

NOTES

Genomic Organization and Molecular Characterization of SM1, a Temperate Bacteriophage of *Streptococcus mitis*

Ian R. Siboo, Barbara A. Bensing, and Paul M. Sullam*

Veterans Affairs Medical Center and University of California, San Francisco, California 94121

Received 9 June 2003/Accepted 7 September 2003

The direct binding of *Streptococcus mitis* to human platelets is mediated in part by two proteins (PblA and PblB) encoded by a lysogenic bacteriophage (SM1). Since SM1 is the first prophage of *S. mitis* that has been identified and because of the possible role of these phage-encoded proteins in virulence, we sought to characterize SM1 in greater detail. Sequencing of the SM1 genome revealed that it consisted of 34,692 bp, with an overall G+C content of 39 mol%. Fifty-six genes encoding proteins of 40 or more amino acids were identified. The genes of SM1 appear to be arranged in a modular, life cycle-specific organization. BLAST analysis also revealed that the proteins of SM1 have homologies to proteins from a wide variety of lambdoid phages. Bioinformatic analyses, in addition to N-terminal sequencing of the proteins, led to the assignment of possible functions to a number of proteins, including the integrase, the terminase, and two major structural proteins. Examination of the phage structural components indicates that the phage head may assemble using stable multimers of the major capsid protein, in a process similar to that of phage r1t. These findings indicate that SM1 may be part of a discrete subfamily of the *Siphoviridae* that includes at least phages r1t of *Lactococcus lactis* and SF370.3 of *Streptococcus pyogenes*.

Virulence factors produced by pathogenic bacteria are often encoded within lysogenic bacteriophages. Cytotoxic proteins, such as SbeA of *Staphylococcus aureus* or the shiga toxins of *Shigella dysenteriae* and *Escherichia coli*, were some of the first such factors identified (31). Recent research has shown that, in addition to toxins, a number of other virulence determinants are also encoded within phages. These include proteins linked to colonization and invasion, as well as proteins that mediate resistance to phagocytosis (6, 31). Thus, bacteriophages are able to impart to their host a wide array of genes that encode various virulence effectors. Moreover, the presence of these genes on mobile genetic elements could provide a mechanism for the dissemination of these virulence determinants to other strains or species.

Streptococcus mitis is a leading cause of infective endocarditis (16). Despite the prominent role of *S. mitis* in endocarditis, few virulence determinants of this organism have been identified. Our laboratory's previous studies have examined the direct binding of human platelets by *S. mitis* (4, 28, 29). This interaction is thought to be important both for the initiation of endocardial infection and for the subsequent development of endocardial vegetations. Our investigators recently identified two cell wall-associated proteins (PblA and PblB) of *S. mitis* that contribute to platelet binding by this organism (4). Of note, these two proteins are encoded by an inducible prophage (SM1) that is present in the genome of *S. mitis* strain SF100 (5). Although the precise mechanism by which PblA and PblB

enhance binding has not been defined, the localization of these proteins suggests that they may mediate the direct attachment of *S. mitis* to the platelet surface. Thus, PblA and PblB appear to be highly unusual examples of phage-encoded virulence factors that serve as adhesins for human tissue.

In our previous work, we showed that the mature SM1 phage particle was composed of an isometric, icosahedral head and a long noncontractile tail (5). The phage genome was estimated to be approximately 35 kb in length. Sequence analysis of a portion of the genome revealed that the putative virulence genes *pblA* and *pblB* reside in a region that encodes morphogenetic proteins. Moreover, Western blotting studies demonstrated that, in addition to being localized to the bacterial cell wall, alternate forms of PblA and PblB were components of the phage particle.

Since no other phages of *S. mitis* have been reported to date and because PblA and PblB are rare examples of phage structural proteins that are associated with bacterial adherence, we sought to examine SM1 in greater detail. Given the association between the Pbl proteins and platelet binding, we also sought to identify any other potential virulence factors encoded by the phage. For these reasons, we have completed the sequencing of the SM1 genome and have begun to assign functions to the various open reading frames (ORFs) encoded therein.

