Journal of Medical Microbiology (2006), **55**, 157–164 DOI 10.1099/jmm.0.46224-0

Characterization of group A streptococci (*Streptococcus pyogenes*): correlation of M-protein and *emm*-gene type with T-protein agglutination pattern and serum opacity factor

Dwight R. Johnson,[1] Edward L. Kaplan,[1] Amy VanGheem,[1] Richard R. Facklam[2] and Bernard Beall[2]

[1]Department of Pediatrics and the World Health Organization Collaborating Center for Reference and Research on Streptococci, University of Minnesota, Minneapolis, MN 55455, USA

[2]Centers for Disease Control and Prevention, Division of Bacterial and Mycotic Diseases, Respiratory Diseases Branch, 1600 Clifton Rd, Mailstop CO2, Atlanta, GA 30333, USA

Correspondence
Edward L. Kaplan
kapla001@umn.edu

Received 1 July 2005
Accepted 23 September 2005

Strain characterization of group A streptococci (GAS) has traditionally been based on serological identification of M protein. Additional tests to determine T-protein serotype and production of streptococcal serum opacity factor (SOF) provide important information both to aid in and to supplement M-protein serotyping. Advances in DNA-sequencing technology in the late twentieth century resulted in the development of a method for determining the M type of GAS from the sequence of the gene encoding M protein, the *emm* gene. Although *emm*-sequence typing has largely replaced M typing in many laboratories, information provided by T typing and SOF determination continues to provide valuable supplementary information for strain characterization. A comprehensive summary of the correlation of T pattern and SOF production with M type was last published in 1993, several years before *emm* typing became widely available. Since then, the ease of M-type identification afforded by *emm* typing has resulted in an increase in the number of confirmed M/*emm* types of more than 50 %. However, comprehensive information about T-protein serotype and the correlation of SOF production with these new M/*emm* types is not widely available. This report presents a comprehensive summary of this information, not only for newly described types, but also updated information for previously described types. This information was extracted from combined records from streptococcal reference laboratories at the University of Minnesota and at the Centers for Disease Control and Prevention in Atlanta. Data from more than 40 000 strains (representing uncomplicated GAS infections, systemic invasive infections and strains associated with non-suppurative sequelae, collected from the US and diverse locations worldwide) were analysed.

INTRODUCTION

For more than 80 years, extensive clinical, epidemiological and laboratory research efforts have focused on understanding group A streptococcal infections and their sequelae. Despite these many decades of research, a complete understanding of the pathogenetic mechanisms of *Streptococcus pyogenes* (group A streptococci, GAS) remains elusive. Both laboratory research and epidemiological studies will continue to be crucial for understanding the role of unique microbial properties along with human susceptibility factors in pathogenesis.

Abbreviations: CDC, Centers for Disease Control and Prevention; GAS, group A streptococci; SOF, serum opacity factor; UMN, University of Minnesota.

Successful studies depend on accurate subspecies characterization of GAS. Demonstration that M protein is an important GAS virulence factor and that antibody to this protein confers protection to the human host led to the acceptance of M typing as the preferred method for GAS characterization throughout most of the twentieth century.

The M-typing process is enhanced by inclusion of two other GAS-characterization methods: T-protein serotyping and detection of streptococcal serum opacity factor (SOF). However, even with utilization of T typing and SOF determination, scarcity of M antisera has prevented the identification of many GAS strains. This problem was more completely addressed in the mid 1990s by the development of *emm*-gene sequence typing, a system that provides a practical alternative to serotyping to identify M-protein type.

T serotyping and SOF determination add valuable parameters for strain identification. For example, it is well known that strains may share the same M/*emm* type, but may differ in clonal type. This is especially true when strains are recovered from diverse locations (Beall *et al.*, 2000). Marked differences in T serotypes between strains sharing the same *emm* sequence type or M serotype almost invariably correlate to markedly different clonal types (Beall *et al.*, 2000), and such differences could have potentially important pathogenetic significance. The most recent comprehensive review of the correlation of T pattern and SOF production with M type was published in 1993 (Johnson & Kaplan, 1993). Since then, the number of M/*emm* types has increased by more than 50 %, but important T-type/SOF correlation information has not been widely available. This report presents an updated and comprehensive correlation of T type and SOF/*sof* with M/*emm* type, based on analysis of more than 40 000 GAS isolates obtained worldwide during the past 50 years.

METHODS

Strains. Group A streptococcal isolates received at the University of Minnesota (UMN) streptococcal reference laboratory from 1953 to 2004 and at the Centers for Disease Control and Prevention (CDC) streptococcal reference laboratory between 1995 and 2004 were included in the analyses. Data from the UMN were based on the characterization of more than 35 000 GAS isolates from individuals with uncomplicated group A streptococcal infections (pharyngitis, pyoderma/impetigo), from systemic invasive infections, and also strains associated with non-suppurative sequelae of group A infections (rheumatic fever, acute post-streptococcal glomerulonephritis). Data from the CDC were based on analysis of more than 7500 isolates. Both the CDC and the UMN GAS collections summarized in this report include isolates from the USA as well as from diverse locations worldwide.

