<td>90</td>
<td>50.1 (43.4–59.3)</td>
<td>34% (28–39)</td>
</tr>
<tr>
<td>Western Europe</td>
<td>Invasive Pharyngitis impetigo Others’</td>
<td>4820</td>
<td>1988–2006</td>
<td>107</td>
<td>12.7 (12.2–13.3)</td>
<td>73% (71–74)</td>
</tr>
<tr>
<td>Serbia</td>
<td>Pharyngitis, impetigo invasive Others’</td>
<td>145</td>
<td>2001–2007</td>
<td>31</td>
<td>12.2 (10.3–15.1)</td>
<td>70% (73–86)</td>
</tr>
<tr>
<td>Poland</td>
<td>Invasive</td>
<td>41</td>
<td>1997–2005</td>
<td>23</td>
<td>10.3 (6.8–20.7)</td>
<td>68% (54–82)</td>
</tr>
<tr>
<td>Israel</td>
<td>Pharyngitis, impetigo invasive Others’</td>
<td>819</td>
<td>1996–2005</td>
<td>71</td>
<td>32.5 (29.6–36)</td>
<td>43% (40–47)</td>
</tr>
<tr>
<td>India</td>
<td>Pharyngitis, impetigo Sequelae Throat carrier</td>
<td>313</td>
<td>2000–2007</td>
<td>88</td>
<td>43.4 (37.2–52.1)</td>
<td>38% (33–43)</td>
</tr>
<tr>
<td>Nepal</td>
<td>Impetigo, Throat carrier</td>
<td>120</td>
<td>1998–1999</td>
<td>45</td>
<td>30.5 (25.1–38.8)</td>
<td>46% (37–55)</td>
</tr>
<tr>
<td>China</td>
<td>Pharyngitis, impetigo invasive infections Sequelae Throat carrier</td>
<td>261</td>
<td>1995–2005</td>
<td>46</td>
<td>11.3 (9.5–14)</td>
<td>72% (67–78)</td>
</tr>
<tr>
<td>South Korea</td>
<td>Pharyngitis, Throat carrier</td>
<td>676</td>
<td>2001–2006</td>
<td>31</td>
<td>11.0 (10.1–12.2)</td>
<td>82% (79–85)</td>
</tr>
<tr>
<td>Japan</td>
<td>Pharyngitis, invasive infections Others’</td>
<td>712</td>
<td>1996–2006</td>
<td>36</td>
<td>8.0 (7.2–9)</td>
<td>85% (82–87)</td>
</tr>
<tr>
<td>Fiji</td>
<td>Pharyngitis, Throat carrier, impetigo invasive infections</td>
<td>535</td>
<td>2005–2007</td>
<td>67</td>
<td>30.7 (27.4–34.9)</td>
<td>46% (42–51)</td>
</tr>
<tr>
<td>Australia</td>
<td>Pharyngitis, impetigo Throat carrier</td>
<td>547</td>
<td>1990–2006</td>
<td>69</td>
<td>34.2 (30.3–39.2)</td>
<td>42% (38–46)</td>
</tr>
</tbody>
</table>

*Others’ includes (depending on the studies): ears, vagina, cervix, vulva, catheter, eye and urine.
*Underlined emm-types are included in the 26-valent vaccine candidate.
### Table 2. Group A streptococcus M-protein epidemiology (cont.).