Phage DNA was extracted from purified phage particles and sequenced to sixfold coverage from shotgun subclone libraries of the SM1 genome (Genome Therapeutics, Waltham, Mass.). Closure of the gaps between contigs was accomplished by sequencing gap-specific PCR products generated using the SM1 genome as a template. The full phage genome was found to consist of 34,692 bp and has a G+C content of 39 mol%. The genomic sequence was analyzed using version 2.1 of Gene

* Corresponding author. Mailing address: Division of Infectious Diseases, VA Medical Center (111W), 4150 Clement St., San Francisco, CA 94121. Phone: (415) 221-4810, ext. 2550. Fax: (415) 750-0502. E-mail: sullam@itsa.ucsf.edu.

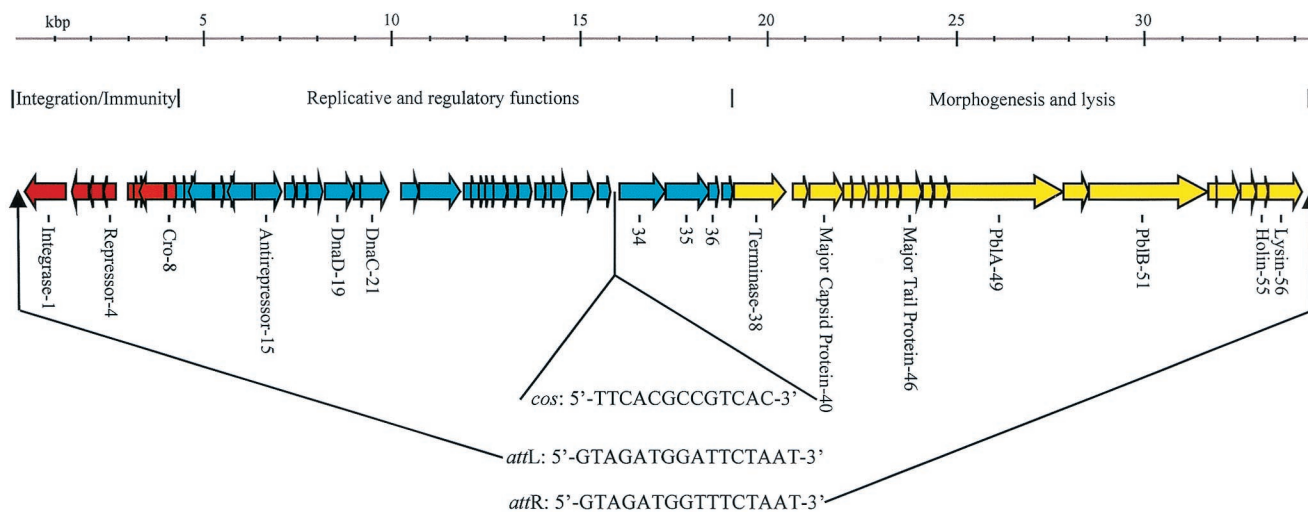


FIG. 1. Map of *S. mitis* phage SM1. Modular regions of the genome encoding distinct functions are indicated above the map. Genes encoding proteins identified by BLAST identity or N-terminal sequencing are labeled below the map. The phage *att* and *cos* sites are indicated along with their respective positions on the genomic map.

MarkHMM for prokaryotes (http://opal.biology.gatech.edu/GeneMark/gmhmm2_prok.cgi) (21) to define potential coding regions. The algorithm identified 56 ORFs that encode proteins of 40 or more amino acids (Fig. 1). These ORFs are grouped into three discrete modular regions that reflect the phage life cycle: (i) integration and immunity, (ii) replicative and early lytic functions, and (iii) morphogenesis and lysis. The integration/immunity module, comprised of eight genes, is located immediately downstream of the left-end *att* site (Fig. 1).