T serotyping. T-agglutination serotyping was performed at the UMN following published protocols (Moody *et al.*, 1965). Prior to 1990, the T-typing antisera used were obtained either from the Central Public Health Laboratories in London or from the CDC in Atlanta. Since 1990, T-typing antisera have been obtained from Sevapharma in Prague, and T-typing antisera used at the CDC were prepared there.

M serotyping. Streptococcal extracts for serogroup determination and for M typing were prepared at the UMN by slight modifications of the classical method of Lancefield (Johnson *et al.*, 1996; Moody *et al.*, 1965). Prior to 1972, M-type precipitin reactions were done by the glass capillary tube method (Swift *et al.*, 1943); subsequent to that date, all M-typing reactions were performed by the Ouchterlony double-diffusion technique (Rotta *et al.*, 1971). M-type-specific antisera used were either prepared in the UMN laboratory using standard methods (Johnson *et al.*, 1996) or obtained from the CDC. Data presented from the CDC in this report were based entirely on *emm* typing; M typing was not performed during the period of time in which the 7500 isolates were tested.

Streptococcal SOF detection. Prior to 1988, detection of SOF was performed at the UMN by the agar plate method (Maxted *et al.*, 1973). Since 1988, SOF detection has been performed in 96-well microplates, and the results are read spectrophotometrically (Johnson & Kaplan, 1988). Culture supernatants, acid extracts and SDS extracts were used as the source of SOF (Rehder *et al.*, 1995). SOF detection was done at the CDC by the agar plate method using culture supernatants (Maxted *et al.*, 1973).

***sof*-gene detection and sequencing.** Routine *sof* gene-detection and sequencing performed at the CDC was based upon PCR and sequence analysis of a variable-length 450–650-base PCR fragment using methods and primers described previously (Beall *et al.*, 2000). Detection of the *sof* gene and *sof* sequencing were not done at the UMN during the time period of this report.

***emm*-sequence typing.** *emm*-sequence typing has been in use at the CDC since 1995. More than 7500 *emm*-typing results obtained at the CDC are included in this report. Typing by *emm* sequencing was initiated at the UMN in 1997 and used in conjunction with M serotyping for characterization of GAS. Approximately 1400 of the 35 000 UMN results are based on *emm* typing. Preparation of GAS lysates, *emm*-gene PCR and *emm* sequencing were performed as previously described (Beall *et al.*, 1996), with modifications at http://www.cdc.gov/ncidod/biotech/strep/protocols.htm.

Collection of data. The T-pattern associations presented in this report are based on analysis of more than 42 000 strains (approx. 7500 CDC results and 35 000 UMN results). However, T-pattern frequency estimates were determined from the analysis of approximately 21 000 results that were available in easily quantifiable computerized databases; this subset included nearly 14 000 UMN results obtained since 1986 and more than 7500 isolates characterized at the CDC since 1995.

RESULTS AND DISCUSSION

Relevance of T typing and SOF determination

Accurate classification of group A streptococcal strains is essential for epidemiological studies, for basic laboratory research and, potentially, for future vaccine development. For most of the twentieth century, serotyping by M-protein precipitation was the laboratory classification technique most widely used. Two additional classification methods, T-protein serotyping and detection of SOF, are useful adjuncts to M typing. However, even with the support of T typing and SOF determination, many M-typing attempts are unsuccessful. Only a few reference laboratories have had extensive sets of M-typing antisera available, because the M-protein-specific antisera required are difficult and expensive to prepare. The introduction of *emm*-gene sequence typing has permitted M-type determination of nearly all strains of GAS by any laboratory with DNA-sequencing capability. However, even in an age of molecular GAS typing, information about T pattern and SOF production remains relevant for several reasons.

T-pattern and SOF/*sof* information can be a valuable asset to efficient and cost-effective determination of *emm*-sequence type. In situations requiring large-scale *emm*-type identification of epidemiologically related GAS isolates, the sequencing materials and labour costs can be high, and the time required to generate results can be quite long. To improve sequencing efficiency and economy, restriction digests of *emm* amplicons from all strains to be characterized can be subjected to gel electrophoresis and the resulting enzyme restriction patterns compared. Strains with common

T patterns, SOF production and *emm*-restriction patterns can then be grouped together, and *emm* sequencing can be performed on a small sample. The *emm* type of the group can then be deduced based on the result of the sequenced subset. Among epidemiologically associated isolates, common T pattern, SOF production and *emm*-gene restriction-enzyme patterns have been shown to quite accurately predict a mutually shared *emm* type (Beall *et al.*, 1998; Espinosa *et al.*, 2003; see also http://www.cdc.gov/ncidod/biotech/strep/protocols.htm).

T-pattern and SOF-production information is also useful for validating the accuracy of *emm*-typing results. The correlations shown in Table 1, based on the analysis of large numbers of isolates from varied geographical areas and collected over a 50-year period of time, likely represent a majority of T-pattern/SOF/M-type/*emm*-type combinations expected to be encountered during routine studies. Results that deviate significantly from those in Table 1, especially SOF/*sof*–*emm* discordance, should be confirmed.