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Theoretical coverage (95% CI)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1</td>
<td>28</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>77</td>
<td>89</td>
<td>6</td>
<td>75</td>
<td>84% (83–86)</td>
<td>[64]</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>28</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>75</td>
<td>89</td>
<td>77</td>
<td>84% (83–85)</td>
<td>[64]</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>75</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>77</td>
<td>77</td>
<td>9</td>
<td>86% (83–89)</td>
<td>[121]</td>
</tr>
<tr>
<td>22</td>
<td>53</td>
<td>49</td>
<td>58</td>
<td>83</td>
<td>8</td>
<td>44</td>
<td>59</td>
<td>12</td>
<td>33</td>
<td>35% (27–43)</td>
<td>[11]</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>3</td>
<td>74</td>
<td>st62</td>
<td>18</td>
<td>25</td>
<td>st463</td>
<td>5</td>
<td>28</td>
<td>36% (31–42)</td>
<td>[51,122]</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>3</td>
<td>89</td>
<td>12</td>
<td>87</td>
<td>4</td>
<td>83</td>
<td>81</td>
<td>77</td>
<td>70% (69–71)</td>
<td>[123–127]</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>1</td>
<td>58</td>
<td>3</td>
<td>4</td>
<td>50</td>
<td>28</td>
<td>53</td>
<td>82</td>
<td>58% (50–66)</td>
<td>[128]</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>81</td>
<td>44</td>
<td>85</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>28</td>
<td>51% (36–66)</td>
<td>[129]</td>
</tr>
<tr>
<td>1</td>
<td>81</td>
<td>14</td>
<td>89</td>
<td>5</td>
<td>28</td>
<td>4</td>
<td>12</td>
<td>18</td>
<td>75</td>
<td>57% (54–61)</td>
<td>[130]</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>49</td>
<td>74</td>
<td>81</td>
<td>11</td>
<td>71</td>
<td>77</td>
<td>42</td>
<td>75</td>
<td>23% (19–28)</td>
<td>[90,131]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[Brahmari, P., Pers. Comm.]</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>44</td>
<td>71</td>
<td>79</td>
<td>82</td>
<td>11</td>
<td>73</td>
<td>77</td>
<td>100</td>
<td>21% (14–28)</td>
<td>[132]</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>60</td>
<td>63</td>
<td>4</td>
<td>58</td>
<td>8</td>
<td>95</td>
<td>18</td>
<td>28</td>
<td>48% (42–54)</td>
<td>[133–135]</td>
</tr>
<tr>
<td>44</td>
<td>12</td>
<td>22</td>
<td>75</td>
<td>1</td>
<td>6</td>
<td>78</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>68% (64–71)</td>
<td>[136,137]</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>4</td>
<td>28</td>
<td>89</td>
<td>3</td>
<td>11</td>
<td>6</td>
<td>49</td>
<td>58</td>
<td>73% (70–76)</td>
<td>[138,139]</td>
</tr>
<tr>
<td>70</td>
<td>25</td>
<td>33</td>
<td>11</td>
<td>93</td>
<td>76</td>
<td>69</td>
<td>56</td>
<td>73</td>
<td>89</td>
<td>26% (23–30)</td>
<td>[63]</td>
</tr>
<tr>
<td>55</td>
<td>100</td>
<td>11</td>
<td>22</td>
<td>44</td>
<td>71</td>
<td>103</td>
<td>78</td>
<td>91</td>
<td>STn1033</td>
<td>25% (22–29)</td>
<td>[50,140,141]</td>
</tr>
</tbody>
</table>

*Others* includes (depending on the studies): ears, vagina, cervix, vulva, catheter, eye and urine.

Underlined *emm*-types are included in the 26-valent vaccine candidate.
disease in Europe, the USA and Canada. The release of this vaccine in the USA resulted in a reduction in the seven vaccine serotypes and associated invasive disease [100]. However, a large increase in the number of nonvaccine serotypes has since occurred, resulting in a 45% increase in nonvaccine serotype-related invasive diseases. This increase was reported to be due to the expansion of already established pneumococcal clones rather than the emergence of new clones [102]. Thus, the lack of association of some bacterial serotypes with disease in a population may not be a function of the pathogenicity of a particular strain; it is probably a function of the competition between strains for an appropriate niche in the host.

Clearly, serotype replacement is a potential issue for any multivalent vaccine. If GAS strain replacement does occur postvaccination, effective coverage would require constant strain surveillance and the re-evaluation and construction of the multivalent vaccine so that the appropriate strains are incorporated. However, the cost of such constant vaccine scrutiny may outweigh the economic benefit of vaccine release in developed countries, while rendering vaccination completely unaffordable in developing countries.

Nevertheless, the pneumococcal vaccine has had a positive impact on disease reduction [102]. Although many consultations are still warranted with regards to the cost—benefit of the 26-valent vaccine, it is evident from the level of coverage provided by the 26-valent vaccine in group 1 countries that it would have a significant positive effect on the reduction of GAS disease burden and associated economic costs.

