Mol Gen Genet (2000) 262: 965–976

ORIGINAL PAPER

A. Geyer · K.-H. Schmidt

Genetic organisation of the M protein region in human isolates of group C and G streptococci: two types of multigene regulator-like (*mgrC*) regions

Received: 16 March 1999 / Accepted: 23 September 1999

Abstract In addition to β-haemolytic streptococci belonging to Lancefield group A (*Streptococcus pyogenes*, GAS), human isolates of group C (GCS) and group G (GGS) streptococci (*S. dysgalactiae* subsp. *equisimilis*) have been implicated as causative agents in outbreaks of purulent pharyngitis, of wound infections and recently also of streptococcal toxic shock-like syndrome. Very little is known about the organisation of the genomic region in which the *emm* gene of GCS and GGS is located. We have investigated the genome sequences flanking the *emm* gene in GCS by sequencing neighbouring fragments obtained by inverse PCR. Our sequence data for GCS strains 25287 and H46A revealed two types of arrangement in the *emm* region, which differ significantly from the known types of *mga* regulon in GAS. We named this segment of the genome *mgrC* (for *m*ultigene *r*egulon-like segment in group *C* streptococci). In strains belonging to the first *mgrC* type (prototype strain 25287) the *emm* gene is flanked upstream by *mgc*, a gene that is 61% identical to the *mga* gene of GAS. A phylogenetic analysis of the deduced protein sequences showed that Mgc is related to Mga proteins of various types of GAS but forms a distinct cluster. Downstream of *emm*, the *mgrC* sequence region is bordered by *rel*. This gene encodes a protein that functions in the synthesis and degradation of guanosine 3′,5′ bipyrophosphate (ppGpp) during the stringent regulatory response to amino acid deprivation. In the second *mgrC* type (prototype strain H46A), the genes *mgc* and *emm* are arranged as in type 1. But an additional ORF (*orf*) is inserted in opposite orientation between *emm* and *rel*. This *orf* shows sequence homology to *cpdB*, which is present in various microorganisms and encodes 2′,3′ cyclo-nucleotide 2′-phosphodiesterase. PCR analysis showed that these two *mgrC* arrangements also exist in GGS. Our sequence and PCR data further showed that both types of *mgrC* region in GCS and GGS are linked via *rel* to the streptokinase region characterised recently in strain H46A. A gene encoding C5a peptidase, which is present at the 3′ end of the *mga* regulon in GAS, was not found in the *mgrC* region identified in the GCS and GGS strains investigated here.

Key words Group C streptococci · Group G streptococci · Multigene regulator · Streptokinase gene · *Streptococcus dysgalactiae* subsp. *equisimilis* · M protein gene

Communicated by W. Goebel

A. Geyer · K.-H. Schmidt (✉)
Institute of Medical Microbiology,
University Hospital, Friedrich-Schiller-University,
Semmelweissstr. 4, D-07740 Jena, Germany
E-mail: khschmidt@bach.med.uni-jena.de
Tel.: +49-3641-934223; Fax: +49-3641-933474

Introduction

Besides *Streptococcus pyogenes*, which belongs to Lancefield group A (GAS), human isolates of group C β-haemolytic streptococci have been implicated as causative agents in outbreaks of purulent pharyngitis (Efstratiou et al. 1989; Fox et al. 1993; Turner et al. 1993; Bisno et al. 1996). They have also been isolated from wound infections, and a few case reports were published recently which describe a link between human isolates of groups C and G and streptococcal toxic-shock syndrome (Keiser and Campbell 1992; Miller et al. 1996; Natoli et al. 1996; Roth et al. 1999). In this paper we use the abbreviations GCS and GGS for human isolates of group C and G streptococci, respectively, which were designated as *Streptococcus dysgalactiae*, subsp. *equisimilis*. Such strains are classified as β-haemolytic and their streptokinase activates human plasminogen (Vandamme et al. 1996).

It has been shown that GCS and GGS isolated from human sources can express M proteins that show high sequence homology in the conserved C-terminal region to M proteins of class CI in group A strepto-

cocci (Ben Nasr et al. 1994; Schnitzler et al. 1995; Geyer et al. 1999). The M proteins of GCS (Geyer et al. 1999) and GGS (Ben Nasr et al. 1994) show similar affinities for several plasma proteins as do M proteins of GAS (Heden and Lindahl 1993; Schmidt and Wadström 1990; Schmidt et al. 1993a, 1995; Åkesson et al. 1994; Carlsson-Wistedt et al. 1995). This implies that GCS and GGS use mechanisms similar to those found in GAS to impede phagocytosis during infection.

In strains of GAS, genes encoding M and M-like proteins, such as *mrp*, *emm*, and *enn*, as well as the gene encoding the C5a peptidase (*scpA*), are closely linked and arranged in tandem. These genes are co-ordinately regulated by the upstream *mga*, a gene that encodes the *trans*-acting multigene regulator Mga (Perez-Casal et al. 1991; Podbielski et al. 1993, 1995; Katerov et al. 1994; Andersson et al. 1996). All these genes are part of a pathogenicity island formerly called the *vir* regulon, but since 1995 the region is referred to as the "*mga* regulon" by agreement among researchers involved in this topic (Scott et al. 1995). The protein Mga is a positive regulator, and inactivation of the coding gene down-regulates expression of M and M-like proteins. The genes encoding M and M-like proteins follow *mga* in tandem at its 3′ end (Podbielski et al. 1995). The architecture of the *mga* regulon varies in different strains of GAS and the length of the *mga*-regulon depends on the number of *emm* and *emm*-like genes located between the genes *mga* and *scpA* (Haanes et al. 1992; Podbielski 1993).

In contrast to GAS, the linkage relationships of *emm* genes in GCS and GGS have not hitherto been reported. Although the gene *scpG*, which is homologous to the C5a peptidase gene *scpA* of GAS, has been identified in GGS, the genes *emmG* and *scpG* were found not to be closely linked (Cleary et al. 1991; Sriprakash and Hartas 1996). So far, attempts to identify an *mga*-like gene in GCS or GGS have failed (Podbielski et al. 1992; Sriprakash and Hartas 1996). A large 8931-bp segment consisting of genes adjacent to the streptokinase gene *skc* in the chromosome of the GCS strain H46A was recently described (Mechold et al. 1993), but these authors did not find a linkage between the *skc* region in strain H46A and the segment carrying the *emm* gene.

In this paper we present sequence data from GCS showing that an *mga*-like gene is found upstream of *emm*. We named this gene *mgc*. At the 3′ end *emm* is flanked in some strains by *rel* and in a second group of strains by *cpdB* and *rel*. On the basis of these results we defined two types of *mga*-regulon-like genome segments in GCS which we have named *mgrC* regions. The *mgrC* segments were significantly different from the known types of *mga* regulons in GAS. PCR analysis revealed that in GGS both types of *mgrC* regions also occur. Our sequence and PCR data further showed that the gene encoding C5a peptidase is not located in the *mgrC* segment of the GCS and GGS genomes.

Materials and methods

Bacterial strains, plasmids, and culture conditions

Human clinical isolates of *S. dysgalactiae* subsp. *equisimilis* group C strain (GCS) 25287, T-type 4 (Geyer et al. 1999), and the group G strains (GGS) 39072, 39086 and 40104 were used in this study. The GCS strain H46A, a streptokinase-producing strain (Christensen 1945), as well as the GAS strains M1 40/58 and M3 4/55, were also used. All strains were obtained from the strain collection of the Streptococcal Laboratory at the University of Jena, Germany. The strains were stored at −70° C in Todd-Hewitt broth supplemented with 20% calf serum and 20% glycerol, or in lyophilised form. For isolation of total DNA, streptococci were cultured at 37° C overnight in 50 ml of Todd-Hewitt Broth. The *Escherichia coli* strains TG1, JM109 (Sambrook et al. 1989), Epicurian *coli* XL1-Blue MRF'Kan (Stratagene, Heidelberg, Germany) and TOP10 (Invitrogen BV, NV Leek, Netherlands) were used for subcloning experiments. The pGEM-T Easy vector system (Promega, Madison, Wis.), the pCR-Script vector (Stratagene) and the pCR TOPO XL plasmid (Invitrogen) were used as plasmid vectors for subcloning of PCR products.

Recombinant DNA techniques and sequencing

Isolation of plasmid DNA from *E. coli*, DNA restriction, ligation and related experiments were performed according to standard techniques (Sambrook et al. 1989) or as recommended by the manufacturers of cloning kits. Total DNA from streptococci was isolated according to Caparon and Scott (1991). For progressive DNA sequencing starting from the *emm* gene of GCS strain 25287 (Geyer et al. 1999, AC X93464), templates were produced by the inverse PCR technique described by Triglia et al. (1988). In one case, *Hin*dIII fragments were produced because a *Hin*dIII site is located in the C-repeat region of the *emm* sequence of strain 25287. The fragments were ligated to circularise the DNA and the mixture was used as template for inverse PCR, using primers based on the *emm* gene.