T-pattern variation can also be an indicator of strain variation within an M/*emm* type. As more strains are examined, especially from less-studied regions of the world, additional divergent T-pattern/M-/*emm*-type results will undoubtedly be found. If T-pattern results that differ from those presented in Table 1 are confirmed, it could indicate that the isolate(s) in question are clonally distinct from previously described members of that M/*emm* type, a finding that could have important epidemiological and pathogenetic significance. Evidence of this has emerged from *sof*-gene sequencing studies that have found a discordance between *emm*-type and SOF-inhibition typing results. Strains with concordant *emm* and SOF types typically had 'classical' T patterns: T patterns long recognized as associated with that M type. When strains with discordant *emm*-/SOF-type associations were discovered, some gave T patterns that were not common for that *emm* type. Further, when it was observed that certain pairs of previously designated M types shared complete or nearly complete homology in their 5′ *emm*-gene sequences (i.e. *emm* 27L/77, 44/61, 50/62 and 65/69; see Table 1), each member of the pair had a unique T pattern that distinguished it from the other (Facklam & Beall, 1997). It is clear that T-pattern deviation can be an indicator of strain diversity within an *emm* type.

SOF production (and *sof*-gene detection) also provide useful additional strain information. If one performs phylogenetic analysis of aligned segments of all known deduced *emm*-gene products, these segments segregate almost entirely into *sof*-gene-positive and *sof*-gene-negative clusters (Beall *et al.*, 2000; Facklam *et al.*, 2002; McGregor *et al.*, 2004; Teixeira *et al.*, 2001; Whatmore *et al.*, 1994). Also, *sof*-positive strains have a distinct pattern of genes in the neighbourhood of *emm* compared to that of *sof*-negative strains (McGregor *et al.*, 2004). Several studies have also suggested that SOF production has a high correlation with M-protein class; strains which do not produce SOF tend to possess class I M proteins, and SOF-producing strains carry class II M proteins (Bessen *et al.*, 1989; Bessen & Fischetti, 1990). It is becoming increasingly clear that the presence or absence of SOF/*sof* indicates fundamental differences in M-protein structures, and also indicates fundamental differences in *mga*-locus gene composition and arrangement. Both of these differences are likely to have profound, although as-yet undefined, biologic significance.

Finally, T-pattern and SOF information about a GAS strain, especially when combined with knowledge of M or *emm* type, provides an important link to information from studies published over the many decades of basic clinical and epidemiological GAS research, when these serological methods were the primary tools available for strain classification.

Factors affecting T-type interpretation

Several factors can cause difficulties in the interpretation of T-serotyping results. Many isolates carry combinations of T proteins, resulting in T-agglutination 'patterns'. Several of these patterns are reproducible and widely recognized (e.g. 3/13/B3264, 5/27/44, 8/25/Impetigo 19). However, many GAS isolates react non-specifically with T antisera, leading to agglutination with many, or even all, sera. Unless these cross-reactions are removed by trypsinization, incorrect T-agglutination patterns will be obtained. Unfortunately, excessive trypsinization may remove true T-protein reactions and affect the final T pattern. For example, a strain with the very common 3/13/B3264 agglutination pattern may lose some, or even all, of the reactions making up that pattern depending on the intensity of trypsinization. Therefore, when comparing individual results with those in Table 1 or when interpreting T-typing results obtained by different persons, in different laboratories, or at different times, it is important to be aware that different combinations of common patterns may, in fact, represent the same type.

Factors affecting SOF determination

Determination of SOF production by group A streptococcal strains has been a widely used and valuable tool for strain characterization. SOF production occurs in approximately half of all known M/*emm* types, and this production correlates highly with specific M type (Top & Wannamaker, 1968) and *emm* type (Beall *et al.*, 1996, 2000; Whatmore *et al.*, 1994).

The sensitivity of tests used to determine SOF production varies considerably. It is most common to test for SOF in Todd–Hewitt broth-culture supernatants, but SOF can also be detected in Lancefield HCl extracts and in SDS extracts of streptococcal cells. The highest sensitivity is achieved by using several methods collectively, but, of the single tests, use of SDS extracts appears to give the lowest percentage of false-negative results (Rehder *et al.*, 1995). The probability of a false-negative result is increased if only a single source of SOF is used.

Table 1. M-/*emm*-type, SOF/*sof* and T-pattern correlations for GAS isolates

The 'M type' column contains entries for only those serotypes that have received official confirmation and M-type designation. Currently, this includes M types 1–93. The '*emm* type' column contains designations for all confirmed and officially designated *emm*-type sequences (Facklam & Edwards, 1979; Facklam *et al.*, 1999, 2000, 2002).

The 'T-agglutination patterns' column gives the T type or T pattern observed for each M/*emm* type, followed in parentheses by the number of isolates with that pattern. When multiple T patterns are found within the same M/*emm* type, patterns considered to be distinctly different are separated by semicolons. Closely related patterns are separated by commas followed by the cumulative number of isolates with those patterns. T patterns not followed by a number in parentheses indicate either a single observation or a very small number of observations. NT indicates T non-typable; Misc (miscellaneous) indicates T patterns that are, as yet, incompletely identified or confirmed. T patterns obtained with reference strains for each M type or *emm* type are indicated by bold, underlined font.

Abbreviations: NA, not applicable; N, negative; P, positive; ND, not done; Neg, negative; Pos, positive. '–' is used in the 'No. of isolates' column for the first member of those pairs of previously designated M types found to have the same *emm*-sequence type [27L/77, 38/40, 44/61, 50/62 and 65/69 (see text)]. Since each of these pairs represents a single *emm* type, the total number of isolates and the T-pattern frequencies for these types are entered only once and arbitrarily listed under the second member of the M-type pair.