C-terminal vaccine design

In an effort to develop a vaccine that has the potential to provide protection against a majority of, if not all, circulating GAS strains, the highly conserved CRR of the M protein has been targeted by several groups [103–105]. Unlike the N-terminal and sub-N-terminal domains of the M protein, which presents high levels of variation between different and also the same emm-types, the CRR is present in all M proteins and its sequence is highly conserved. Although not responsible for naturally acquired serotype specific immunity against GAS, antibodies specific for CRR have been found in the blood and also mucosal secretions of people from communities endemic for GAS disease [107,108]. These titer increase with age, which has been attributed to repeated GAS infections [109]. Owing to the numerous roles the CRR is believed to have in the pathogenesis, such as resistance to phagocytosis and also cellular adherence, it is thought that antibodies to this region would in turn reduce bacterial survival and protect against disease.

Early vaccine studies used peptides spanning the entire CRR of the M protein, encompassing all three repeat domains [105,110]. When delivered intranasally in a murine model, these peptides stimulated both systemic and mucosal humoral responses, which could reduce colonization and also protect against disease. Similar results were also observed when the CRR from M6 was expressed on the surface of vaccinia virus and delivered to mice intranasally [6]. Further investigation of the CRR suggested that there may be safety risks associated with using the entire domain due to the presence of cross-reactive T-cell epitopes [111]. Subsequently, Hayman et al. identified a 20-mer peptide (p145) spanning a region of the CRR that was capable of inducing opsonic antibodies, yet this epitope still reacted with T cells from RF patients. Further mapping of this peptide identified the specific B-cell epitope responsible for the protective immunity, which was termed J8-i. By omitting the potentially cross-reactive epitopes [112]. A longer homolog that encompasses J8-i was also identified termed J14-i. Initial studies with these peptides demonstrated that the protective B-cell epitope they contained was conformationally restricted; hence the peptides needed to be presented in their native α-helical conformation to induce a functional opsonic immune response. A novel strategy employed by Rel et al. was to position J8-i and J14-i within flanking regions of an α-helical peptide derived from yeast, CGN4, to produce the chimeric peptides J8 and J14, respectively [113]. Both chimeric peptide vaccine candidates have been predicted to be safe in vivo and, when coupled with diptheria toxoid (DT), are immuno-genic in mouse models, providing protection against systemic challenge with multiple GAS serotypes [114]. The mechanism of immunity induced by these peptides is eluded to be B-cell mediated, inducing opsonic antibody against GAS. In a mucosal GAS model [115], J14-DT provides protection against both colonization and systemic disease, suggesting a dual method of protection [116].

The conjugate vaccine J8-DT delivered with human-approved adjuvant alum is currently in preclinical assessment and expected to reach Phase 1 clinical trials in the near future [Batzloff M. Pers. Comm.]. Molecular epidemiological data acquired from two group 3 countries that are endemic for GAS (Fiji and India) are supportive of the potential of this CRR derived vaccine candidate to provide protection against the majority of circulating GAS strains. In data collected from Fiji, 96% of strains contain either J14 or the closely related cross-reactive variant J14.1 in the C3 repeat. The remaining strains either contained new J14 types previously not encountered or a variant J14 type usually found in the downstream C repeats. How the presence of these new J14 types in the C3 repeat affects vaccine efficacy is yet to be determined through cross-reactivity studies with J14-specific antibodies. Similar data were also recorded in Brazil and Brussels, where 94% of 51 strains tested presented either J14 or J14.1 in the C3 repeat [179]. In addition, all of the M proteins sequenced contained at least two C repeats. Thus, this vaccine offers the possibility of providing equal levels of protection against the high diversity of circulating emm-types found in group 1, 2 and 3 countries.

While in theory J14 has great potential in the protection against GAS and associated diseases, there are two questions that are still to be addressed with regards to vaccine efficacy. First, what is the risk associated with relying on a single protective epitope in a protein that already displays some level of sequence divergence? The binding of the CRR to many host proteins such as factor H, IgG and albumin is thought to be dependent on the CRR, assuming its native α-helical conformation [177]. It has been hypothesized that due to these functionality constraints, the CRR may be resistant to amino acid changes normally acquired due to immune selective pressure. However, there are more than 59 J14 variants in the C-repeat regions; increased immune selection pressure may provide the impetus for faster evolution of the repeat
sequences. The recent identification of new J14 types present in the C3 repeats [69] suggests that this could be a possibility that may need to be considered and addressed. Hence, it is quite possible that the release of the conserved epitope-based peptide vaccine targeting the CRK may also require a format of bacterial surveillance to monitor evolution of the epitope following vaccine-induced immune pressure.

