Fibronectin-Binding Protein Gene Recombination and Horizontal Transfer between Group A and G Streptococci

Rebecca J. Towers,1,2,3 Daniel Gal,1,2 David McMillan,4 Kadaba S. Sriprakash,4 Bart J. Currie,1,2 Mark J. Walker,3 Gursharan S. Chhatwal,3 and Peter K. Fagan1,2,6*

Menzies School of Health Research,1 Cooperative Research Centre for Aboriginal Health,5 Institute of Advanced Studies, Charles Darwin University,2 Darwin, Northern Territory, Department of Biological Sciences, University of Wollongong, Wollongong, New South Wales,3 and Queensland Institute of Medical Research, Brisbane, Queensland,4 Australia, and German National Research Centre for Biotechnology, Braunschweig, Germany5

Received 20 April 2004/Returned for modification 8 June 2004/Accepted 30 July 2004

We report evidence of interspecies gene transfer between the important virulence factor genes sfbI and gfbA. Because the identified group G streptococcus gfbA types possess DNA cassettes that can be identified in a number of group A streptococcus strains, it appears that homologous recombination is occurring between these species.

Group A streptococcus (GAS) is a human-specific pathogen primarily infecting the throat and skin. Group G streptococcus (GGS) is generally regarded as a less important organism, which cohabits with GAS at the same tissue sites. However, some GGS, specifically Streptococcus dysgalactiae subsp. equisimilis, are capable of causing serious and life-threatening diseases traditionally associated with GAS (1, 4, 9, 20, 28). These diseases include septicemia (10), endocarditis (25), septic arthritis (12), pneumonia, meningitis, pharyngitis, otitis media, acute poststrep-tococcal glomerulonephritis (22), and cellulitis (7). Although no definitive cases of GAS-associated acute rheumatic fever have been reported, antibodies raised against Australian GGS strains have recently been shown to cross-react with human heart muscle myosin, which has reported to indicate a potential to elicit an autoimmune response that may trigger acute rheumatic fever (13).

Identification of genes encoding C5a peptidase, M protein (emm) (5, 24, 26), streptococcal pyrogenic exotoxin G (SpeG) (23), streptolysin O, and streptokinase in GGS is consistent with the observation that these two bacteria have a common disease spectrum. Sequence analysis of these genes from various S. dysgalactiae subsp. equisimilis isolates suggested that horizontal genetic transfer between GAS and GGS may have contributed to the diversity of these genes. More recently, multilocus sequence typing revealed that horizontal acquisition may also have contributed to variations in selection-neutral housekeeping genes in GGS (16). The multilocus sequence typing data also suggests that the overall direction of genetic flow is predominantly from GAS to GGS. In order to test the generality of horizontal acquisition in contributing to diversity of virulence genes, we chose to investigate sfbI, encoding streptococcal fibronectin-binding protein I (SfbI; also known as protein F1) (14, 27).

SfbI is an important and well-characterized virulence factor in GAS that is involved in adherence to throat and skin cells (21, 27) and mediates GAS internalization in epithelial cells (15, 19). A detailed study of gene architecture of sfbI from different GAS strains revealed mosaic structures suggesting lateral genetic transfer between strains of GAS and recombination contributing to sfbI diversity (29). sfbI is situated in a chromosomal locus called the FCT region (named FCT for fibronectin, collagen binding, and T antigen) under the transcriptional regulation of RofA, which is also encoded in this region (3).

The sfbI homolog gfbA has been identified in GGS and is reportedly present in about 30% of GGS strains (17), although only one gfbA gene has been sequenced and characterized. The present study showed that approximately 10% of S. dysgalactiae subsp. equisimilis isolates from humans in the Northern Territory of Australia have the GGS sfbI homolog gfbA and that this gene is located adjacent to rofA, which is also the case in the GAS FCT locus. Comparison of derived gfbA sequences with the sequences of known sfbI genes revealed mosaicism, with sections of gfbA homologous to corresponding sections of different sfbI types.