M type	*emm* type	SOF	*sof* gene	No. of isolates	T-agglutination patterns
1	1	N	Neg	4107	**1** (3,979); NT (110); 1/3/13/B3264 (18)
2	2	P	Pos	742	**2** (528); 2/28 (146); 8/25/Imp19 (56); NT (12)
3	3	N	Neg	2171	**3/13/B3264** (1,959); NT (202); 1 (10)
4	4	P	Pos	1543	**4** (1,478); NT (30); 8/25/Imp19 (16); 4/28 (14); 3/13 (5)
5	5	N	Neg	616	**5/27/44** (442); NT (166); 11/12 (8)
6	6	N	Neg	1412	**6** (1,283); NT (129)
7	NA	NA	NA	0	M-7 does not exist. Determined to be a group C streptococcus (See text).
8	8	P	Pos	18	**8/25/Imp19** (15); T14 (2); NT (1)
9	9	P	Pos	142	**9** (97); 14 (20); 5/9 (17); 9/3/B3264 (3); NT (4); 11/12
10	NA	NA	NA	0	M-10 does not exist. Determined to be the same as M-12 (See text).
11	11	P	Pos	339	**11**, 11/12 (311); NT (26); 28 (2)
12	12	N	Pos	2585	**12**, 11/12 (2,396); NT (189)
13	13	P	Pos	4	**3/13/B3264** (4)
14	14	N	Neg	13	**14** (6); NT (5); 11 (2)
15	15	N	Neg	4	3/13/B3264 (3); 8/25 (1); **23**, **23/8/14**
16	NA	NA	NA	0	M-16 does not exist. Determined to be a group G streptococcus (See text).
17	17	N	ND	0	**23**, 23/8/14
18	18	N	Neg	456	**18** (127); NT (192); 8/Imp19/27 (125); 14 (7); 8/25, 5/27/44, 23 (5); **9/18**
19	19	N	Neg	9	23 (7); **NT** (2); **9/18**
20	NA	NA	NA	0	M-20 does not exist. Determined to be a group C streptococcus (See text).
21	NA	NA	NA	0	M-21 does not exist. Determined to be a group C streptococcus (See text).
22	22	P	Pos	681	12 (391); 12/3/13/B3264, 13/B3264 (219); 11/12 (50); NT (17); **22** (4)
23	23	N	ND	2	**23**, NT
24	24	N	ND	6	**4** (4); NT (2)
25	25	P	Pos	38	**8/25/Imp19** (20); NT (9); 1 (5); 14 (3); 11/12
26	26	N	ND	4	**NT** (4)
27G	27G	P	Pos	26	**5/27/44** (26); NT (see comment in text)
27L	27L/77	P	Pos	–	**5/27/44** (see *emm*-77 and comment in text)
28	28	P	Pos	1412	**28**, 4/28 (1,329); NT (35); 11/28, 12/28, 8/28 (37); 3/13/B3264 (9); 4
29	29	N	Neg	22	28, **4/28** (13); **NT** (9)
30	30	N	Neg	2	**23**, **NT**
31	31	N	Neg	28	8/25/Imp19 (14); **NT** (14)
32	32	N	Neg	9	9 (7); **NT** (2); **8/14/23**
33	33	N	Neg	72	**3/13/B3264** (60); NT (8); 8/25 (4)
34	34	N	Neg	2	28, **3/13/28**, NT
35	NA	NA	NA	0	M-35 does not exist. Determined to be the same as M-49 (See text).
36	36	N	Neg	3	**NT**, 8/12
37	37	N	Neg	0	**NT**

Table 1. cont.