Second, what would be the consequences of complete eradication of GAS globally? If successful, a highly conserved vaccine could ultimately result in the eradication of GAS from the bacterial flora of the upper respiratory tract. If this does occur, what are the downstream effects of removing a well-described member of the pharyngeal flora on the remaining inhabitants? Will this new environmental niche provide an opportunity for colonization by new commensal or other pathogenic bacterial species? Or will GAS eradication result in the redistribution of existing bacteria in the upper respiratory tract? As highly conserved GAS vaccines are transitioning into clinical trials, the wider ramifications of the eradication of streptococci need to be contemplated.

Expert commentary

While the impact of morbidity and mortality due to GAS extends throughout the world, the severity is particularly felt in the populations of developing countries where the rates of RF and RHD are high. The stark contrast in the epidemiology of streptococcal research between the three cross-continental groups is probably due to complex interplay between the circulating M types, population density, infection rates, seasonal differences and horizontal cross-species genetic exchanges.

The contrasting features of epidemiology offer challenges for vaccine design. The immune response of the M molecule is known to be protective and some epitopes elicit cross-reactive antibodies responsible for autoimmune diseases. This has necessitated separation of these epitopes in the design of a safe vaccine candidate. In this article, two distinct approaches of M-based vaccine designs are reviewed. While the multivalent N-terminal epitope vaccine has the advantages of a ‘designer’ vaccine by avoiding unnecessary immune responses, vaccine-induced GAS strain replacement may be a major drawback. On the other hand, there is already population-based evidence for the minimal effect of immune pressures due to herd immunity resulting from the vaccine based on the conserved epitopes. For instance, in GAS-endemic populations, the age-related acquisition of immunity to GAS infection is concordant with increased antibody titers against the conserved region. However, the quality of antibody response to the conserved region is different in the natural population compared with vaccinated individuals; the titers in natural population are low and acquired slowly in adults, while the vaccine is expected to elicit higher titers of antibody responses rapidly in the younger population. How does this affect pressures of herd immunity on strain replacement?

Five-year view

There is more optimism now than ever for a safe and effective vaccine against GAS. It is likely that the conserved epitope-based vaccine would cover both GAS and closely related commensal bacteria, such as GGS. However, the N-terminal vaccine has the distinct advantage of being a designer vaccine and in eliciting very high immune responses. If indeed both these candidates graduate to higher levels of clinical trials, it will be an obvious step to link both of these in one vaccine. In fact, a previous study attempted to achieve this with a limited number of N terminal epitopes and showed protection in mouse model [118].

Acknowledgements

The authors sincerely acknowledge PV Brahmacari, Tidayoshi Ikebe, Sunjo Kim, Vera Mjas, Stanford Shulman, Andrew Steer, Kimiko Utsukawa and Harno Watanabe for sharing details about their M-protein epidemiology studies. The authors also thank Michele Druau and for her excellent statistical support and Mimi Keesing for her excellent design work.