The 38 GGS isolates used in this study were collected between 1995 and 2003 from remote Aboriginal communities in the tropical Top End of the Northern Territory, Australia, or as clinical samples from the Royal Darwin Hospital, Darwin, Northern Territory, Australia. In Aboriginal communities of the Top End, streptococcal infections are endemic (11), and the rates of poststreptococcal sequelae are among the highest in the world (6, 18). These GGS isolates are of special importance, because it has been hypothesized that pharyngeal carriage of GGS may trigger rheumatic fever in this population (13). All bacteria selected for use in this study were large colony-forming beta-hemolytic streptococci that reacted with group G-specific antisera (Oxoid, Basingstoke, England). They were identified as S. dysgalactiae subsp. equisimilis using commercially available gram-positive identification test cards in a VITEK system.

* Corresponding author. Mailing address: Division of Infectious Disease, Menzies School of Health Research, P.O. Box 41096, Casuarina, Northern Territory 0811, Australia. Phone: 61-8-89227793. Fax: 61-8-89275187. E-mail: peterf@menzies.edu.au.
The entire gfbA gene was present in 4 of the 38 GGS isolates designated 59G, NS3402, NS3404, and NS3404 (Fig. 1). Size variation in the length of gfbA (Fig. 1A) was attributed to either differences in the numbers of fibronectin-binding repeats or recombination within the proline-rich domain. All four GGS gfbA genes identified in this study were adjacent to the gene encoding the global regulator RofA (Fig. 1B). In fact, the presence of rofA was detected in all of the 38 GGS strains tested (results not shown). RofA is a global regulator of GAS (8) found in the FCT region, a locus which exhibits high rates of recombination (3). Therefore, it is possible that an analogous FCT region under the control of RofA is present in all of recombinatorial hot spots. Further investigation of the downstream genes that are influenced by RofA in GAS have been detected in human GGS isolates (2, 24, 30).

Partial sequence from the variable 5′ end of gfbA (Fig. 2A) and the published gfbA sequence (accession number U31115) (17) were compared with previously reported sfbI sequences from GAS. Overall, the architectural arrangements of predicted GfbA proteins are the same as that of SfbI (Fig. 2). From the N to C terminus, the proteins include a signal sequence, an aromatic amino acid-rich domain (Aro domain), a proline-rich domain comprising a proline-rich region (Pro region) and several proline-rich repeats, an upstream fibronectin-binding domain, a fibronectin-binding repeat region, and an anchor. The C-proximal half beginning with the proline-rich repeats is highly conserved and differs primarily in the number of repeats, but the N-proximal Aro and Pro domains are highly variable. In GAS, 15 Aro domain sequence types (Aro types) and 2 Pro region sequence types (ProA and ProB) were identified.

The five gfbA sequences were aligned against the previously described GAS sfbI sequences. The alignment revealed four gfbA types (ST35 to ST38) with three novel Aro types in GGS (Aro16 to Aro18). Similar to sfbI in GAS, the gfbA gene in GGS exhibits a mosaic structure within the Aro domain. To better illustrate this structure, the Aro domain has been divided into four main sections (Fig. 2A). These sections represent distinct DNA cassettes with local homology between Aro types, which lie between highly conserved “junction” sequences. For instance, in GAS 1750, the Aro domain sections 1, 2, 3, and 4 may have evolved from GAS Aro5, Aro6, Aro6, and Aro5, respectively. Likewise, sections 1 to 4 in GGS 59G appear to have been derived from GAS Aro1, Aro2 or Aro4, Aro2, and Aro3, respectively (Fig. 2B).

Three of the GGS isolates contain Aro18 gfbA genes. Two of the three isolates are clonal (NS3402 and NS3404) as determined by random amplification of polymorphic DNA and pulsed-field gel electrophoretic analysis. NS3572 is not clonal, although the pulsed-field gel electrophoresis fingerprint indicates a limited degree of shared loci (results not shown). This Aro type (Aro18) has evolved from mosaic

### Table 1. Primer sequences used to detect gfbA and rofA in GGS

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td></td>
</tr>
<tr>
<td>sfbI-F1</td>
<td>5′-GTCTTTCTTGACAATAACGTGTAAGCT-3′</td>
</tr>
<tr>
<td>rofA-F</td>
<td>5′-GCCAATAACTGAGGTTAGC-3′</td>
</tr>
<tr>
<td>Reverse</td>
<td></td>
</tr>
<tr>
<td>sfbI-R3</td>
<td>5′-GTATCTCAAGAACGTTACTGCAC-3′</td>
</tr>
<tr>
<td>sfbI-R6</td>
<td>5′-GCCACATAGGGTACTGAGCACAAG-3′</td>
</tr>
<tr>
<td>rofA-R</td>
<td>5′-GGTITCTGCTTATTAGG1-3′</td>
</tr>
</tbody>
</table>