M type	*emm* type	SOF	*sof* gene	No. of isolates	T-agglutination patterns
38	38/40	N	Neg	−	**NT** (see *emm*-40 and comment in text)
39	39	N	Neg	6	3/13/B3264 (6); **NT**
40	38/40	N	Neg	1	**NT** (See *emm*-38 and comment in text)
41	41	N	Neg	95	**3/13/B3264** (78); **NT** (17); Misc (4)
42	42	N	Neg	6	**3/13/B3264** (6)
43	43	N	Neg	88	**3/13/B3264** (68); NT (18); Misc (2)
44	44/61	P	Pos	−	**5/27/44** (See *emm*-61 and comment in text)
45	NA	NA	NA	0	M-45 does not exist. It is the same as M-24 (See text).
46	46	N	Neg	1	**4**; 9
47	47	N	Neg	0	**23**
48	48	P	Pos	42	**4/28** (36); NT (6); Misc (3)
49	49	P	Pos	185	**14** (98); NT (56); 8/14/25/Imp19 (18); 3/13/B3264 (9); 12; Misc (3)
50	50/62	N	ND	−	**NT** (See *emm*-62 and comment in text)
51	51	N	Neg	0	**14**
52	52	N	Neg	20	**3/13/B3264** (15); 14; 8/25/Imp19
53	53	N	Neg	59	**3/13/B3264** (39); **NT** (7); 1 (3); Misc (10)
54	54	N	Neg	7	28 (3); 3/13/B3264 (2); **NT** (2); **23**
55	55	N	Neg	8	**8/25/Imp19** (5); NT (3)
56	56	N	Neg	15	**NT** (9); 3/13/B3264 (2); 3/13/28 (4)
57	57	N	Neg	7	**NT** (5); **8/25**; 3
58	58	P	Pos	158	**8/25/Imp19** (81); NT (45); 2/28 (18); 2/8/25 (11); 12 (2); 14/25
59	59	P	Pos	41	**11/12** (27); NT (11); 14 (2); Misc (1); **14/25/Imp19**
60	60	P	Pos	26	**4** (16); 4/28 (6); NT (4)
61	44/61	P	Pos	198	5/27/44 (146); 11/12 (24); NT (21); 8/25/Imp19 (7) **11**, **9/11** (See *emm*-44 and text)
62	50/62	P	Pos	18	**12**, 12/B3264 (15); 11/12 (3) (See *emm*-50 and comment in text)
63	63	P	Pos	22	**4** (9); 6 (8); NT (4); Misc (1)
64	64	N	Neg	11	NT (5); 3, 3/13/28 (4); 14; 11/12; **8/14** [previously incorrectly identified as SOF positive (Johnson & Kaplan, 1993)]
65	65/69	N	Neg	−	**8/25/Imp19** (see *emm*-69 and comment in text)
66	66	P	Pos	44	12/13/B3264 (24); **12**, 11/12 (19); 4
67	67	N	Neg	3	B3264, 3/13; NT; **3/13/28**, **28**
68	68	P	Pos	34	3/13/B3264 (16); 12, 11/12 (5); **1** (6); 6 (3); 4 (2); NT (2)
69	65/69	N	Neg	9	8/25/Imp19, 2/25 (3); **3/13/B3264** (3); NT (3) (see *emm*-65 and comment in text)
70	70	N	Neg	2	**28**; 3/13
71	71	N	Neg	14	14 (6); NT (5); 3/13, **9/3/13/B3264**; 5/11
72	72	N	Neg	0	**12**, **12/B3264**, 12/3/13, 3/13/B3264
73	73	P	Pos	165	**3/13/B3264** (156); NT (6); Misc (3)
74	74	N	Neg	36	3/13/B3264 (18); NT (8); **9** (5); 8 (5)
75	75	P	Pos	912	**8/25/Imp19** (885); NT (18); 14 (4); 13 (3); 6; 2/8/14
76	76	P	Pos	78	8/25/Imp19 (40); **12**, 11/12, 12/B3264 (30); 22; NT
77	27L/77	P	Pos	505	**3/13/B3264** (209); 28, 13/28 (194); 9, 9/13/28 (45); NT (19); 8/25, 8/28 (10); 5/27/44 (8); 2 (7); 11/12 (3); Misc (10) (see *emm*-27L and comment in text)
78	78	P	Pos	186	**11**, 11/12 (158); NT (16); 3/13, 3/13/B3264/5/27/44 (10); 5/11/27; 14/25
79	79	P	Pos	7	11/12 (5); 4 (2); **25/Imp19**, **NT**
80	80	N	Neg	17	**14**, 14/B3264 (14); 12 (2); 3
81	81	P	Pos	71	**3/13/B3264** (37); NT (11); 12, 12/B3264 (5); 8 (5); 4 (4); 6 (3); 14 (2); 23; Misc (3)
82	82	P	Pos	202	**5/27/44** (154); NT (32); 3/13/B3264 (11); 11/12 (3); 4 (1); 8/25 (1)
83	83	N	Neg	77	**3/13/B3264** (67); NT (10)
84	84	P	Pos	0	**8/25/Imp19**
85	85	P	Pos	32	B3264, **3/13/B3264** (24); NT (5); 1 (2); 8/25 (1)
86	86	N	Neg	4	**3/13** (2); 1/13/27/28 (1); NT (1)
87	87	P	Pos	94	**28** (80); NT (8); 11/12 (4); Imp19 (1); 6 (1)
88	88	P	Pos	7	**NT** (3); 28 (2); 4/8 (1); 8/9 (1)

Table 1. cont.