Of the eight genes, only three (genes 1, 4, and 8) could be ascribed a function based on similarity to proteins of known function (PSI-BLAST [2]; <http://www.ncbi.nlm.nih.gov/BLAST/>). The protein encoded by gene 1 (gp1) is 79% identical to the integrase of the *Streptococcus pneumoniae* phage MM1 and is therefore likely to have a similar role in SM1. The product of gene 4 shows high similarity to several repressor proteins, having highest identity (55%) to the repressor of phage MM1. Motif analysis (<http://motif.genome.ad.jp/>) of gp4 revealed a DNA-binding helix-turn-helix motif in the N terminus of the protein, which is also characteristic of phage repressors. The position of gene 4 in the SM1 genome is typical of many other genes encoding phage repressors (19). Specifically, it is the last gene in a group of genes immediately downstream of the left-end *att* site that is transcribed from right to left (Fig. 1). Downstream of the putative repressor is a noncoding region that may contain the promoter-operator binding sites mediating the phage lytic-lysogenic switch. This overall organization is characteristic of several other *Siphoviridae* from low-GC-content gram-positive bacteria, including phages MM1 and r1t (22, 23). Thus, the regulatory region of SM1 conserves much of the genetic organization present within this phage family.

The immunity regions of many phages of low-GC-content gram-positive bacteria contain a Cro-like protein encoded divergently, but immediately upstream, from the repressor (19). However, SM1 appears to have a Cro-like protein encoded by gene 8, in a position that separates it from the putative promoter-operator region by three genes (genes 5 to 7). Although

it is currently not known whether this separation affects the functional regulation of lysis and lysogeny, it does represent an unusual organizational feature of the SM1 genome.

The region of the phage genome extending from the end of the integration/immunity module toward the terminase gene defines the module related to replication and early lytic functions. Proteins encoded within this region show varying degrees of identity to proteins from numerous other phages, including *Streptococcus thermophilus*, *S. pneumoniae*, *Streptococcus pyogenes*, *Streptococcus agalactiae*, and *Lactococcus lactis*. Although it was not possible to assign probable functions to many of the proteins in this region, due to a lack of similarity to proteins of known function, motif and PSI-BLAST searches did reveal possible functions for a number of gene products. gp15, encoded by one of the most 5'-proximal genes in the region, has a high percentage of identity to antirepressors, including that of phage LambdaSa2 of *S. agalactiae* 2603V/R (77%) and a phage of *S. pyogenes* MGAS8232 (72%). These antirepressors function in other phages to relieve the inhibition mediated by the cI-like repressor proteins by binding directly to these repressors subsequent to initiation of the SOS response upon DNA damage (10, 26, 27). Therefore, gp15 may facilitate the entry of SM1 into a lytic phase of development upon induction by DNA-damaging agents.

The products of genes 19 and 21 are homologues of DNA replication proteins from a wide range of species. Analyses using the ProDom (8) and Pfam (3) algorithms indicated that gp19 and gp21 contain DnaD and DnaC motifs, respectively. This suggests that gp19 and gp21 function in genome replication. DnaD has been shown to mediate interaction with the replication initiation protein DnaA, with DnaC, and also with DnaD (7, 17). Therefore, gp19 may mediate primosome assembly and replication initiation in a manner similar to DnaD. The ATP-binding DnaC protein has been demonstrated to be an energy-dependent helicase (9). Thus, the gp21 orthologue may serve to unwind the phage genome during replication.

Gene 34 appears to encode a phage structural component

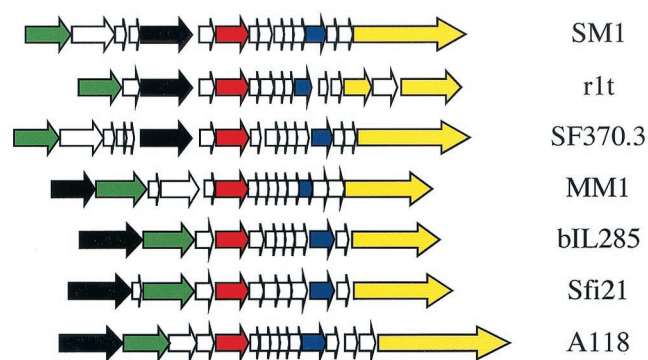


FIG. 2. A schematic representation of the order of structural genes proximal to the major capsid protein in SM1 and other phages of gram-positive hosts. The position of the genes encoding a possible portal protein demonstrate the distinctive organization of the head cluster locus of SM1, r1t, and SF370.3 relative to that of other low-GC-content phages. Colored arrows represent gene homologues. Red = major capsid protein; black = terminase; green = portal; blue = major tail protein; yellow = tape measure protein.

