Supporting Materials and Methods

Bacterial Strain Used for Sequencing. Strain MGAS315 has been characterized extensively (1-6). The organism was recovered in the late 1980s from a patient in the United States with streptococcal toxic shock syndrome (1). This strain is multilocus enzyme electrophoretic type 2 (ET 2), the genotype of virtually all serotype M3 strains recovered from patients with invasive episodes, pharyngitis, and other group A *Streptococcus* (GAS) infections (1). Strain MGAS315 has the *speA3* allele of the *speA* gene, the allele associated with contemporary serotype M3 strains, and the *ssa* gene encoding streptococcal superantigen A (3, 7). It has the *emm3.0* allele found in the majority of M3 strains (5). The culture supernatant proteome of this strain has been analyzed by two-dimensional electrophoresis and matrix-assisted laser desorption ionization time-of-flight (2). The strain can be manipulated genetically, has been studied in mouse models of upper respiratory tract and invasive infection, and is virulent for mice (4, 6). Prior to sequence analysis, the size of the genome of strain MGAS315 was estimated to be ≈1.8 Mb by pulsed-field gel electrophoresis, a size similar to the genomes of the serotype M1 and M18 strains sequenced (8, 9), and indistinguishable from 10 other randomly chosen *speA*-positive M3 strains from diverse localities (data not shown).

Genome Sequencing, Closure, and Annotation. The genome of strain MGAS315 was sequenced by methods described recently (9). Briefly, a small insert (≈2 kb) library was generated from nebulized chromosomal DNA cloned into pCR4Blunt-TOPO vector (Invitrogen). Library clones were amplified by colony PCR and end sequenced. Fluorescent dye terminator chemistry, ABI 3700 automated sequencers (Applied Biosystems), and the FINCH data management suite (Geospiza, Seattle) were used for sequence data acquisition. The DNA sequence reads were assembled into contigs, and aligned with the genome

sequence of GAS strain SF370 (GenBank acc. no. AE004092), and edited with PHRED/PHRAP, CROSSMATCH, and CONSED (P. Green, http://genome.washington.edu). The gaps in the assembly were closed by directed sequencing or transposon mediated sequencing (EZ::TN <TET-1> system; Epicentre, Madison, WI), assisted with the AUTOFINISH function of CONSED. Large repeated elements in the genome (>700 bp) such as the 16s and 23s rRNA operons were amplified from purified chromosomal DNA by PCR with primers to unique flanking sequence, sequenced, and assembled independently with SEQUENCHER (Gene Codes, Ann Arbor, MI). Regions of low sequence quality in the assembled genome were identified by CONSED and directed sequencing was performed to increase the minimum consensus base quality to Q40.

Coding sequences were predicted with GLIMMER2 (>60 bp) and start codon positions were adjusted with RBS FINDER (www.tigr.org). Translated CDSs were compared to the Entrez nonredundant protein database with BLASTP to identify homologues and predict functions. Proteins predicted to be secreted were identified with SIGNALP v2.0 (10), predicted cell-wall-attached proteins were identified using FUZZPRO (11) with a tripartite pattern including an LPXTGX motif (12), and predicted lipoproteins were identified on the basis of the presence of an LA(G,A)C motif (13). Multiple sequence alignments were performed with CLUSTALW (inn-prot.weizmann.ac.il/software) and phylogenetic trees were generated with PHYLIP (evolution.genetics.washington.edu/phylip.html).

CDSs present in the genome of strain MGAS315 but absent in the genome of M1 strain SF370 were identified by TBLASTN comparison of the strain MGAS315 translated CDSs with the M1 genome sequence (expectation value = 0.0001). Similar pairwise comparisons were made between all genomes. A protein was defined as unique if it lacked a homologue in the comparison strain with >50% positive amino acids (identical plus similar) across its complete sequence. Paired comparisons were performed with each of the GAS genomes to identify the subset of genes which were unique and shared relative to the

other GAS genomes.