M type	*emm* type	SOF	*sof* gene	No. of isolates	T-agglutination patterns
89	89	P	Pos	518	**11**, 11/12 (314); 3/13/B3264 (115); NT (86); 4; 28; 27
90	90	P	Pos	4	13, **3/13/B3264** (3); 1
91	91	N	Neg	6	3/13/B3264 (4); 11/28; 12/28; **5/27/44**
92	92	P	Pos	98	8/25/Imp19, **Imp19** (86); 3/13/B3264 (5); NT (5); 28 (2)
93	93	N	Neg	4	**3/13/B3264** (3); NT
	94	P	Pos	81	**3/13/B3264**, **B3264** (77); NT (2); 6; 11
	95	N	Neg	9	**6** (5); **NT** (4)
	96	P	Pos	5	**3** (4); 28
	97	N	Neg	2	**NT** (2); **3/13/14**
	98	N	Neg	7	3 (5); **14/B3264**, **3/13/14**; NT
	99	N	Neg	9	**3**, 3/13/B3264, B3264 (7); NT (2); **14/B3264**
	100	N	Neg	20	28 (8); 6/28 (4); **NT** (7); 13/B3264
	101	N	Neg	45	**3/13/B3264** (31); 8/25/Imp19 (6); NT (6); 14; 28
	102	P	Pos	60	**3/13/B3264** (25); NT (18); 11, **12**, 11/12 (12); 14 (2); Imp19; 1; 12/B; 12/25
	103	P	Pos	30	3/13/B3264 (23); **NT** (4); 12 (2); 6
	104	P	Pos	9	**NT** (5); **3/B3264** (2); 11/12 (2)
	105	N	Neg	7	**1** (3); NT (3); 3/B3264
	106	P	Pos	17	3/13/B3264/5/27/44 (7); 3/13/B3264 (4); **5/27/44** (3); 4 (2); 8/9; **NT**
	107	P	Pos	2	**14**; **NT**
	108	N	ND	16	28 (3); 8/25 (3); 6 (3); NT (3); 5/27/44 (2); **B3264**; 13; **14**
	109	P	Pos	13	**6** (11); 28 (2)
	110	P	Pos	12	**NT** (6); 6 (2); **4** (2); 3/13/B3264; 12
	111	N	Neg	9	**28** (4); **NT** (4); 3/12/14
	112	P	Pos	26	**11/12** (9); **12** (5); 8, 8/25 (7); NT (5)
	113	P	Pos	4	**12** (3); **NT**
	114	P	Pos	129	NT (66); **11/12** (47); 14 (10); **12/B3264**; 5/11; 9; Imp19; 5; 1
	115	N	Neg	13	**NT** (10); Imp19; 9; 13
	116	N	Neg	7	**5/27/44** (3); NT (3); 14/25
	117	P	Pos	7	11/12, **12** (4); NT (2)
	118	P	Pos	38	**3/13/B3264** (22); NT (8); 6 (4); 12 (2); 4; 5/27/44/3
	119	N	Neg	7	**NT** (4); 14; 3/13/B3264; 12
	120	N	Neg	8	**3** (6); 13/14 (2)
	121	N	Neg	4	3/13/B3264/14; **14/B3264**; 13/14; 13/28
	122	N	Neg	10	**3/13/B3264** (8); NT (2)
	123	N	Neg	5	12 (2); **13**; 14; **NT**
	124	P	Pos	9	**11/12**, **12** (8); NT

The identification of the gene encoding SOF, the *sof* gene, and the subsequent development of tests to detect this gene by PCR, have added another tool for GAS characterization (Beall *et al.*, 2000; Rakonjac *et al.*, 1995). There is a nearly complete correlation of *sof*-gene presence with SOF production, making this molecular test a practical alternative to classical phenotypic SOF determination (see Table 1). Of interest is the observation that several SOF-negative isolates tested have been found to carry a *sof* gene (Beall *et al.*, 2000; Rakonjac *et al.*, 1995). Among SOF-negative strains analysed in the present report, only those of M/*emm* type 12 were found to have a *sof* gene, predicted to be truncated and not cell wall associated due to the presence of a frame-shift mutation (Jeng *et al.*, 2003).

SOF inhibition serotyping as a surrogate for M or *emm* typing

Streptococci of different M/*emm* types produce immunologically distinct forms of SOF, and the resulting serum opacity reaction can be inhibited by anti-SOF antibodies in a type-specific manner (Top & Wannamaker, 1968). This type-specific inhibition of SOF was developed into a typing system with very high correlation to M type (Maxted *et al.*, 1973). Many laboratories have used SOF-inhibition typing as a surrogate for M typing when appropriate M antisera were not available. However, recent studies have demonstrated that multiple *sof*-gene types do occur within a single M/*emm* type, and that a single *sof* sequence can be associated

with more than one *emm* type (Beall *et al.*, 2000). Therefore, although SOF-inhibition serotyping remains a useful GAS classification tool, it must be interpreted carefully, with the understanding that in some instances the SOF-inhibition type will not accurately predict M/*emm* type, especially when strains from diverse sources are analysed.

Recognition of new *emm* types

The development of *emm* typing has made GAS M-type identification accessible to virtually all laboratories with DNA-sequencing capability. Further, compared to the laborious serological process required for recognition of new M types (Facklam *et al.*, 2002), recognition and differentiation of possible new *emm* types can be relatively simple. With many laboratories worldwide involved in *emm* typing, a need has arisen to establish criteria for and a process by which new *emm*-sequence types can be proposed and validated in an orderly and systematic manner. This was addressed in 1997 and 1999 by a Working Group composed of representatives of six international streptococcal reference centres (Facklam *et al.*, 1999, 2000, 2002). This Working Group established guidelines for the recognition of new *emm* types, and they also agreed that the official, web-searchable database of *emm*-gene sequences would be maintained at the CDC (http://www.cdc.gov/ncidod/biotech/strep/emmtypes.htm).