The primers A (5′-CACAAGGATCCGCTGCTCTCAAACA-ACTTGAA-3′, forward) and B (5′-GCCCTGCTACTAACCC-TGCTCC-3′, reverse) were used to obtain upstream sequences. These primers hybridise upstream of the *Hin*dIII site. Primers C (5′-GCACCACAAGCAGGTACAAAACC-3′, forward) and D (5′-GTTTTTCAAGAGCAGCTAAGTTGCTG-3′, reverse) bind downstream of the *Hin*dIII site and were used to walk downstream to the *emm* sequence. Sequencing of this region in GCS strain H46A was then done using PCR fragments produced based on the sequence information obtained from strain 25287. DNA sequencing was performed using the BigDye Terminator sequencing kit and the Genetic Analyser ABI Prism 310 (Perkin Elmer-Applied Biosystems). Other primers used were E (5′-AAGCTTAGCCTT-CAGAGGAAGCC-3′, forward), F (5′-CCTAAGGTTAGCCG-TAACGCACTCATG-3′, reverse), G (5′-GCACCACAAGC-AGGTACAAAACC-3′, forward), H (5′-CAACTTCTTCTCCT-GTTAAATTG-3′, reverse), I (5′-GCACCACAAGCAGGTACA-AAACC-3′, forward) and J (5′-GGGCAAGCGACCCGAAG-GAGAG-3′, reverse).

Database searches and sequence analysis

The EMBL Data Library at the Deutsches Krebsforschungszentrum, Heidelberg was accessed and searched by computer. Nucleic acid and protein sequences were analysed using the Genius/Husar program package.

Nucleotide sequence accession numbers

Nucleotide sequences reported in this paper appear in the EMBL/GenBank/DDBJ databases under the Accession Nos. X93464, Y18363, and AJ133440.

Results

Chromosome walking

The basis of this investigation was the *emm* sequence (Accession No. X93464) of strain 25287 (Geyer et al. 1999). This gene has a *Hin*dIII site at the end of the fourth C repeat at position 989. The chromosomal DNA of strain 25287 was digested with *Hin*dIII, religated and inverse PCR was performed according to Triglia et al. (1988) with the primers described above. Primers A and B produced a fragment of 2.8 kb. With primers C and D a 1-kb fragment was obtained. The 2.8-kb fragment was cloned in pCR-Script and completely sequenced. The 1-kb fragment was sequenced after cloning in the vector pGEM-T Easy. The complete sequence of the "*emm* region" deduced from these data is given in Fig. 1. At the 5′ end of the *emm* gene we found two ORFs. The first gene adjacent to *emm* is in the same orientation and encodes a 513-amino acid protein. DNA sequence alignment revealed about 61% homology to sequences of *mga* published for different M types of GAS. The alignment of the corresponding amino acid sequence with the peptide sequence of Mga (Mry) from the M type 6 strain D471 (Perez-Casal et al. 1991) shows a comparable level of homology (Fig. 2). Therefore we named this gene *mgc*, for *m*ultigene *r*egulator in group *C* streptococci, and the encoded protein Mgc.

The second gene, which we named *nrd*, is oriented oppositely and encodes a protein that has segments of high homology to the ribonucleotide reductase (*nrdI*) of *Lactococcus lactis* (Jordan et al. 1997; Accession No. X92690).

Downstream of *emm* (537 bp away) we found – in the same orientation – the beginning of an ORF identical to *rel* gene of the GCS strain H46A, which has been described by Mechold et al. (1993). The *rel* gene encodes a protein responsible for the synthesis and degradation of guanosine 3′,5′ bipyrophosphate (ppGpp) in various bacterial systems (Metzger et al. 1989; Mechold et al. 1993), and is linked to the streptokinase region in H46A (Mechold et al. 1993). We did not find an *scpA*-like gene in this region.

Investigation by PCR of the regions flanking *emm* in different GCS and GGS strains

To investigate how *nrd*, *mgc*, *emm* and *rel* are arranged in other GCS and GGS strains, we compared the PCR patterns using the corresponding chromosomal DNAs as templates. As controls the DNAs of the GAS strains M1 and M3 were used. The PCR primers E, F, G, and H used for these experiments were based on the sequence shown in Fig. 1. Primers E and F characterise the region between *nrd* and *emm*. The PCRs with GCS and GGS strains generated identical patterns with a main band at 2.6 kb (Fig. 3A). No products were amplified from the DNA of the GAS strains. Thus we conclude (and this was confirmed by partial sequencing of the PCR fragments) that the arrangement of the genes located upstream of *emm* in other GCS and GGS strains is similar to that in 25287.

In a second set of PCRs the primers G and H were used. These primers recognise the segment between the end of *emm* and the start of *rel*. As illustrated in Fig. 3B, two different PCR patterns were obtained. With GCS strain 25287 and the GGS strain 39086 a small PCR fragment of 0.7 kb was amplified. This corresponded to the fragment size which we expected from our sequence data for strain 25287. Partial sequencing of these fragments confirmed the linkage between *emm* and *rel*.

But in the case of the GCS strain H46A and of the GGS strains 39072 and 40104 a larger fragment of about 3 kb was found. No PCR fragments were obtained from the GAS control strains. The 3-kb fragment from strain H46A was cloned in the plasmid pCR TOPO XL and sequenced. In this group of strains we found an additional ORF which was located between the *emm* gene and *rel*, and oriented in the opposite direction. The gene was identified by database searches as a *cpdB*-like gene which encodes a protein of 683 amino acids. The protein showed high sequence homology to 2′,3′-cyclic-nucleotide 2′-phosphodiesterases found in various bacterial species. The sequence information obtained enabled us to sequence the whole *emm*-containing segment of strain H46A from *nrd* to *rel* (Accession No. AJ133440). The strain H46A used in this report came from the same stock culture as was used by Mechold et al. (1993) for sequencing of the *skc* region. In this strain and also in strain 25287 the *emm* containing segment is flanked downstream by *rel*. The latter gene starts the following *skc* region. To test whether the *skc* region is generally found adjacent to this genome segment, a PCR was performed with primers I and J. Primer I starts in *emm* and primes in the 3′ direction, while the reverse primer J, deduced from Mechold et al. (1993), starts inside the *skc* gene and amplifies towards *emm*. Figure 3C shows a comparable pattern of PCR fragments to Fig. 3B, but the products are of higher molecular weight. The fragments obtained from strains 25287 and 39086 (3.3 kb) in this experiment were also smaller then the 5.8-kb fragment found in the other strains. This confirms the result in Fig. 3B, showing that some strains lack the *cpdB* gene in that region. The PCR fragments were partially sequenced. With primer J the corresponding partial *skc* sequence including the stop codon and the following intergenic region was obtained. Primer I recognised 170 bp of *emm* and a part of the downstream intergenic sequence. Because appropriate PCR fragments were obtained from all five investigated GCS/GGS strains, we conclude that the *emm*-containing segment and the *skc* region in GCS and GGS are adjacent.

Sequence alignment of Mgc and M proteins from GCS with the related molecules from GAS

To compare characteristic sequences in the *emm*-flanking regions from GAS and GCS, we performed a multiple