having homology to a protein from at least two low-GC-content phages. Of interest, both a lactococcal phage (r1t) and a related, uninducible phage of *S. pyogenes* (SF370.3) contain proteins homologous to gp34 that are in similar positions within their respective genomes (11, 30) (Fig. 2). Specifically, the SM1, r1t, and SF370.3 appear to have a potential portal protein encoded by a gene located upstream of the gene encoding the terminase, while the other phages in this family maintain their head genes as a cluster located downstream of the terminase. Therefore, although the location of gene 35 and its homologues is unusual relative to most low-GC-content phages, it may reflect an organization that is characteristic of a subfamily of related phages.

The products of genes 35 and 36 have no known function, but of note, both have regions of homology to putative neisserial proteins. The N-terminal region of gp35 (amino acids 1 to 249) shows 56% identity to a phage protein from *S. pyogenes* strain 8232 designated as a conserved phage protein, while the C-terminal region (amino acids 297 to 401) has 46% identity to an uncharacterized, putative adhesin of *Neisseria gonorrhoeae*, MafB. Thus, gp35 appears to be a chimeric protein. Beyond its designation as an adhesin, no other information regarding MafB has been published. The product of gene 36 has no known function but is 79% identical to the hypothetical protein NMA0318 of *Neisseria meningitidis* Z249.

The third modular region of the phage includes genes encoding the morphogenetic and lytic proteins. gp38 shows greatest identity (79%) to the terminase of phage r1t. A number of putative structural proteins are encoded downstream of gene 38. PSI-BLAST searches indicated that genes 40 and 46 encode the major head and tail proteins of the phage, respectively. Other components of the phage include PblA and PblB (gp49 and gp51), which may function in the phage as putative tail tape measure and tail fiber/host recognition proteins, respectively. The protein encoded by gene 55 shows high identity to the holin of the *S. pneumoniae* phage VO1 (84%). The putative phage lysin, a homologue of the *N*-acetylmuramoyl-L-alanine amidase of the *S. pneumoniae* phage Dp-1 (72%), is

encoded by gene 56. Most of the remaining genes (gp41 to -45, -47, -48, -50, and -52 to -54) show similarity to phage proteins that have no known function.

Interestingly, phages SM1, r1t, and SF370.3 appear to have specific genes in distinct positions relative to many other low-GC-content phages. The products of genes 38 through 49 of SM1 are homologous with, and are arranged similarly to, that of gp29 through -40 of phage r1t and gp26 through -37 of SF370.3. In addition, the position of the putative structural protein encoded by gene 35 outside of the morphogenetic module is conserved among SM1, r1t, and SF370.3 (Fig. 2). Thus, the arrangement of these genes, as well as the gp34 homologues discussed above, may be indicative of a discrete subfamily.

The phage *att* and *cos* sites are two distinct DNA domains that define the ends of the integrated and mature phage genome, respectively. To identify the *att* sites (*attL* and *attR*) of the integrated phage, inverse PCR (4) was utilized, using primers designed to read out from the phage sequence and into the genome of the lysogen. The junction regions between the integrated phage genome and host chromosomal sequences contained a repeated, 16-bp sequence (5'-GTAGATGGATTCT AAT-3' [*attL*] and 5'-GTAGATGGTTTCTAAT-3' [*attR*]) representing the sequences duplicated as a result of the recombination between *attP* and *attB*. Examination of the sequence of the phage genome revealed that the phage attachment sequence (*attP*) was identical to that of *attL*. Furthermore, sequencing of the region of SF100 chromosomal DNA from which the phage had excised revealed that *attB* was identical to *attR*.