Results and observations

The dramatic impact of *emm* typing becomes apparent when one considers that it took more than 50 years from the designation of M type 1 by Griffith in 1926 to the designation of M-81 in 1976 (Facklam & Edwards, 1979; Griffith, 1926). By contrast, only 6 years after demonstrating in 1996 that *emm* typing could be used as a reliable substitute for M typing, the number of formally recognized M/*emm* types rose to 124, an increase of more than 50 % (Facklam *et al.*, 2002). A comprehensive list of T-type and SOF-production correlations with M type was published in 1993, several years prior to the introduction and widespread use of *emm* typing and the dramatic increase in recognized types (Johnson & Kaplan, 1993). Although limited information about the T types and SOF production associated with these new *emm* types can be found in scattered publications and on the internet, there has been no published, comprehensive source of this information. This report provides an updated compilation of current M and *emm* types, together with information about their corresponding T patterns and SOF production, as extracted from the records of two streptococcal reference laboratories and based on typing results of more than 40 000 GAS strains collected worldwide during the past 50 years.

Table 1 lists M and/or *emm* types, SOF production, *sof*-gene amplification results, the number of isolates analysed, and the T-agglutination patterns observed. Clarifying comments, when necessary, are included in the 'T-agglutination patterns' column or in the legend to Table 1. Note that the numbers given in the 'No. of isolates' column represent approximately 21 000 results available in computerized databases maintained at the UMN since 1986 and at the CDC since 1995, and are only a subset of all results available. Therefore, the relative M-/*emm*-type frequencies may be biased toward more recent years. Also, because isolates sent to these two laboratories include outbreaks or surveys conducted in specific geographic areas, the frequency distribution may be influenced by selection bias.

Note that for completeness, M types 7, 10, 16, 20, 21, 35 and 45 are listed in Table 1, even though they do not exist; either the reference strains for these types are not GAS or the M type is identical to another officially designated M type (Rotta, 1978).

Several instances in which different, previously designated M types were found to share identical or near-identical 5′ *emm* sequences have been described (Beall *et al.*, 1996, 2000; Facklam & Beall, 1997; Whatmore *et al.*, 1994). M-type pairs with shared *emm* types described in Table 1 are Lancefield's M-27 and M-77 (note that Griffith's type 27, designated 27G in the table, has a unique *emm* sequence), M-38 and M-40, M-44 and M-61, and M-50 and M-62. M types 65 and 69 have nearly identical *emm* sequences. Two of these pairs, 27L/77 and 44/61, share identical *emm*-sequence types. In addition, limited experimentation indicates that representative reference strains of these pairs and clinical isolates of corresponding T types test positively with sera prepared against both members of the pairs (Beall *et al.*, 2000). M types 38 and 40 have been reported to have identical corresponding 5′ *emm* sequences, but yet react type-specifically with M-38 and M-40 antisera (Beall *et al.*, 1996). The explanation for this possible lack of concordance between *emm* sequence and M specificity is unknown at this time.

Conclusions

The development of *emm* typing has brought a powerful tool to the study of GAS. It allows characterization of essentially all GAS, with results that can be directly related to the extensive history of GAS research when M typing was the primary tool available, and the relative ease of characterization has resulted in a dramatic increase in the number of identified M and *emm* types. Even with this powerful molecular tool, T-agglutination type and SOF-production information continue to be important and relevant, providing the potential for improved efficiency, validation of results, and increased discrimination of clonal differences within *emm* types. However, dissemination of T-pattern and SOF-production information has not kept pace with developments resulting from *emm*-sequencing technology. The T-pattern/SOF correlations provided in this report, covering not only the large number of new types established as a result of *emm*-typing technology but also updated information about classical types, will provide an important resource for epidemiologists and laboratory scientists as they seek to further unravel the still existing mysteries about the pathogenesis of group A streptococcal

infections and their suppurative and non-suppurative sequelae.

REFERENCES

Beall, B., Facklam, R. & Thompson, T. (1996). Sequencing *emm*-specific PCR products for routine and accurate typing of group A streptococci. *J Clin Microbiol* **34**, 953–958.

Beall, B., Facklam, R. R., Elliott, J. A., Franklin, A. R., Hoenes, T., Jackson, D., Laclaire, L., Thompson, T. & Viswanathan, R. (1998). Streptococcal *emm* types associated with T-agglutination types and the use of conserved *emm* gene restriction fragment patterns for subtyping group A streptococci. *J Med Microbiol* **4**, 1–5.

Beall, B., Gherardi, G., Lovgren, M., Facklam, R., Forwick, B. & Tyrrell, G. (2000). *emm* and *sof* gene sequence variation in relation to serological typing of opacity-factor-positive group A streptococci. *Microbiology* **146**, 1195–1209.

Bessen, D. E. & Fischetti, V. A. (1990). Differentiation between two biologically distinct classes of group A streptococci by limited substitutions of amino acids within the shared region of M protein-like molecules. *J Exp Med* **172**, 1757–1764.

Bessen, D., Jones, K. & Fischetti, V. (1989). Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. *J Exp Med* **169**, 269–283.

Espinosa, L. E., Li, Z., Gomez Barreto, D., Calderon Jaimes, E., Rodriguez, R. S., Sakota, V., Facklam, R. R. & Beall, B. (2003). M protein gene type distribution among group A streptococcal clinical isolates recovered in Mexico City, Mexico, from 1991 to 2000, and Durango, Mexico, from 1998 to 1999: overlap with type distribution within the United States. *J Clin Microbiol* **41**, 373–378.