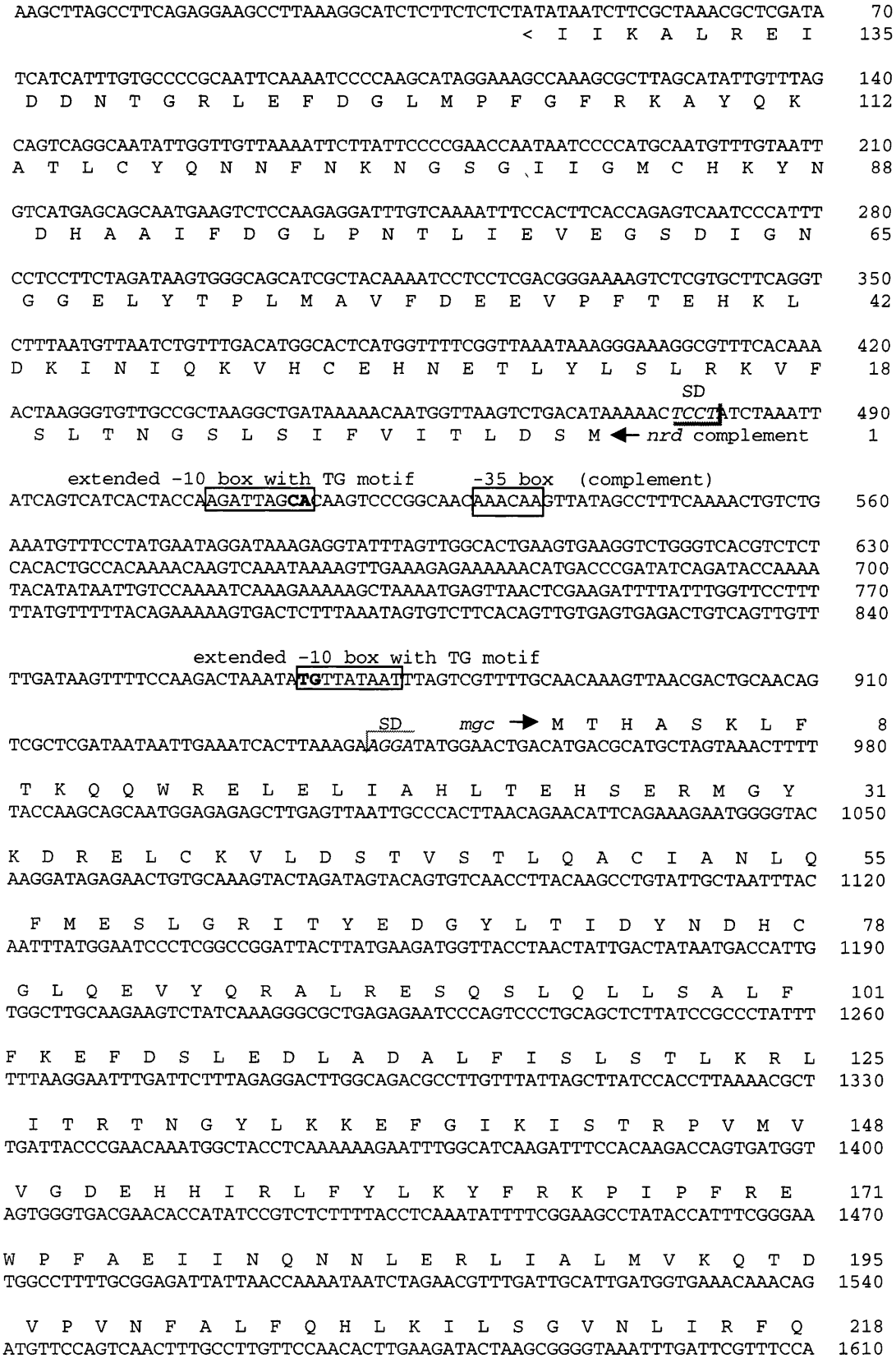
AAGCTTAGCCTTCAGAGGAAGCCTTAAAGGCATCTCTTCTCTCTATATAATCTTCGCTAAACGCTCGATA 70
< I I K A L R E I 135
TCATCATTTGTGCCCCGCAATTCAAAATCCCCAAGCATAGGAAAGCCAAAGCGCTTAGCATATTGTTTAG 140
D D N T G R L E F D G L M P F G F R K A Y Q K 112
CAGTCAGGCAATATTGGTTGTTAAAATTCTTATTCCCCGAACCAATAATCCCCATGCAATGTTTGTAATT 210
A T L C Y Q N N F N K N G S G I I G M C H K Y N 88
GTCATGAGCAGCAATGAAGTCTCCAAGAGGATTTGTCAAAATTTCCACTTCACCAGAGTCAATCCCATTT 280
D H A A I F D G L P N T L I E V E G S D I G N 65
CCTCCTTCTAGATAAGTGGGCAGCATCGCTACAAAATCCTCCTCGACGGGAAAAGTCTCGTGCTTCAGGT 350
G G E L Y T P L M A V F D E E V P F T E H K L 42
CTTTAATGTTAATCTGTTTGACATGGCACTCATGGTTTTCGGTTAAATAAAGGGAAAGGCGTTTCACAAA 420
D K I N I Q K V H C E H N E T L Y L S L R K V F 18
SD
ACTAAGGGTGTTGCCGCTAAGGCTGATAAAAACAATGGTTAAGTCTGACATAAAAACTCCTTATCTAAATT 490
S L T N G S L S I F V I T L D S M ← nrd complement 1
extended -10 box with TG motif -35 box (complement)
ATCAGTCATCACTACCAAGATTAGCACAAGTCCCGGCAACAAACAAGTTATAGCCTTTCAAAACTGTCTG 560
AAATGTTTCCTATGAATAGGATAAAGAGGTATTTAGTTGGCACTGAAGTGAAGGTCTGGGTCACGTCTCT 630
CACACTGCCACAAAACAAGTCAAATAAAAGTTGAAAGAGAAAAAAACATGACCCGATATCAGATACCAAAA 700
TACATATAATTGTCCAAAATCAAAGAAAAAGCTAAAATGAGTTAACTCGAAGATTTTATTTGGTTCCTTT 770
TTATGTTTTTACAGAAAAAGTGACTCTTTAAATAGTGTCTTCACAGTTGTGAGTGAGACTGTCAGTTGTT 840
extended -10 box with TG motif
TTGATAAGTTTTCCAAGACTAAATATGTTATAATTTAGTCGTTTTGCAACAAAGTTAACGACTGCAACAG 910
SD mgc → M T H A S K L F 8
TCGCTCGATAATAATTGAAATCACTTAAAGAAGGATATGGAACTGACATGACGCATGCTAGTAAACTTTT 980
T K Q Q W R E L E L I A H L T E H S E R M G Y 31
TACCAAGCAGCAATGGAGAGAGCTTGAGTTAATTGCCCACTTAACAGAACATTCAGAAAGAATGGGGTAC 1050
K D R E L C K V L D S T V S T L Q A C I A N L Q 55
AAGGATAGAGAACTGTGCAAAGTACTAGATAGTACAGTGTCAACCTTACAAGCCTGTATTGCTAATTTAC 1120
F M E S L G R I T Y E D G Y L T I D Y N D H C 78
AATTTATGGAATCCCTCGGCCGGATTACTTATGAAGATGGTTACCTAACTATTGACTATAATGACCATTG 1190
G L Q E V Y Q R A L R E S Q S L Q L L S A L F 101
TGGCTTGCAAGAAGTCTATCAAAGGGCGCTGAGAGAATCCCAGTCCCTGCAGCTCTTATCCGCCCTATTT 1260
F K E F D S L E D L A D A L F I S L S T L K R L 125
TTTAAGGAATTTGATTCTTTAGAGGACTTGGCAGACGCCTTGTTTATTAGCTTATCCACCTTAAAACGCT 1330
I T R T N G Y L K K E F G I K I S T R P V M V 148
TGATTACCCGAACAAATGGCTACCTCAAAAAAGAATTTGGCATCAAGATTTCCACAAGACCAGTGATGGT 1400
V G D E H H I R L F Y L K Y F R K P I P F R E 171
AGTGGGTGACGAACACCATATCCGTCTCTTTTACCTCAAATATTTTCGGAAGCCTATACCATTTCGGGAA 1470
W P F A E I I N Q N N L E R L I A L M V K Q T D 195
TGGCCTTTTGCGGAGATTATTAACCAAAATAATCTAGAACGTTTGATTGCATTGATGGTGAAACAAACAG 1540
V P V N F A L F Q H L K I L S G V N L I R F Q 218
ATGTTCCAGTCAACTTTGCCTTGTTCCAACACTTGAAGATACTAAGCGGGGTAAATTTGATTCGTTTCCA 1610

Fig. 1 (Contd.)