We had reported previously that an 18-kb *EcoRI* restriction fragment of the SM1 genome could be resolved into 15-kb and 3-kb fragments by heating the products of the restriction digest prior to separation by agarose electrophoresis. This indicated that the *cos* region is present within the 18-kb *EcoRI* restriction fragment. When we compared this and various other restriction patterns predicted from the phage sequence data, we were able to narrow the position of the *cos* site to an intergenic region between genes 33 and 34. Primers that read outward from the genes and toward the *cos* site were used for sequencing from the SM1 genome. Comparison of the sequence obtained with the whole genome sequence revealed that SM1 has a *cos* site with a 13-nucleotide 3' overhang (Fig. 3).

Bioinformatic analysis of the proteins encoded within the SM1 genome indicated the presence of several potential structural proteins. In addition to minor structural components, it was predicted that the products of genes 40 and 46 were the major head and tail proteins, respectively. To identify proteins that make up the phage particle, protein components of SM1 were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and stained with silver (Fig. 4). Wild-type phage (SM1) and a tailless derivative from which the genes encoding PblA and PblB had been deleted (SM1 Δ AB) were isolated as described previously (5). The protein components were separated by SDS-PAGE on 4-to-15% gradient Ready Gels (Bio-Rad) and either stained with silver to visualize all proteins within the purified preparations or used to isolate proteins for N-terminal sequencing. Three major and several minor phage structural proteins were apparent on the stained gels (Fig. 4). N-terminal sequence analysis of the two

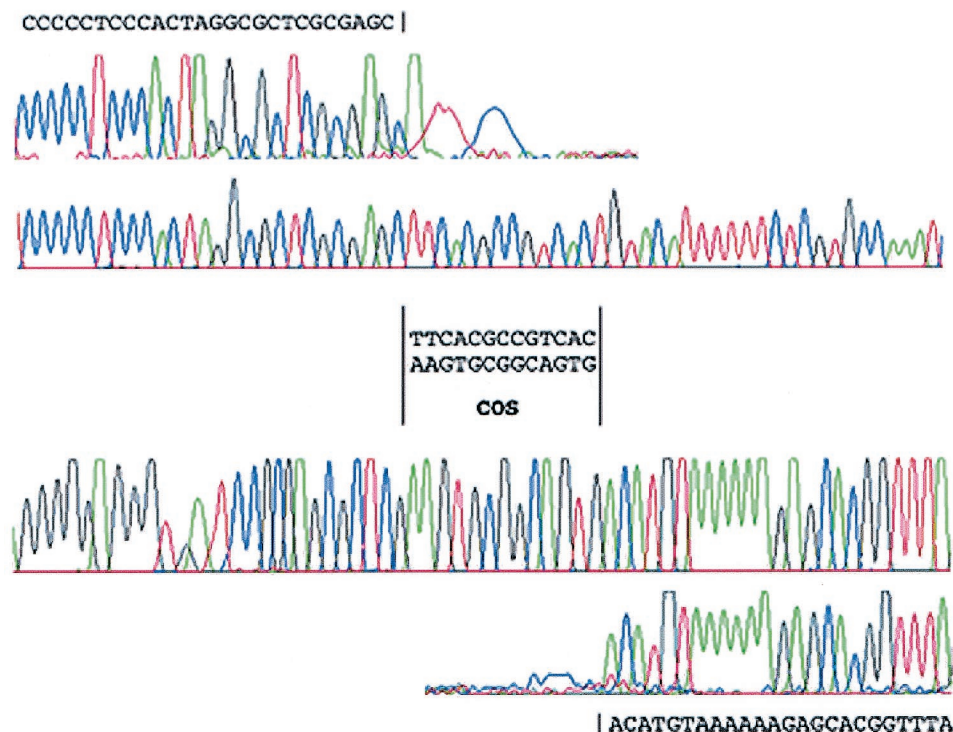


FIG. 3. Sequencing data for the SM1 *cos* site. The chromatograms generated from sequencing of the two ends of the SM1 *cos* site are aligned above and below the traces derived from sequencing of the corresponding region of the integrated phage.

largest major proteins (apparent masses of approximately 190 and 160 kDa [NH₂-XEAQLSKGNL-COOH]) indicated that both are encoded by gene 40. This gene is predicted to encode a 32-kDa protein (gp40). Thus, these large structural proteins may represent stable hexameric and pentameric forms, respectively, of gp40. Furthermore, these sequences revealed that the major head protein had not undergone proteolytic processing. Sequence analysis of the smallest of the three major proteins (NH₂-ATEANVTTAK-COOH) indicated that it is encoded by gene 46. The apparent molecular mass of this protein (28 kDa) is somewhat higher than predicted (19 kDa) for this gene product.