Facklam, R. & Beall, B. (1997). Anomalies in *emm* typing of Group A Streptococci. In *Streptococci and the Host*, pp. 335–337. Edited by E. A. Horaud. New York: Plenum Press.

Facklam, R. & Edwards, L. (1979). A reference laboratory's investigations of proposed M-type strains of *Streptococcus pyogenes*, capsular types of *S. agalactiae*, and new group antigens of streptococci. In *Pathogenic Streptococci*, pp. 251–253. Edited by M. T. Parker. Chertsey, Surrey, UK: Reedbooks.

Facklam, R., Beall, B., Efstratiou, A. & 13 other authors (1999). *emm* typing and validation of provisional M types for group A streptococci. *Emerg Infect Dis* **5**, 247–253.

Facklam, R., Beall, B., Efstratiou, A., Kriz, P., Tyrrell, G., Lovgren, M., Martin, D. & Kaplan, E. (2000). Nomenclature of new *emm* types of group A streptococci. In *Streptococci and Streptococcal Diseases: Entering the New Millennium*, pp. 797–803. Edited by D. R. Martin & J. R. Tagg. Wellington: Securacopy.

Facklam, R., Martin, D. R., Lovgren, M. & 8 other authors (2002). Extension of the Lancefield classification for group A streptococci by addition of 22 new M protein gene sequence types from clinical isolates: *emm* 103 to *emm* 124. *Clin Infect Dis* **34**, 28–38.

Griffith, F. (1926). Types of haemolytic streptococci in relation to scarlet fever. *J Hyg* **25**, 385–397.

Jeng, A., Sakota, V., Li, Z., Datta, V., Beall, B. & Nizet, V. (2003). Molecular genetic analysis of a group A *Streptococcus* operon encoding serum opacity factor and a novel fibronectin-binding protein, SfbX. *J Bacteriol* **185**, 1208–1217.

Johnson, D. R. & Kaplan, E. L. (1988). Microtechnique for serum opacity factor characterization of group A streptococci adaptable to the use of human sera. *J Clin Microbiol* **26**, 2025–2030.

Johnson, D. & Kaplan, E. (1993). A review of the correlation of T-agglutination patterns and M-protein typing and opacity factor production in the identification of group A streptococci. *J Med Microbiol* **38**, 311–315.

Johnson, D. R., Kaplan, E. L., Sramek, J., Bicova, R., Havlicek, J., Havlickova, H., Motlova, J. & Kriz, P. (1996). *Laboratory Diagnosis of Group A Streptococcal Infections*, p. 111. Geneva, Switzerland: World Health Organization.

Maxted, W. R., Widdowson, J. P., Fraser, C. A. M., Ball, L. C. & Bassett, D. C. J. (1973). The use of the serum opacity reaction in the typing of group-A streptococci. *J Med Microbiol* **6**, 83–90.

McGregor, K. F., Spratt, B. G., Kalia, A., Bennett, A., Bilek, N., Beall, B. & Bessen, D. E. (2004). Multilocus sequence typing of *Streptococcus pyogenes* representing most known *emm* types and distinctions among subpopulation genetic structures. *J Bacteriol* **186**, 4285–4294.

Moody, M. D., Padula, J., Lizana, D. & Hall, C. T. (1965). Epidemiologic characterization of group A streptococci by T-agglutination and M-precipitation tests in the public health laboratory. *Health Lab Sci* **2**, 149–162.

Rakonjac, J. A., Robbins, J. C. & Fischetti, V. A. (1995). DNA sequence of the serum opacity factor of group A streptococci: identification of a fibronectin-binding repeat domain. *Infect Immun* **63**, 622–631.

Rehder, C. D., Johnson, D. R. & Kaplan, E. L. (1995). Comparison of methods for obtaining Serum Opacity Factor from group A streptococci. *J Clin Microbiol* **33**, 2963–2967.

Rotta, J. (1978). Group and type (groups A and B) identification of haemolytic streptococci. In *Methods in Microbiology*, pp. 177–198. Edited by T. Bergan & J. R. Norris. London: Academic Press.

Rotta, J., Krause, R. M., Lancefield, R. C., Everly, W. & Lackland, H. (1971). New approaches for the laboratory recognition of M-types of group A streptococci. *J Exp Med* **134**, 1298–1315.

Swift, H. F., Wilson, A. T. & Lancefield, R. C. (1943). Typing group A hemolytic streptococci by M precipitin reactions in capillary pipettes. *J Exp Med* **78**, 127–133.

Teixeira, L. M., Barros, R. R., Castro, A. C., Peralta, J. M., Da Gloria, S. C. M., Talkington, D. F., Vivoni, A. M., Facklam, R. R. & Beall, B. (2001). Genetic and phenotypic features of *Streptococcus pyogenes* strains isolated in Brazil that harbor new *emm* sequences. *J Clin Microbiol* **39**, 3290–3295.

Top, F. & Wannamaker, L. (1968). The serum opacity reaction of *Streptococcus pyogenes*: the demonstration of multiple, strain-specific lipoproteinase antigens. *J Exp Med* **127**, 1013–1034.

Whatmore, A., Kapur, V., Sullivan, D., Musser, J. & Kehoe, M. (1994). Non-congruent relationships between variation in *emm* gene sequences and the population structure of group A streptococci. *Mol Microbiol* **14**, 619–631.