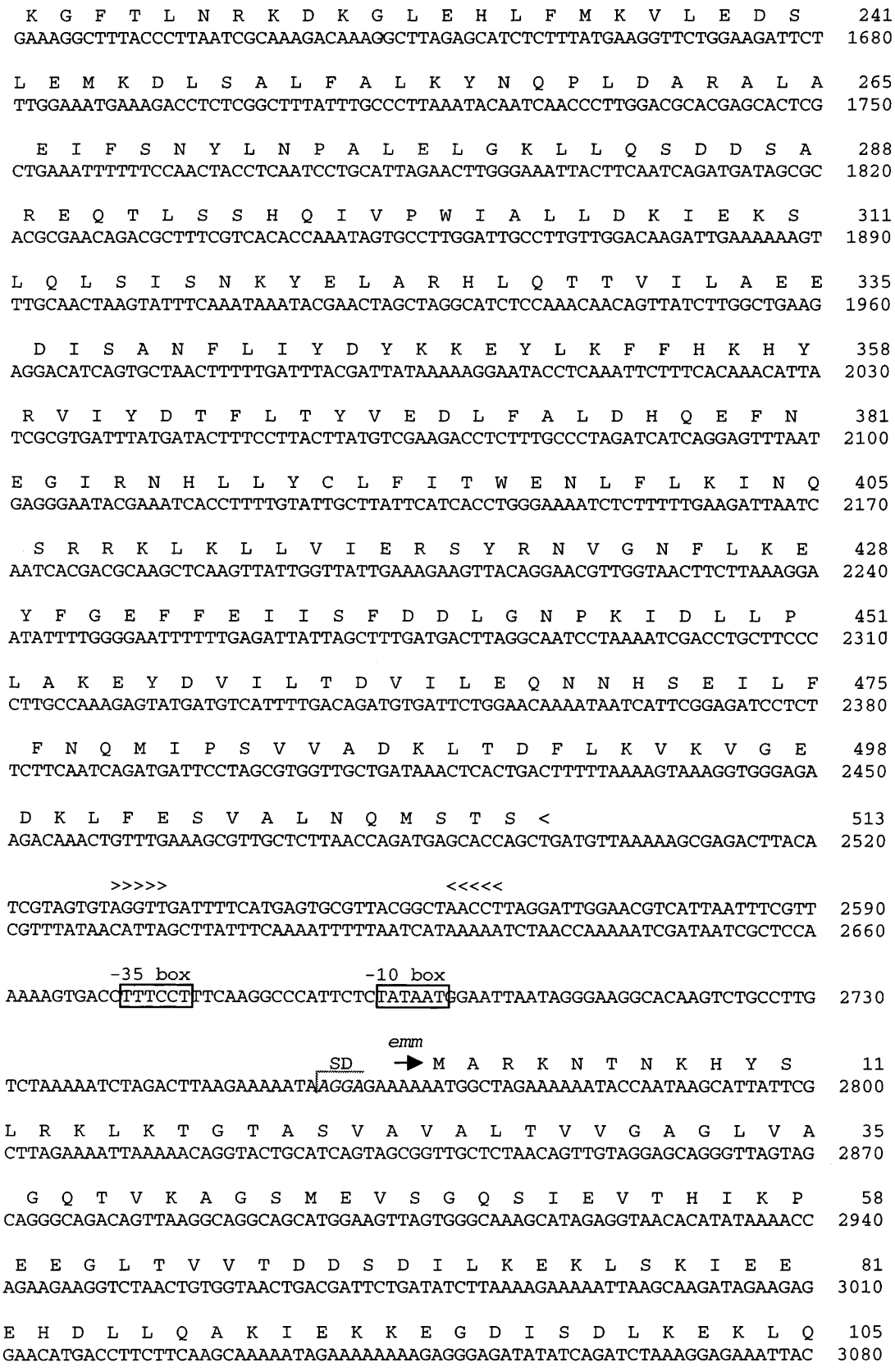
K G F T L N R K D K G L E H L F M K V L E D S 241
GAAAGGCTTTACCCTTAATCGCAAAGACAAAGGCTTAGAGCATCTCTTTATGAAGGTTCTGGAAGATTCT 1680
L E M K D L S A L F A L K Y N Q P L D A R A L A 265
TTGGAAATGAAAGACCTCTCGGCTTTATTTGCCCTTAAATACAATCAACCCTTGGACGCACGAGCACTCG 1750
E I F S N Y L N P A L E L G K L L Q S D D S A 288
CTGAAATTTTTTCCAACTACCTCAATCCTGCATTAGAACTTGGGAAATTACTTCAATCAGATGATAGCGC 1820
R E Q T L S S H Q I V P W I A L L D K I E K S 311
ACGCGAACAGACGCTTTCGTCACACCAAATAGTGCCTTGGATTGCCTTGTTGGACAAGATTGAAAAAAGT 1890
L Q L S I S N K Y E L A R H L Q T T V I L A E E 335
TTGCAACTAAGTATTTCAAATAAATACGAACTAGCTAGGCATCTCCAAACAACAGTTATCTTGGCTGAAG 1960
D I S A N F L I Y D Y K K E Y L K F F H K H Y 358
AGGACATCAGTGCTAACTTTTTGATTTACGATTATAAAAAGGAATACCTCAAATTCTTTCACAAACATTA 2030
R V I Y D T F L T Y V E D L F A L D H Q E F N 381
TCGCGTGATTTATGATACTTTCCTTACTTATGTCGAAGACCTCTTTGCCCTAGATCATCAGGAGTTTAAT 2100
E G I R N H L L Y C L F I T W E N L F L K I N Q 405
GAGGGAATACGAAATCACCTTTTGTATTGCTTATTCATCACCTGGGAAAATCTCTTTTTGAAGATTAATC 2170
S R R K L K L L V I E R S Y R N V G N F L K E 428
AATCACGACGCAAGCTCAAGTTATTGGTTATTGAAAGAAGTTACAGGAACGTTGGTAACTTCTTAAAGGA 2240
Y F G E F F E I I S F D D L G N P K I D L L P 451
ATATTTTGGGGAATTTTTTGAGATTATTAGCTTTGATGACTTAGGCAATCCTAAAATCGACCTGCTTCCC 2310
L A K E Y D V I L T D V I L E Q N N H S E I L F 475
CTTGCCAAAGAGTATGATGTCATTTTGACAGATGTGATTCTGGAACAAAATAATCATTCGGAGATCCTCT 2380
F N Q M I P S V V A D K L T D F L K V K V G E 498
TCTTCAATCAGATGATTCCTAGCGTGGTTGCTGATAAACTCACTGACTTTTTAAAAGTAAAGGTGGGAGA 2450
D K L F E S V A L N Q M S T S < 513
AGACAAACTGTTTGAAAGCGTTGCTCTTAACCAGATGAGCACCAGCTGATGTTAAAAAGCGAGACTTACA 2520
>>>>> <<<<<
TCGTAGTGTAGGTTGATTTTCATGAGTGCGTTACGGCTAACCTTAGGATTGGAACGTCATTAATTTCGTT 2590
CGTTTATAACATTAGCTTATTTCAAAATTTTTAATCATAAAAATCTAACCAAAAATCGATAATCGCTCCA 2660
-35 box -10 box
AAAAGTGACCTTTCCTTTCAAGGCCCATTCTCTATAATGGAATTAATAGGGAAGGCACAAGTCTGCCTTG 2730
emm
SD M A R K N T N K H Y S 11
TCTAAAAATCTAGACTTAAGAAAAATAAGGAGAAAAAAATGGCTAGAAAAAATACCAATAAGCATTATTCG 2800
L R K L K T G T A S V A V A L T V V G A G L V A 35
CTTAGAAAATTAAAAACAGGTACTGCATCAGTAGCGGTTGCTCTAACAGTTGTAGGAGCAGGGTTAGTAG 2870
G Q T V K A G S M E V S G Q S I E V T H I K P 58
CAGGGCAGACAGTTAAGGCAGGCAGCATGGAAGTTAGTGGGCAAAGCATAGAGGTAACACATATAAAACC 2940
E E G L T V V T D D S D I L K E K L S K I E E 81
AGAAGAAGGTCTAACTGTGGTAACTGACGATTCTGATATCTTAAAAGAAAAATTAAGCAAGATAGAAGAG 3010
E H D L L Q A K I E K K E G D I S D L K E K L Q 105
GAACATGACCTTCTTCAAGCAAAAATAGAAAAAAAAGAGGGAGATATATCAGATCTAAAGGAGAAATTAC 3080