Key issues

- Recent estimates place group A streptococcus (GAS) diseases in the top ten infectious disease burden; most cases occurring in settings of poverty.
- Sparse baseline epidemiological studies in countries where streptococcal disease is considered endemic may hinder advances in vaccine research and eventual uptake.
- Contrasting epidemiological features in tropical regions challenges the accepted association between pharyngitis resulting from ‘rheumatogenic GAS’ and rheumatic fever/rheumatic heart disease. In fact, the link between a defined emm-type and one disease pathology is not universal.
- Antigenic diversity of the major surface virulence protein, M protein, and its geographical distribution suggests that three groups of diversity exist.
- The selective pressures driving these differences in terms of the diversity of circulating emm-types remain to be found. Herd and protective immunities are primarily type-specific.
- The current 26-valent vaccine may be currently efficacious against GAS diseases in group 1 (i.e., Canada, the USA, Mexico, Western Europe, Korea and Japan) countries, but is not likely to be so in group 2 (i.e., countries from Eastern Europe such as Serbia and Poland, and China) and group 3 countries (i.e., Brazil, Ethiopia, Israel, India, Nepal, Australia and Fiji).
- The vaccine based on the conserved region could be effective against virtually all M types.
- We do not know the effect vaccine-generated immune pressure will have on the selection of novel variants. To determine this, we need baseline information on circulating GAS emm-types before and after the introduction of the vaccine.
- Will decreased GAS colonization rates significantly change nasopharyngeal flora?
Financial & competing interests disclosure
Pierre R Smeesters is a 'chargé de recherche' FNRS and has been supported by an ESPID Fellowship Award. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.
No writing assistance was utilized in the production of this manuscript.

References
Papers of special note have been highlighted as:
• of interest
•• of considerable interest
•• Excellent extensive review.
•• Comprehensive review of group A streptococcus (GAS) virulence factors.
Differences among group A streptococcus epidemiological landscapes


42 Fae KG, Diefenbach da Silva D, Bilate AM et al. PDA1, HSP40 and vimentin, proteins identified by 2-DE in the valvular tissue, are the target antigens of peripheral and heart infiltrating T cells from chronic rheumatic heart disease patients. *J. Autoimmun.* 31(2), 136–141 (2008).


- Interesting comparison between throat and skin isolates from different locations.


- Original and argued questioning on the exclusive role of throat infections in the pathogenesis of rheumatic fever.


- Large epidemiological surveillance.


- Interesting study showing that the presence of most virulence factors does not increase per se the observed clinical virulence.


- Complete overview of the European epidemiology.


** Large epidemiological study in a country with a high number of circulating emm-types.


- Recent and complete overview of North American pharyngitis epidemiology.


Watson RF, Rothbard S, Swift HF. Type-specific protection and immunity following intranasal inoculation of monkeys with group-A hemolytic streptococci. J. Exp. Med. 84(2), 127–142 (1946).


- Results of the 26-valent Phase I clinical trial.


- Interesting paper illustrating the high number of circulating emm-types in India.


95 Davies MR, Tran TN, McMillan DJ, Gardiner DL, Currie BJ, Srirakshak KS. Inter-species genetic movement may blur the epidemiology of streptococcal diseases in endemic regions. Microbes Infect. 7(9–10), 1128–1138 (2005).


111 Pioneering studies elucidating to the potential of peptides encompassing conserved M-protein epitopes as mucosally delivered vaccine antigens.


115 Describes identification of the minimal B-cell epitope within the conserved repeat region of the M protein and the strategies employed to maintain the conformation of this epitope to promote immunogenicity.


Complete and interesting review of the epidemiological situation in Israel.


McDonald MF, Towers RJ, Fagan P, Carapetis JR, Currie BJ. Molecular typing of Streptococcus pyogenes from remote Aboriginal communities where rheumatic fever is common and pyoderma is the predominant streptococcal infection. Epidemiol. Infect. 135(8), 1398–1405 (2007).


Affiliations

- Pierre R Smeeers, MD, PhD
  Bacterial Pathogenesis Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia

- Laboratoire de Génetique et Physiologie Bactérienne, Institut de Biologie et de Médicine Moléculaire, Faculté des Sciences, Université Libre de Bruxelles, Belgium

- Infectious Diseases Department, Hôpital Universitaire des Enfants Reine Fabiola, Université libre de Bruxelles, Belgium
  Tel.: +32 2650 9776
  pimeercie@ulb.ac.be

- David J McMillan, PhD
  Bacterial Pathogenesis Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia

- Griffith Medical Research College, Herston, Queensland, Australia
  Tel.: +61 073 845 3698
  david.mcmillan@qimr.edu.au

- Kadaba S Srirakash, PhD
  Bacterial Pathogenesis Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia

- Griffith Medical Research College, Herston, Queensland, Australia
  Tel.: +61 073 845 3698
  sril.srirakash@qimr.edu.au

- Melina M Georgousakis, PhD
  Bacterial Pathogenesis Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia
  melina.gegousakis@qimr.edu.au