To confirm the function of these proteins, we compared by SDS-PAGE the structural components of SM1 and SM1ΔAB (Fig. 4). The 190- and 160-kDa proteins described above were present in both SM1 and SM1ΔAB. However, the protein identified as gp46, in addition to four other proteins that have yet to be characterized (Fig. 4), were absent from SM1ΔAB. These findings corroborate the bioinformatic data indicating that gp46 is the major tail protein and that gp40 is the SM1 major capsid protein.

Complete sequencing and extensive bioinformatic analyses have revealed several interesting characteristics of the SM1 genome. As has been described for other members of the *Siphoviridae*, the ORFs of SM1 appear to be arranged in discrete modules that reflect the phage life cycle. However, we did find some notable differences in gene arrangements between SM1 and typical members of the phage family. In λ, the integrase is separated from the repressor by 14 genes. However, like many phages of low-GC-content gram-positive bacteria,

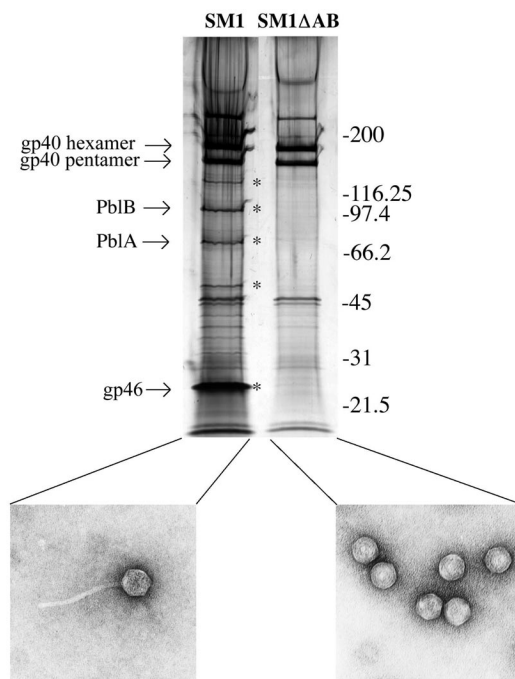


FIG. 4. Profile of the SM1 structural proteins and electron micrographs of phage particles. The protein components of SM1 and SM1ΔAB were separated by SDS-PAGE and stained with silver as described in the text. The phage isolate is indicated above each lane, while an electron micrograph of the material present in the sample is shown below the gel. Proteins marked with asterisks represent components of the tail. Molecular masses, in kilodaltons, are indicated to the right of the gel.