Fig. 1 (Contd.)

```
  N  L  K  D  D  K  S  L  A  E  A  G  Y  A  N  S  Y  K  H  H  Q  E  Q    128
AGAATTTAAAGGACGATAAAAGTTTAGCTGAAGCGGGTTATGCTAATTCTTATAAACATCATCAAGAACA  3150

  L  A  E  K  D  K  D  I  S  D  L  K  E  K  L  Q  N  L  K  D  D  K  S    151
ACTTGCTGAAAAAGATAAAGATATATCAGATCTAAAGGAGAAATTACAGAACTTAAAGGACGATAAAAGT  3220

L  A  E  A  G  Y  A  N  S  Y  K  H  H  Q  E  Q  L  A  E  K  D  K  L  N    175
TTAGCTGAAGCGGGTTATGCTAATTCTTATAAACATCATCAAGAACAACTTGCTGAAAAAGATAAACTTA  3290

  A  E  K  D  D  Q  I  K  Q  L  E  K  Q  K  Q  I  L  N  A  S  S  K  G    198
ATGCTGAAAAAGATGATCAAATTAAACAACTAGAAAAGCAAAAACAAATCCTAAACGCTAGCAGTAAAGG  3360

  T  A  R  D  L  E  A  V  R  K  A  K  A  Q  V  E  A  A  L  K  Q  L  E    221
TACAGCACGAGACCTTGAAGCTGTTCGCAAAGCAAAAGCACAAGTTGAAGCTGCTCTCAAACAACTTGAA  3430

E  Q  N  K  I  S  D  A  S  R  Q  G  L  R  R  D  L  N  A  S  R  E  A  K    245
GAACAAAACAAAATCTCAGATGCAAGTCGTCAAGGTCTTCGCCGTGACTTGAACGCATCACGTGAAGCTA  3500

  K  Q  L  E  A  E  H  Q  K  L  E  E  Q  N  K  I  S  E  A  S  R  K  G    268
AGAAACAATTAGAAGCTGAACACCAAAAACTTGAAGAACAAAACAAGATTTCAGAAGCAAGCCGTAAGGG  3570

  L  R  R  D  L  D  A  S  R  E  A  K  K  Q  V  E  K  D  L  A  N  L  T    291
TCTTCGCCGTGACTTGGATGCATCACGTGAAGCTAAGAAACAAGTTGAAAAAGATTTAGCAAACTTGACT  3640

A  E  L  D  N  L  I  R  L  K  E  D  K  Q  I  S  D  A  S  R  K  G  L  R    315
GCTGAACTTGATAACTTGATAAGGCTTAAAGAAGACAAACAAATCTCAGACGCAAGTCGTAAAGGTCTTC  3710

  R  D  L  D  A  S  R  E  A  K  K  Q  V  E  K  A  L  E  E  A  N  S  N    338
GTCGTGACTTGGACGCATCACGTGAAGCTAAGAAACAAGTTGAAAAAGCTTTAGAAGAAGCAAACAGCAA  3780

  L  A  A  L  E  K  L  N  K  E  L  E  E  S  K  K  L  T  E  K  E  K  A    361
CTTAGCTGCTCTTGAAAAACTTAACAAAGAGCTTGAAGAAAGCAAGAAATTAACCGAAAAAGAAAAAGCT  3850

E  L  Q  A  K  L  E  A  E  A  K  L  E  A  E  A  K  A  L  K  E  K  L  A    385
GAGCTACAAGCAAAACTTGAAGCAGAAGCAAAACTTGAAGCAGAAGCAAAAGCACTCAAAGAAAAATTAG  3920

  K  Q  A  E  E  L  A  K  L  R  A  G  K  A  S  D  S  Q  T  P  D  A  K    408
CAAAACAAGCTGAAGAACTTGCAAAACTAAGAGCTGGAAAAGCATCAGACTCACAAACCCCTGATGCAAA  3990

  P  G  N  K  V  V  P  G  K  G  Q  A  P  Q  A  G  T  K  P  N  Q  N  K    431
ACCAGGAAACAAAGTTGTTCCAGGTAAAGGTCAAGCACCACAAGCAGGTACAAAACCTAACCAAAACAAA  4060

E  P  M  K  E  T  K  R  Q  L  P  S  T  G  E  A  T  N  P  F  F  T  A  A    455
GAGCCAATGAAGGAAACTAAGAGACAGTTACCATCAACAGGTGAAGCAACTAACCCATTCTTCACAGCGG  4130

  A  L  A  V  M  A  T  A  G  V  A  A  V  V  K  R  K  E  E  N  <          475
CAGCCCTTGCTGTTATGGCAACAGCTGGAGTAGCAGCAGTTGTAAAACGCAAAGAAGAAAACTAAAAAAA  4200
                     >>>>>>>>>>>       <<<<<<<<<<<
CATTCCTATTCCCAAAATAAGACAAACTAGCTTTCATTAGCTAGTTTGTTTTTTTATGAAGCCAAGCAGA  4270
AAAGCAGCTGAGGAAAATAGACAATAAAACAGATGCCCTCTCCCTAGCTTTAATGGATTAGAGAGAAGGC  4340
ATCTGAGAGTAGTGTCTTAGGCAAGGTGCTAGAGGTCGTCTTCTAGCGAGAGGAAACTTAGTTGCGCTCC  4410
TGTTTGCGACTAAAGACGGCAACTAACATGGCCAAAAAATGGCACTTTTTGACATGTAACCTTTTGTCAT  4480
CATAAAATCTCCTCTTTTTCCTAAAAATGTTACCTACTCATTATAGCTCATGAAAGCGATTGTATCAATT  4550
GAAAATGCGAACTATTTTTAGGTTTTATAAAGAAAATATTTGGTTAATCAAAAAAGAAATGACCTAGTCA  4620
AAAACTTTCGATGACGAATGCTTTCTCTAACCGCTTTCTTTAAGGATTCCTCCTTTTGATTTTGGGGTGA  4690

                                        rel
extended -10 box with TG motif    SD  →  M  A  K  E  I  N  L  T  G  E    10
AAGTGTTAAAATAGGGAAAAGAAATCTTAGACGAGGAAATAAATGGCAAAAGAAATCAATTTAACAGGAG  4760

  E  V  V  A  L  A  A  K  Y  M  N  E  T  D  A  A  F  V  K  K  A still open end
AAGAAGTTGTTGCCTTAGCAGCCAAATACATGAATGAGACCGATGCGGCCTTTGTGAAAAAAGCTT      4826
```

Fig. 1 DNA sequence of the region adjacent to *emm* in GCS strain 25287 (Accession No. Y18363) comprising the genes *nrd*, *mgc*, and *emm*, which is linked to *rel*. The start of each ORF is marked by an *arrow*, putative promoter sequences are *boxed*, Shine-Delgarno sequences are shown in *italics*, and inverted repeats are delineated by *wedges*

alignment (with the Clustal program) using peptide sequences of Mga and Mgc as well as of M proteins. The Mga-M and Mgc-M pairs from the same GAS or GCS strain, respectively, were analysed (Fig. 4A and B). In the phylogenetic tree Mgc sequences are clearly separated from the cluster of Mga proteins. But within the GAS, the analysis splits the Mga sequences into two groups – those from opacity factor-positive and negative strains. In contrast, the corresponding M protein sequences of GCS (MC, MC1) showed high homology to M proteins of class CI of GAS and did not form a distinct branch outside of the GAS (Fig. 4B). The intergenic sequences located upstream of *mgc* and *emm* of strain 25287 did not show any similarity to corresponding segments from GAS (not shown).

Definition of two types of *mgrC* region in GCS and GGS

From the results obtained above at least two different types of *mga* regulon-like segments can be proposed in

Fig. 2 Alignment of the amino acid sequences of Mgc (Accession No. Y18364; this study) and Mga of M type 6 strain D471 (Perez-Casal et al. 1991; AC M58461). The proteins are 51.3% identical and 64.9% similar. Similar results were obtained with Mga proteins of other GAS strains

```
MGC      - MTHASKLFTKQQWRELELIAHLTEHSERMGYKDRELCKVLDSTVS -45
MGA6     - M-YVSKLFTSQQWRELKLISYVTENADAIGVKDKELSKALNISML -44

MGC      - TLQACIANLQFMESLGRITYEDGYLTIDYNDHCGLQEVYQRALRE -90
MGA6     - TLQTCLTNMQFMKEVGGITYKNGYITIWYHQHCGLQEVYQKALRH -89

MGC      - SQSLQLLSALFFKEFDSLEDLADALFISLSTLKRLITRTNGYLKK -135
MGA6     - SQSFKLLETLFFRDFNSLEELAEELFVSLSTLKRLIKKTNAYLTH -134

MGC      - EFGIKISTRPVMVVGDEHHIRLFYLKYFRKPIPFREWPFAEIINQ -180
MGA6     - TFGITILTSPVQVSGDEHQIRLFYLKYFSEAYKISEWPFGEILNL -179

MGC      - NNLERLIALMVKQTDVPVNFALFQHLKILSGVNLIRFQKGFTLNR -225
MGA6     - KNCERLLSLMIKEVDVRVNFTLFQHLKILSSVNLIRYYKGYSAVY -224

MGC      - KDKGLEHLFMKVLEDSLEMKDLSALFALKYNQPLDARALAEIFSN -270
MGA6     - DNKKTSHRFSQLIQSSLETQDLSRLFYLKFGLYLDETTIAEMFSN -269

MGC      - YLNPALELGKLLQS--DDSAREQTLSSHQIVPWIALLDKIEKSLQ -313
MGA6     - HVNDQLEIGYAFDSIKQDSPT----GCRKVTNWIHLLDELEINLN -310

MGC      - LSISNKYELARHLQTTVILAEEDISANFLIYDYKKEYLKFFHKHY -358
MGA6     - LSVTNKYEVAVILHNTTVLKEEDITANYLFFDYKKSYLNFYKQEH -355

MGC      - RVIYDTFLTYVEDLFALDHQEFNEGIRNHLLYCLFITWENLFLKI -403
MGA6     - PHLYKAFVAGVEKLMRSEKEPISTELTNQLIYAFFITWENSFLKV -400

MGC      - NQSRRKLKLLVIERSYRNVGNFLKEYFGEFFEIISFDDLGNPKID -448
MGA6     - NQKDEKIRLLVIERSFNSVGNFLKKYIGEFFSITNFNELDALTID -445

MGC      - LLPLAKEYDVILTDVILEQNNHSEILFFNQMIPSVVADKLTDFLK -493
MGA6     - LEEIEKQYDVIVTDVMVGKSDELEIFFFYKMIPEAIIDKLNVFLN -490

MGC      - VKVGEDKLFESVALNQM------------------------STS -513
MGA6     - ISFAD-----SLPLDKPIKNPLDFHRKELTLPTPPNKLHAPPSTT -530
```