Cloning of *sla*, *ssa*, and *speK* Genes. The genes encoding mature Sla (amino acid residues 28-191), SpeK (27-259), and SSA (45-191) were cloned from strain MGAS315 with paired primers 5'-ACCATGGAAGGGATAAATGATAAAATGG-3'/5'-CGAATTCTTAACATCCTATAGAACCTAC-3' and 5'-ACCATGGATACGTACAATACAAATGATG-3'/5'-CGAATTCTAATCTTTAGAAAAATCTTCG-3', respectively. The PCR products were digested with *Nco*I and *EcoR*I and cloned into pET-His2 to yield recombinant plasmids pSla and pSpeK. The vector pET-His2 was obtained by inserting the smaller *Xba*I/*EcoR*I-fragment of pET-His (GenBank acc. no. L20317) into pET-21b (Novagen) linearized by *Xba*I and *EcoR*I. Recombinant Sla (rSla) and SpeK (rSpeK) made by the clones had 11 amino acid residues MHHHHHHLETM fused to the amino-terminus of the mature proteins. The cloned genes were sequenced to rule out spurious mutations.

Purification of rSla, rSpeK, and rSSA. rSla and rSpeK were purified from *Escherichia coli* BL21 (DE3) containing pSla and pSpeK, respectively. The bacteria were grown overnight at 37°C in 2 liters of LB broth supplemented with 100 mg/liter of ampicillin; harvested by centrifugation; suspended in 30 ml of 10 mM Tris•HCl (THCl), pH 8.3; and sonicated for 15 min on ice. The cell debris was removed by centrifugation at 20,000 x g for 15 min, and the supernatant was loaded onto a Ni-NTA agarose (Qiagen) column (1.5 x 3 cm). The column was washed with 50 ml of 1.0 M NaCl in THCl, the protein was eluted with a 40 ml linear gradient of 0-0.25 M imidazole in THCl. Recombinant proteins were identified by SDS-PAGE and peak fractions were pooled. The protein was dialyzed against 3 liters of THCl at 4°C overnight and loaded on a DEAE sepharose column (1.5 x 5 cm) equilibrated with THCl. The proteins were eluted with a 60 ml linear gradient of 0-0.15 M NaCl (rSla) or 80 ml of 50 mM NaCl (rSpeK) in THCl. The purified proteins were

dialyzed against 3.5 liters THCl. Purified rSpeK and rSla were free of contaminating proteins as assessed by Coomassie blue stained SDS-PAGE. rSSA was purified by procedures described previously for rSpeC (14). All reagents and glassware used for toxin purification and biological assays were pyrogen free.

Assay for Superantigen Activity. The mitogenicity of rSpeK was determined with [^{3}H]thymidine incorporation assays (14). Rabbit splenocytes and PBMCs were seeded into 96-well plates at 2 x 10^5 cells per well. Serial 10-fold dilutions of rSpeK were added to the wells in quadruplicate. Purified rSpeC (14) and PBS were added as controls. The splenocytes were grown at 37ºC for 3 days and pulsed with 1 μCi [^{3}H]thymidine (Amersham) per well overnight. Cells were harvested onto fiberglass filters, and [^{3}H]thymidine incorporation was measured with a scintillation counter (Beckman Instruments).

Analysis of T cell Repertoire. PBMCs obtained from 5 healthy human donors were isolated, cultured with anti-CD3 (20 ng/ml) or rSpeK (100 ng/ml), and analyzed for T cell receptor Vβ receptor repertoire by flow cytometry (15). Cells were stained in 96-well, round-bottomed plates with a panel of biotinylated monoclonal antibodies directed against human Vβ1, 2, 3, 5.1, 5.2, 5.3, 7, 7.2, 8, 11, 12, 13.1, 13.2, 14, 16, 17, 20, 21.3, and 22 (Immunotech, Westbrook, ME), Vβ9, Vβ23 (PharMingen) and Vβ6.7 fluorescein isothiocyanate (FITC) (Endogen, Woburn, MA). The cells were washed, suspended in staining solution, and incubated with anti-CD3 allophycocyanin (APC), anti-CD4-peridinin chlorophyll protein (PerCP) (Becton Dickinson), anti-CD8 (FITC) (Becton Dickinson), and streptavidin phycoerythrin (PE) (Southern Biotechnologies, Birmingham, AL). The cells were fixed and analyzed by four color flow cytometry (FACSCalibur) (Becton Dickinson). Negative control reagents were used to verify the staining specificity of the antibodies used.