**a** The primer locations within the genes and PCR strategy were described previously by Towers et al. (29).
FIG. 2. Variation seen in the sfbI/gfbA gene family in GAS and GGS. (A) Schematic showing the range of percentage DNA identity for each domain (above) and for each section of the aromatic amino acid-rich domain (below). Gaps introduced to aid alignment have been treated as single base changes. The diagram is not drawn to scale. The lower schematic defines DNA cassettes or sections used to illustrate the gene mosaic structure of sfbI/gfbA. UFBD, upstream fibronectin-binding domain; Fn-binding repeats, fibronectin-binding repeats. (B) Analysis demonstrating similarity between DNA cassettes found in sfbI and gfbA. Sections 1 to 4 correspond to the region encoding the aromatic amino acid-rich domain (described in the text). Section 5 corresponds to the proline-rich region type (ProA/ProB). Sections with the greatest percent identity to the corresponding sfbI Aro types in pairwise comparison are shown in bold type. GAS strain BL16 is an example used to demonstrate the analogous mosaic arrangement in GAS which has been previously described (29). Aro16 ProB1 was determined by analyzing the previously characterized gfbA from GGS strain 1750 (16).
arrangements. Aro domain sections 1, 3, and 4 have evolved from GAS Aro4, a novel Aro type, Aro6, and Aro8, respectively. Comparing NS3572 to NS3404, there are five base changes in the Aro-rich domain, only two of which are silent. Further differences are apparent in the proline-rich repeats with NS3404 having three nonidentical repeats. In contrast, NS3572 contains a single proline-rich repeat, which is identical to those seen in some GAS strains. These differences can be explained by accumulations of point mutations and duplication or deletion events within the genome. These differences can also be explained, however, by ongoing interspecies lateral transfer rearranging mosaic cassettes. It is not possible from these data to determine whether recombinational diversification is occurring in GAS prior to horizontal transfer or whether it is ongoing within GGS. A majority of the mosaic cassettes have been identified in GAS isolates which cohabit with *S. dysgalactiae* subsp. *equisimilis*. When the GGS aromatic-rich domain gene sequences are aligned against the known GAS aromatic-rich domain gene sequences using pairwise comparisons, the different GGS-derived Aro type groups among GAS isolates. In fact, there is very little similarity between the GGS sequences (54.3 to 67.3% in pairwise analysis). This observation, in light of the fact that the GGS sequences show less similarity between themselves than with the homologous GAS sequences, supports the hypothesis that it is unlikely that the differences seen are the result of intragenomic or intergenomic rearrangement with other GAS.

This study is the first demonstration that ongoing lateral transfer leading to genetic diversity may extend to a streptococcal virulence gene with limited strain distribution and that there is ongoing mosaic rearrangement within the *sfbI/gfbA* gene family in GAS or GGS. This horizontal gene transfer and gene rearrangement are of particular significance in light of the fact that the regulatory genes involved in the control of these pathogenicity factors (i.e., *rafA*) have also been identified in *S. dysgalactiae* subsp. *equisimilis*. We believe that taken together, these results indicate that selective pressure may result in the emergence of virulent human pathogenic GGS strains with the potential to cause diseases classically associated with GAS.

**Nucleotide sequence accession numbers.** All newly determined nucleotide sequences were deposited in GenBank. The nucleotide sequences and accession numbers follow: GAS isolate 59G (ST36), AJ605760; GAS isolate NS3402 (ST37), AJ605743; GAS isolate NS3404 (ST37), AJ605744; and GAS isolate NS3572 (ST38), AJ605745. The alignment of the GGS *gfbA* variable regions with published GAS *sfbI* Aro types may be accessed through the EMBL nucleotide database (Align 000635).

This work was supported in part by the Cooperative Research Centre for Aboriginal Health (Australia), the National Health and Medical Research Council of Australia, and the BMBF Network “Pathogenomik” (Germany). P.K.F. is the recipient of an Alexander von Humboldt Foundation Fellowship.

We thank M. McDonald and P. Dasari for their contributions to the MSHR Northern Territory Streptococcal Collection, the source of the strains mentioned in this paper.

**REFERENCES**