TABLE 1. Regions of homology in the SM1 genome

ORF	Start	Stop	MW	pI	Putative function	Best match ^{a,b}	Other BLAST homologies ^b
1	1230	103	43.6	9.4	Integrase	Integrase, <i>S. pneumoniae</i> phage MM1, CAB96616 (79%)	<i>L. lactis</i> phage bIL312, NP_076801 (38%); <i>S. pyogenes</i> M1 GAS, NP_269567 (35%); <i>S. agalactiae</i> 2603V/R phage LambdaSa1, NP_687574 (37%)
2	1791	1351	16.5	10.7		<i>L. lactis</i> phage TPW22, AAF12708 (72%)	<i>Clostridium acetobutylicum</i> , NP_348535 (53%)
3	2184	1804	15.0	5.6		<i>S. pyogenes</i> MGAS315, NP_665260 (45%)	<i>S. pyogenes</i> phage ϕ NIH1.1, NP_438116 (44%); <i>S. agalactiae</i> 2603V/R, NP_688872 (44%); <i>S. pneumoniae</i> phage MM1, NP_150135 (38%)
4	2541	2188	13.3	4.9	Repressor	Repressor, <i>S. pneumoniae</i> MM1 NP_150136 (55%)	<i>S. pyogenes</i> MGAS315, NP_664488 (50%); <i>S. pyogenes</i> MGAS8232, NP_606594 (50%); <i>S. agalactiae</i> 2603V/R phage LambdaSa2, NP_688871 (50%); <i>Listeria innocua</i> phage A118, NP_471937 (46%); <i>S. pyogenes</i> phage ϕ NIH1.1, NP_438117 (44%); <i>Listeria monocytogenes</i> phage 2389, NP_511010 (44%); <i>Enterococcus faecalis</i> V583, NP_816474 (44%)
5	2829	2957	4.8	6.2		—	—
6	2972	3184	8.2	5.1		—	—
7	3840	3160	26.1	6.3		—	—
8	3895	4104	7.8	8.7	Repressor	Cro homolog <i>S. agalactiae</i> 2603V/R, NP_688867 (80%)	<i>S. pyogenes</i> phage 315.1, NP_664490 (50%); <i>S. pyogenes</i> MGAS8232, NP_606596 (50%)
9	4108	4305	7.3	5.1		—	—
10	4316	4468	5.7	4.5		—	—
11	5081	4422	25.3	5.7		<i>S. pyogenes</i> SSI-1, NP_802143 (60%).	<i>S. pyogenes</i> MGAS315, NP_664778 (60%)
12	5138	5368	8.8	6.4		—	—
13	5390	5518	5.0	5.0		—	—
14	6158	5481	26.1	5.4		<i>S. pyogenes</i> MGAS8232, NP_607411 (37%)	—
15	6213	6932	27.2	5.9	Antirepressor	Antirepressor protein, <i>S. agalactiae</i> 2603V/R phage LambdaSa2, NP_688866 (77%)	<i>S. pyogenes</i> MGAS8232, AAL97097 (72%); <i>N. gonorrhoeae</i> , CAC28354 100% identity over the first 128 amino acids
16	7015	7287	10.6	9.4		<i>S. pneumoniae</i> phage MM1, NP_150139 (82%)	—
17	7337	7594	10.1	5.4		<i>S. thermophilus</i> phage Sfi21, NP_597801 (41%)	—
18	7610	8038	16.6	4.8		—	—
19	8058	8855	30.6	6.0	Replication initiation	DnaD homolog, <i>C. acetobutylicum</i> , NP_348556 (60%)	<i>S. aureus</i> phage phi 11, NP_803268 (49%); Orf 20 <i>S. aureus</i> phage ϕ PV83, NP_061610 (45%)
20	8843	9001	5.8	5.1		—	—
21	8995	9765	29.8	5.8	Helicase	DnaC homolog orf 5, <i>S. thermophilus</i> phage 7201, NP_038305 (43%)	Orf 12 <i>L. lactis</i> phage rlt, NP_695040 (36%); DnaC <i>L. lactis</i> phage bIL309, NP_076710 (33%); DnaC <i>L. lactis</i> phage bIL286, NP_076651 (33%)
22	10112	10585	18.0	5.6		<i>E. faecalis</i> V583, NP_817053 (47%)	lin1252, <i>L. innocua</i> , NP_470589 (46%); <i>S. pyogenes</i> phage 315.3, NP_664935 (43%)
23	10585	11943	54.3	9.7		lin1253, <i>L. innocua</i> , NP_470590 (41%)	<i>S. pyogenes</i> phage 315.3, NP_664934 (40%); <i>E. faecalis</i> V583, NP_817052 (37%)
24	11963	12289	12.3	4.4		<i>S. pneumoniae</i> phage MM1, NP_150152 (80%)	—
25	12323	12583	10.2	4.2		—	—
26	12580	12921	13.4	8.7	Potential HNH nuclease	<i>Microbulbifer degradans</i> 2-40, ZP_00066196 (50%)	—
27	12918	13178	10.0	9.4		<i>S. thermophilus</i> phage Sfi21, NP_597808 (37%)	<i>S. agalactiae</i> 2603V/R, NP_688856 (37%); <i>S. thermophilus</i> phage DT1, NP_049431 (36%)
28	13175	13591	16.0	9.7		—	—
29	13674	13925	9.7	9.2		<i>S. pneumoniae</i> phage MM1, NP_758428 (73%)	<i>S. agalactiae</i> 2603V/R, NP_688855 (69%); <i>S. pyogenes</i> phage MGAS8232, NP_607814 (60%); <i>S. pyogenes</i> phage ϕ NIH1.1, NP_438130 (60%)
30	13903	14067	5.8	6.1		—	—