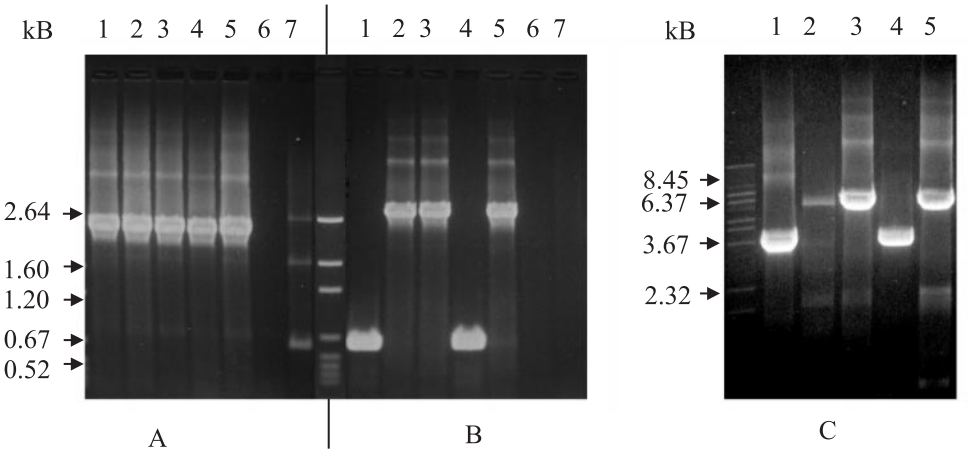

Fig. 3A–C Comparison by PCR of the *mgrC* locus in various GCS (lanes 1–2), GGS (lanes 3–5) and GAS strains (lanes 6 and 7). Lanes 1, 25287; lanes 2, H46A; lanes 3, 39072; lanes 4, 39086; lanes 5, 40104; lanes 6, M1 40/58; lanes 7, M3 4/55. **A** Total DNA from each strain was used for PCR with primers E and F to characterise the region between *nrd* and *emm*. **B** Primers G and H were used to amplify the region between *emm* and *rel*. The molecular weight standard was loaded in the central lane. **C** Primers I and J amplify the region located between *emm* and *skc*

the genomes of GCS and GGS (Fig. 5A, B). We termed this region *mgrC*, for *m*ultigene *r*egulon-like genomic segment in group *C* streptococci. The word "like" is included because the role of *mgc* as a regulator gene in GCS has yet to be demonstrated. The *mgrC* segment can be classified as the small type 1 (Fig. 5A) or the large type 2 (Fig. 5B). In Fig. 5 the two *mgrC* types are compared with two representative types of *mga* regulons in GAS. In the case of the GCS and GGS strains investigated here, *nrd* is always upstream of *mgc* and oppositely oriented (compare with PCR, Fig. 3A). In *mgrC* type 1 (prototype strain 25287) the genes immediately adjacent to *mgc* are *emm* and *rel*. (Fig. 5A). The small *mga* regulon in GAS also consists of three genes; however, here *scpA* borders the region at the 3′ end. In the larger *mgrC* type 2 (prototype strain H46A) *mgc* and *emm* are arranged as in type 1. But here *cpdB* is inserted between *emm* and *rel* and in opposite orientation (Fig. 5B). The other genes located downstream of *emm* are different in GCS and GAS. In both *mgrC* types, *rel* links the *mgrC* segment with the streptokinase region characterised recently by Mechold et al. (1993).

Discussion

The genomic region which flanks the *emm* gene in GCS and GGS has not been characterised previously. In this report we identify the genes that flank the *emm* gene in GCS and GGS strains. Sequence data for the GCS strains 25287 and H46A revealed the existence of a genome segment which shows partial similarity to, but also clear differences from, the *mga* regulon in GAS. We refer to this region as the *mgrC* region. In *mgrC* the two genes upstream of *emm* are the oppositely oriented ORF *nrd*, which encodes a protein with many regions of similarity to the ribonucleotide reductase (*nrdI*) of *L. lactis* (Jordan et al. 1997), and *mgc*, which is immediately adjacent to, and oriented in the same direction as, *emm*. The *mgc* sequence is about 61% homologous to various *mga* genes of GAS strains. In GCS strain 25287 *emm* is closely linked at the 3′ end to *rel*, which encodes a protein that functions in the synthesis and degradation of guanosine 3′,5′ bipyrophosphate (ppGpp; Metzger et al. 1989; Mechold et al. 1993). The *rel* gene flanks the *skc* region which itself comprises several genes: *orf1*, *skc* in opposite orientation, *lrp*, *abc* and *dexB*. This region has been characterized in detail by Mechold et al. (1993). The gene *skc* encodes streptokinase of GCS, *abc* encodes a protein belonging to the superfamily of ABC transporters (Higgins 1992), *dexB* encodes an alpha-glucosidase-like protein (Russel and Ferretti 1990), while *orf1* and *lrp* encode proteins of unknown function.

We also identified the gene *cpdB* in strain H46A. *cpdB* is located between *emm* and *rel* and its sequence shows homology to genes present in various microorganisms which encode 2′,3′ cyclo-nucleotide 2′-phosphodiesterases. The analysis of three additional GGS strains revealed that there exist two groups of GCS/GGS strains; one carries the gene *cpdB*, the other group does not. Because of the close linkage between the genes *mgc* and *emm* we postulate that there are two types of *mgrC* region in human isolates of GCS and GGS: type 1 lacks *cpdB* (prototype strain 25287) and in type 2 *cpdB* is inserted between *emm* and *rel* (prototype strain H46A). This is the first report of an *mgrC* region in GCS/GGS strains that is similar to the *mga* regulon of GAS. The crucial difference between *mgrC* and the *mga* regulon lies in the fact that they are located in completely different genomic environments. In GAS strains, three ORFs have been identified 5′ from *mga* and in the same orientation; these include *isp*, which encodes an immunogenic protein conserved in GAS (McIver et al. 1996). In the GCS and GGS strains tested here we identified, upstream of *mgc* and in opposite orientation, an *nrdI*-like gene (*nrd*) which shows no sequence homology to the *isp* region located 5′ to *mga* in GAS. At the 3′ end the tandemly arranged *mgc* and *emm* genes in GCS and GGS are linked to *rel*, which flanks the *skc* region. Moreover, in some GCS and GGS strains a *cpdB* gene is inserted between between *emm* and *rel*. This gene is not present in the *mga* regulon of GAS. In GAS the different types of *mga* regulons start with *mga*, which is followed

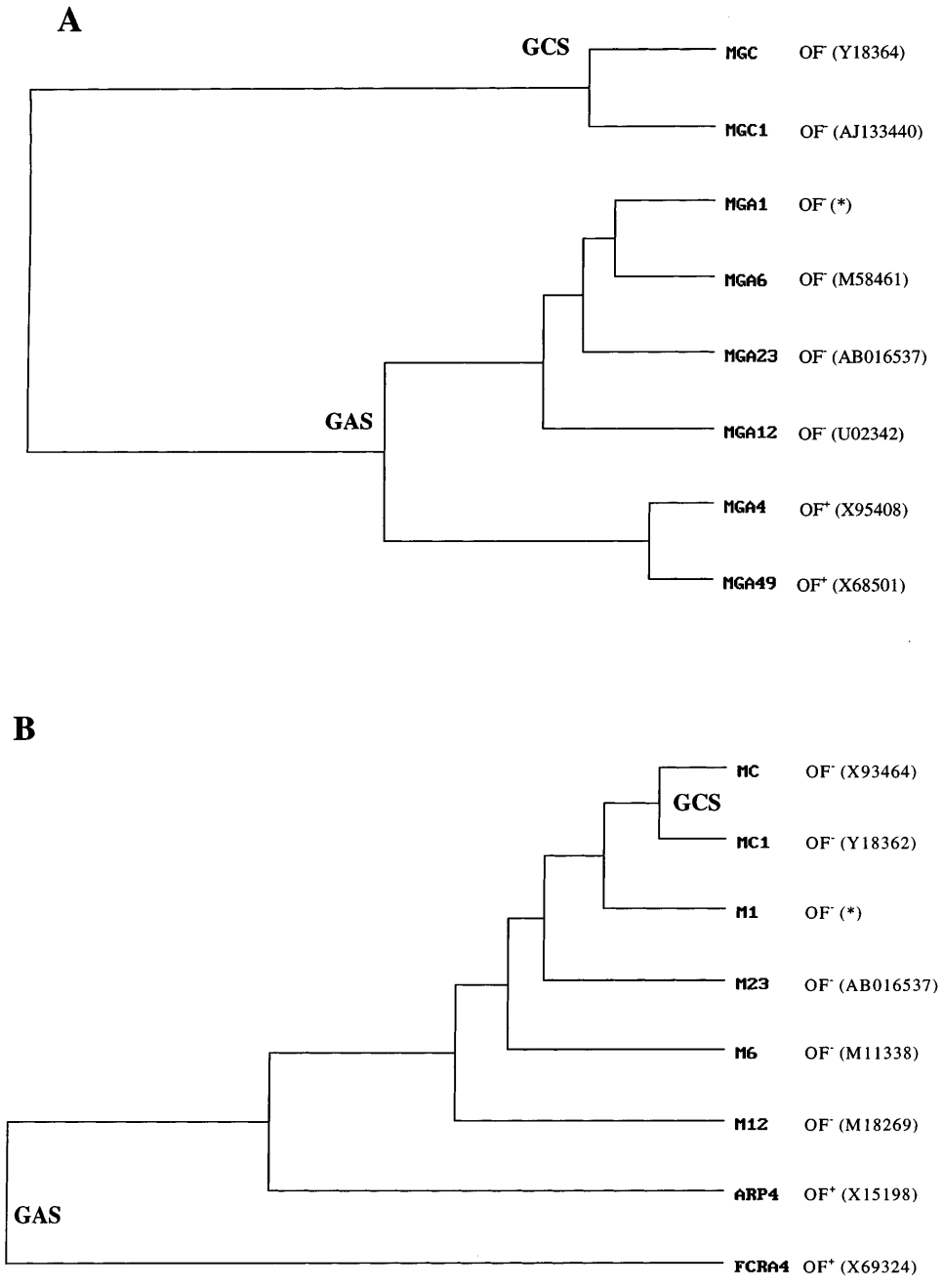

Fig. 4A Phylogenetic relationships between the amino acid sequences of Mga proteins from different M types of GAS strains and Mgc and Mgc1 of the GCS strains 25287 and H46A. **B** The M proteins of the corresponding strains are compared. The opacity factor status of each strain is given as OF^+ and OF^-. The Accession Nos. of the sequences from the NCBI database are given in *parentheses*. Mgc and Mgc1 form a distinct cluster, but the corresponding M proteins MC and MC1 are highly homologous to the M proteins of GAS. The sequence of Mga1 (*asterisk*) was taken from the database of the Streptococcal Genome Sequencing Project funded by USPHS/NIH Grant No. AI38406 (B. A. Roe, S. P. Linn, L. Song, X. Yuan, S. Clifton, M. McShan and J. Ferretti). Other sequences were taken from: Hollingshead et al. (1986); Robbins et al. (1987); Frithz et al. (1989); Podbielski (Accession No. X69324); Harbaugh et al. (1993); Geyer et al. (1999); Perez-Casal et al. (1991); Hong (Accession No. AB016537); Chen et al. (1993); and Podbielski et al. (1995)