Pyrogenicity and Lethality Models of Toxic Shock Syndrome. American Dutch belted rabbits were used to assess the pyrogenicity and toxicity of rSpeK and rSSA (15). Rabbits were injected in the marginal ear veins with rSpeK, rSSA, rSpeC (5 μg/kg), or PBS, and temperatures were recorded at 4 hours. At the 4 hour time point, endotoxin from *Salmonella typhimurium* serovar Typhimurium (10 μg/kg, 1/50 lethal dose, 50% endpoint) was injected intravenously to assess the ability of rSpeK and rSSA to enhance host susceptibility to lethal endotoxin shock (endotoxin enhancement). Animals were monitored for symptoms of streptococcal toxic shock syndrome (STSS), and mortality was recorded over a 2-day period.

The miniosmotic pump model of STSS (16) was used to assess the lethality of the rSpeK and rSSA in a model presumed to resemble infection with exotoxin-producing GAS. The pumps release a constant amount of exotoxin into the subcutaneous tissue over a period of 7 days. American Dutch belted rabbits were anesthetized with ketamine and xylazine (Phoenix Pharmaceuticals, St. Joseph, MO), and miniosmotic pumps (Alza Pharmaceuticals, Palo Alto, CA) filled with 500 μg of rSpeK, rSSA, or rSpeC in 200 μl of PBS were implanted subcutaneously into the left flank. Rabbits were monitored for signs of STSS, and mortality was recorded over a 15-day period.

Phospholipase A_2 (PLA_2) Assay. PLA_2 activity was assayed with a commercially available kit (Cayman Chemical, Ann Arbor, MI) that measures the hydrolysis of phospholipids at the *sn*-2 position, yielding a free fatty acid and a lysophospholipid. The assay uses the 1,2-dithio analog of diheptanoyl phosphatidylcholine as substrate (17). Bee venom PLA_2 supplied with the kit was used as positive control (data not shown), and rSpeK was used as a negative control.

Detection and Chromosomal Mapping of *speK* and *sla*. PCR was used to screen GAS strains for the presence of genes encoding SpeK and Sla in 121 serotype M3 strains from diverse localities and times (Table 2). The collection included strains recovered from patients with pharyngitis, acute rheumatic fever, impetigo, and severe invasive infections in eight of the United States (n = 29 isolates), Ontario, Canada (n = 22 isolates), United Kingdom (n = 1), former East Germany (n = 67 isolates), former Soviet Union (n = 1), and New Zealand (n = 1). Sixty-eight isolates were recovered between 1920 and 1984, and 53 isolates were collected between 1987 and 2001. The strains include organisms isolated from patients in a large epidemic of scarlet fever that occurred in Ottawa, Canada, between September 1940 and February 1942, and two country-wide outbreaks of serotype M3 GAS that occurred in the former East Germany between 1968 and 1990 (18-20). The epidemiology of these outbreaks has been well described previously and many of the GAS isolates were characterized previously by multilocus enzyme electrophoresis (18-20).

Internal primers and chromosomal DNA were used for the PCR amplification of *speK* and *sla* (Table 3). To determine if these genes were encoded by phage Φ315.4, and if this phage was integrated at the same chromosomal locus in the test strains as in reference strain MGAS315, PCR was performed with combinations of primers for *speK* and *sla*, and contiguous phage and chromosomal genes (Table 3).