by one or more *emm* or *emm*-like genes. Downstream these genes are flanked by *scpA* (Perez-Casal et al. 1991; Podbielski 1993; Podbielski et al. 1995; Katerov et al. 1994; Andersson et al. 1996). In GCS/GGS we find a similar tandem arrangement of *mgc* and *emm*, but no other *emm*-like genes were found in the *mgrC* region. The architecture of *mgrC* downstream of *emm* in GCS/GGS is also completely different from that of the *mga* regulon in GAS. In both *mgrC* types, *rel*, which flanks the *skc* region, is the last gene, while *scpA* terminates the *mga* regulon found in GAS. Moreover, in GAS, the ORFs that have been described 3′ to *scpA* – *orf1* and *orf2* (Podbielski et al. 1996) – show no homology to any of the genes in the *skc* region. In GAS the existence of a *skc* region with the same arrangement of genes as described for the GCS strain H46A has been reported (Frank et al. 1995). But in GAS the *skc* region seems not to be located in a genomic segment near the *mga* regulon – unlike the case in GCS/GGS, where it is closely linked to the *mgrC* region. Recent PCR experiments by Sriprakash and Hartas (1996), using GGS genomic DNA and primers based on published sequences from the *mga* regulon of a type M49 strain (Podbielski et al. 1996), did not reveal DNA fragments with sequence homology to *mga* genes of GAS. The apparent divergence (61% similarity) between the DNA sequences of *mgc* in GCS and *mga* in GAS and the complete lack of homology between the upstream regions of the two genes explain

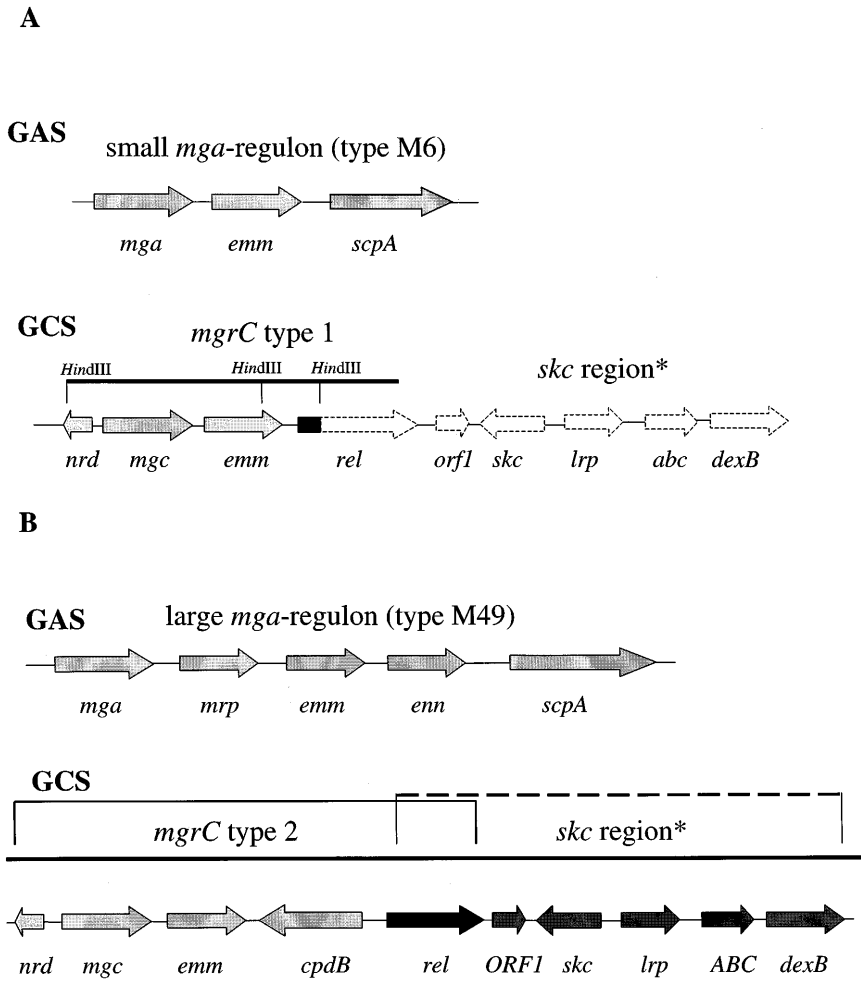

Fig. 5A, B Organisation of two types of *mgrC* regions in GCS and GGS deduced from sequence and PCR data. **A** The small *mga* regulon described for type 6 GAS (Perez-Casal et al. 1991) compared with the *mgrC* type 1 segment deduced from GCS strain 25287. *mgrC* type 1 is characterised by the tandem arrangement of *mgc*, *emm* and *rel*. The following genes in the streptokinase region are indicated. The open region in *rel* indicates that this gene was partially sequenced in strain 25287. **B** The arrangement of genes in a large *mga* regulon of GAS type 49 (Podbielski et al. 1995) and in the *mgrC* type 2 segment deduced from GCS strain H46A. In *mgrC* type 2 the gene *cpdB* is inserted in the opposite orientation between *emm* and *rel*. In both cases *rel* borders on the *skc* region (*asterisk*) sequenced by Mechold et al. (1993) in strain H46A. The whole *mgrC* type 2 segment up to *rel* of the same strain was sequenced in our laboratory (Accession No. AJ133440)

the failure of this experiment. Our approach, which involved inverse PCR starting from the *emm* gene of GCS strain 25287, has provided sequence information for the genes adjacent to *emm*.

It has been speculated that the specific arrangement characteristic of the *mga* regulon may be responsible for the increased virulence of GAS strains compared to other β-haemolytic streptococci (Podbielski et al. 1992). The expression of M and M-like proteins in GAS is positively regulated by *mga*, which encodes the protein Mga. In phagocytosis experiments, both in vivo and in vitro, the GCS strain H46A was found to be less virulent than a number of GAS strains (Schmidt et al. 1993b). However, in recent years, GCS as well as GGS strains have been isolated from patients suffering from serious infections like toxic-shock syndrome (Keiser and Campbell 1992; Miller et al. 1996; Natoli et al. 1996; Roth et al. 1999). Furthermore, we have identified the M protein of the GCS human isolate strain 25287 as a virulence factor for chicken embryos (Geyer et al. 1999). Thus, a regulatory function for *mgc* in the expression of M protein, and possibly of additional proteins, in group C and G streptococci can be assumed. To clarify whether *mgc* really functions as a positive regulator gene in GCS and GGS, like *mga* in GAS, detailed studies of regulation and expression are underway in our laboratory.

The phylogenetic analysis revealed that the Mgc proteins of strains 25287 and H46A clearly form a distinct cluster separated from the Mga proteins of GAS. It seems likely that *mgc* of GCS/GGS and *mga* of GAS have undergone a considerable period of independent evolutionary development, as underlined by the fact that they now show only 61% identity. In contrast, M proteins of GCS and GAS did not form distinct clusters, which makes horizontal transfer of *emm* genes between GAS and GCS/GGS a likely possibility (Simpson et al. 1992; Schnitzler et al. 1995; Sriprakash and Hartas 1996). A comparison between the 5′ intergenic sequences of *mgc* and *emm* in the GCS strain 25287 and the corresponding regions of *mga* and *emm* in GAS did not reveal any homology, not even in the putative promoter sequences. This also emphasizes that the *mgrC* region of GCS/GGS and the *mga* regulon of GAS have followed different evolutionary trajectories.

Acknowledgements We thank Ms. Roswitha John for excellent technical assistance. The work was supported by the grant Schm877/1-3 from the Deutsche Forschungsgemeinschaft (DFG) Germany. We thank Horst Malke (Jena) for his helpful review of this manuscript.