Detection of Anti-SSA, -SpeK, and -Sla Antibodies in Sera Obtained from Humans with GAS Infections. Western immunoblotting was used to test for the presence of specific antibodies to rSSA, rSpeK, and rSla in paired acute and convalescent sera from patients with GAS pharyngitis and invasive episodes. Acute-phase sera were obtained upon diagnosis, and convalescent sera were drawn 10-110 days after GAS diagnosis.

Reference List

1. Musser, J. M., Hauser, A. R., Kim, M. H., Schlievert, P. M., Nelson, K. & Selander, R. K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2668-2672.
2. Lei, B., Mackie, S., Lukomski, S. & Musser, J. M. (2000) *Infect. Immun.* **68**, 6807-6818.
3. Musser, J.M., Kapur, V., Kanjilal, S., Shah, U., Musher, D. M., Barg, N. L., Johnston, K. H., Schlievert, P. M., Henrichsen, J., Gerlach, D., Rakita, R. M., Tanna, A., Cookson, B. D. & Huang, J. C. (1993) *J. Infect. Dis.* **167**, 337-346.
4. Kapur, V., Maffei, J. T., Greer, R. S., Li, L.-L., Adams, G. J. & Musser, J. M. (1994) *Microb. Pathol.* **16**, 443-450.
5. Hoe, N., Nakashima, K., Grigsby, D., Pan, X., Dou, S. J., Naidich, S., Garcia, M., Kahn, E., Bergmire-Sweat, D. & Musser, J. M. (1999) *Emerging Infect. Dis.* **5**, 254-263.
6. Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Ireland, R. M., Reid, S. D., Adams, G. G. & Musser, J. M. (2000) *Infect. Immun.* **68**, 6542-6553.
7. Reda, K. B., Kapur, V., Mollick, J. A., Lamphear, J. G., Musser, J. M. & Rich, R. R. (1994) *Infect. Immun.* **62**, 1867-1874.
8. Ferretti, J. J., McShan, W. M., Ajdic, D., Savic, D. J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A. N., Kenton, S. et al. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 4658-4663.
9. Smoot, J. C., Barbian, K. D., Van Gompel, J. J., Smoot, L. M., Chaussee, M. S., Sylva, G. L., Sturdevant, D. E., Ricklefs, S. M., Porcella, S. F., Parkins, L. D. et al. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 4668-4673
10. Nielsen, H., Engelbrecht, S., Brunak, S. & von Heijne, G. (1997) *Protein Eng.* **10**, 1-6.
11. Rice, P., Longden, I. & Bleasby, A. (2000) *Trends Genet.* **16**, 276-277.
12. Janulczyk, R. & Rasmussen, M. (2001) *Infect. Immun.* **69**, 4019-4026.
13. Sutcliffe, I. C. & Russell, R. R. B. (1995) *J. Bacteriol.* **177**, 1123-1128.
14. McCormick, J. K., Tripp, T. J., Olmstead, S. B., Matsuka, Y. V., Gahr, P. J., Ohlendorf, D. H. & Schlievert, P. M. (2000) *J. Immunol.* **165**, 2306-2312.
15. Dinges, M. M. & Schlievert, P. M. (2001) *Infect. Immun.* **69**, 7169-7172.
16. Lee, P. K., Deringer, J. R., Kreiswirth, B. N., Novick, R. P. & Schlievert, P. M. (1991) *Infect. Immun.* **59**, 879-884.
17. Reynolds, L. J., Hughes, L. L. & Dennis, E. A. (1992) *Anal. Biochem*. **204**, 190-197.
18. Musser, J. M., Nelson, K., Selander, R. K., Gerlach, Huang, J. C., Kapur, V. & Kanjilal, S. (1993) *J. Infect. Dis.* **167**, 759-762.
19. Bynoe, E. T. (1943) *Can. J. Public Health* **34**, 272-281.
20. Kohler, W., Gerlach, D. & Knoll, H. (1987) *Zentralbl. Bakteriol. Hyg. A* **266**, 104-115.