References

Åkesson P, Schmidt KH, Cooney J, Björck L (1994) M1 and protein H: IgGFc- and albumin-binding streptococcal surface proteins encoded by adjacent genes. Biochem J 300:877–886

Andersson G, McIver K, Heden LO, Scott JR (1996) Complementation of divergent *mga* genes in group A Streptococcus. Gene 175:77–81

Ben Nasr A, Wistedt A, Ringdahl U, Sjöbring U (1994) Streptokonase activates plasminogen bound to human group C and G streptococci through M-like proteins. Eur J Biochem 222:267–276

Bisno AL, Collins CM, Turner JC (1996) M proteins of group C streptococci isolated from patients with acute pharyngitis. J Clin Microbiol 34:2511–2515

Caparon MG, Scott JR (1991) Genetic manipulation of pathogenic streptococci. Methods Enzymol 204:556–586

Carlsson-Wistedt A, Ringdahl U, Müller-Esterl W, Sjöbring U (1995) Identification of a plasminogen-binding motif in PAM, a bacterial surface protein. Mol Microbiol 18:569–578

Chen C, Bormann N, Cleary PP (1993) VirR and Mry are homologous *trans*-acting regulators of M protein and C5a peptidase expression in group A streptococci. Mol Gen Genet 241:685–693

Christensen LR (1945) Streptococcal fibrinolysin: a proteolytic reaction due to a serum enzyme activated by streptococcal fibrinolysin. J Gen Physiol 28:363–383

Cleary PP, Peterson J, Chen C, Nelson C (1991) Virulent human strains of group-G streptococci express a C5a peptidase enzyme similar to that produced by group-A streptococci. Infect Immun 59:2305–2310

Efstratiou A, Teare EL, McGhie D, Colman G (1989) The presence of M proteins in outbreak strains of *Streptococcus equisimilis* T-type 204. J Infect 19:105–111

Fox K, Turner J, Fox A (1993) Role of beta-hemolytic group C streptococci in pharyngitis: incidence and biochemical characteristics of *Streptococcus equisimilis* and *Streptococcus anginosis* in patients and healthy controls. J Clin Microbiol 31:804–807

Frank C, Steiner K, Malke H (1995) Conservation of the organisation of the streptokinase region among pathogenic streptococci. Med Microbiol Immunol 184:139–146

Frithz E, Heden LO, Lindahl G (1989) Extensive sequence homology between IgA receptor and M proteins in *Streptococcus pyogenes*. Mol Microbiol 3:1111–1119

Geyer A, Roth A, Vettermann S, Günther E, Groh A, Straube E, Schmidt KH (1999) M protein of a *Streptococcus dysgalactiae* human wound isolate shows multiple binding to different plasma proteins and shares epitopes with keratin and human cartilage. FEMS Immunol Med Microbiol 26: 11–24

Haanes EJ, Heath DG, Cleary PP (1992) Architecture of the *vir* regulons of group A streptococci parallels opacity factor phenotype and M protein class. J Bacteriol 174:4967–4976

Harbaugh MP, Podbielski A, Hugl S, Cleary PP (1993) Nucleotide substitutions and small-scale insertion produce size and antigenic variation in group A streptococcal M1 protein. Mol Microbiol 8:981–991

Heden LO, Lindahl G (1993) Conserved and variable regions in protein Arp, the IgA receptor of *Streptococcus pyogenes*. J Gen Microbiol 139:2067–2074

Higgins CF (1992) ABC transporters: from microorganisms to man. Annu Rev Cell Biol 8:67–113

Hollingshead SK, Fischetti VF, Scott JR (1986) Complete nucleotide sequence of type 6 M protein of the group A streptococcus: repetitive structure and membrane anchor. J Biol Chem 261:1677–1686

Jordan A, Aslund F, Pontis E, Reichard P, Holmgreen A (1997) Characterisation of *Escherichia coli* NrdH. A glutaredoxin-like protein with a thioredoxin-like activity profile. J Biol Chem 272:18044–18050

Katerov V, Schalen C, Totolian AA (1994) Sequencing of genes within the *vir* regulon of *Streptococcus pyogenes* type M15 – an opacity factor-positive serotype with low opacity factor expression. Mol Gen Genet 245:78–85

Keiser P, Campbell W (1992) "Toxic strep syndrome" associated with group C streptococcus. Arch Intern Med 152:882–884

McIver KS, Subbarao S, Kellner EM, Heath AS, Scot JR (1996) Identification of *isp*, a locus encoding an immunogenic secreted protein conserved among group A streptococci. Infect Immun 64:2548–2555

Mechold U, Steiner K, Vettermann S, Malke H (1993) Genetic organisation of the streptokinase region of the *Streptococcus equisimilis* H46A chromosome. Mol Gen Genet 241:129–140

Metzger S, Sarubbi E, Glaser G, Cashel M (1989) Protein sequences encoded by the *relA* and the *spoT* genes of *Escherichia coli* are interrelated. J Biol Chem 264:9122–9125

Miller CW, Prescott JF, Mathews KA, Betschel SD, Yager JA, Guru V, DeWinter L, Low DE (1996) Streptococcal toxic shock syndrome in dogs. J Am Vet Med Assoc 209:1421–1426

Natoli S, Fimiani C, Faglieri N, Laurenzi L, Calamaro A, Frasca AM, Arcuri E (1996) Toxic shock syndrome due to group C streptococci. A case report. Intensive Care Med 22:985–989

Perez-Casal J, Caparon MG, Scott JR (1991) Mry, a trans-acting positive regulator of the M protein gene of *Streptococcus pyogenes* with similarity to the receptor proteins of two-component regulatory systems. J Bacteriol 173:2617–2624

Podbielski A (1993) Three different types of organisation of the *vir*-regulon in group A streptococci. Mol Gen Genet 237: 287–330

Podbielski A, Kaufhold A, Lütticken R (1992) The *vir*-regulon of *Streptococcus pyogenes*: Coordinated expression of important virulence factors. Immunität und Infektion 20:161–168

Podbielski A, Flosdorff A, Weberheynemann J (1995) The group A streptococcal *virR49* gene controls expression of four structural *vir* regulon genes. Infect Immun 63:9–20

Podbielski A, Woischnik M, Pohl B, Schmidt KH (1996) What is the size of the group A streptococcal *vir* regulon? The Mga regulator affects expression of secreted and surface virulence factors. Med Microbiol Immunol 185:171–181

Robbins JC, Spanier JG, Jones SJ, Simpson WJ, Cleary PP (1987) *Streptococcus pyogenes* type 12 M protein gene regulation by upstream sequences. J Bacteriol 169:5633–5640

Roth S, Andrassy K, Schmidt KH, Günther E, Ritz E (1999) Febrile lady with acute renal failure and desquamatic erythema. Amer J Kidney Diseases 34:1–6

Russel RRB, Ferretti JJ (1990) Nucleotide sequence of the dextran glucosidase (*dexB*) gene of *Streptococcus mutans*. J Gen Microbiol 136:803–810

Sambrook J, Fritsch EF, Maniatis TE (1989) Molecular cloning: a laboratory manual (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schmidt KH, Wadström T (1990) A secreted receptor related to M protein of *Streptococcus pyogenes* binds to fibrinogen, IgG and albumin. Zentralbl Bacteriol 273:216–228

Schmidt KH, Mann K, Cooney J, Köhler W (1993a) Multiple binding of type 3 streptococcal M protein to human fibrinogen, albumin and fibronectin. FEMS Immunol Med Microbiol 7:135–144

Schmidt KH, Wiesner J, Gerlach D, Reichardt W, Ozegowski JH, Köhler W (1993b) Susceptibility of chicken embryos to group A streptococci: Correlation with fibrinogen binding. FEMS Immunol Med Microbiol 7:231–240

Schmidt KH, Gerlach D, Wollweber L, Reichardt W, Mann KH, Ozegowski JH, Fleischer B (1995) Mitogenenicity of M5 protein extracted from *Streptococcus pyogenes* cells is due to streptococcal pyrogenic exotoxin C and mitogenic factor MF. Infect Immun 63:4569–4575

Schnitzler N, Podbielski A, Baumgarten G, Mignon M, Kaufhold A (1995) M or M-like protein gene polymorphism in human group G streptococci. J Clin Microbiol 33:356–363

Scott S, Cleary P, Caparon MG, Kehoe M, Heden L, Musser JM, Hollingshead S, Podbielski A (1995) New name for the positive regulator of the M protein of group A streptococcus. Mol Microbiol 17:779

Simpson WJ, Musser JM, Cleary PP (1992) Evidence consistent with horizontal transfer of the gene (*emm12*) encoding serotype M12 protein between group A and group G pathogenic streptococci. Infect Immun 60:1890–1893

Sriprakash KS, Hartas J (1996) Lateral genetic transfers between group A and G streptococci for M-like genes are ongoing. Microb Pathog 20:275–285

Triglia T, Peterson GM, Kemp DJ (1988) A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res 16:8186

Turner JC, Fox A, Fox K, Addy C, Garrison CZ, Herron B, Brunson C, Betcher G (1993) Role of group C beta-hemolytic streptococci in pharyngitis: epidemiologic study of clinical features associated with isolation of group C streptococci. J Clin Microbiol 31:808–811

Vandamme P, Pot B, Falsen E, Kersters K, Devriese LA (1996) Taxonomic study of Lancefield streptococcal groups C, G and L (*Streptococcus dysgalactiae*) and proposal of *S. dysgalactiae* subsp. *equisimilis* subsp. nov. Int J Syst Bacteriol 46:774–781