Continued on following page

TABLE 1—Continued

ORF	Start	Stop	MW	pI	Putative function	Best match ^{a,b}	Other BLAST homologies ^b
31	14087	14512	16.7	5.8		<i>S. pyogenes</i> phage 315.5, NP_665134 (52%)	<i>S. pyogenes</i> phage φNIH1.1, NP_438135 (49%); <i>S. pyogenes</i> phage 315.2, NP_664754 (49%); <i>S. thermophilus</i> phage O1205, NP_695102 (43%)
32	14625	15191	21.8	9.5	Integrase/recombinase	Recombinase <i>S. pyogenes</i> phage 315.1, NP_664508 (53%)	<i>S. pyogenes</i> phage MGAS8232, NP_606615 (52%); site-specific recombinase, <i>S. agalactiae</i> 2603V/R phage LambdaSa2, NP_688849 (45%)
33	15043	15747	13.1	10.0	Potential HNH nuclease	<i>S. pyogenes</i> phage MGAS8232, NP_607557 (68%)	Orf 3, <i>L. lactis</i> phage φ31, AAC39305 (68%); Orf 26, <i>L. lactis</i> phage rlt, NP_695054 (68%); <i>S. pyogenes</i> phage 315.3, NP_664927 (67%)
34	15988	17220	45.0	5.5		ORF5, <i>L. lactis</i> phage φ31, AAC39307 (71%)	Structural protein, <i>L. lactis</i> phage rlt, NP_695055 (70%); <i>S. agalactiae</i> 2603V/R, NP_687610 (64%); putative structural protein, <i>S. pyogenes</i> MGAS8232, NP_607555 (63%)
35	17210	1842	46.1	6.7		<i>S. agalactiae</i> 2603V/R, NP_687611 (66%)	<i>S. pyogenes</i> MGAS8232, NP_607554 (52%); <i>S. pyogenes</i> SSI-1, NP_802006 (52%); Orf 28 <i>L. lactis</i> phage rlt, NP_695056 (48%); <i>N. gonorrhoeae</i> , AAD31039 (43%)
36	18411	18590	6.5	4.6		<i>N. meningitidis</i> Z2491, NP_283152 (74%)	—
37	18753	19004	9.5	9.2		<i>E. faecalis</i> V583, NP_815180 (51%)	Lin0103, <i>L. innocua</i> , NP_469449 (50%); <i>L. lactis</i> phage ul36, NP_663675 (40%)
38	19068	20447	51.7	6.0	Terminase	Terminase (Orf 29), <i>L. lactis</i> phage rlt, NP_695057 (73%)	<i>S. pyogenes</i> MGAS8232, NP_607550 (70%); putative terminase <i>S. pyogenes</i> phage 315.3, NP_664919 (70%); <i>S. agalactiae</i> 2603V/R, NP_687612 (68%)
39	20597	21031	16.1	5.0		<i>S. agalactiae</i> 2603V/R, NP_687613 (51%)	<i>S. pyogenes</i> MGAS315, NP_664918 (50%); Orf 30, <i>L. lactis</i> phage rlt, NP_695058 (43%)
40	21044	21946	32.8	5.2	Major capsid Protein	Putative structural protein, phage-associated <i>S. pyogenes</i> M1 GAS NP_269540 (62%)	Putative structural protein, phage-associated <i>S. pyogenes</i> MGAS8232, NP_607548 (62%); structural protein <i>L. lactis</i> phage rlt, NP_695059 (55%); <i>S. agalactiae</i> 2603V/R phage LambdaSa1, NP_687614 (62%)
41	21949	22170	8.0	4.3		<i>S. pyogenes</i> , NP_269539 (38%)	Orf 32, <i>L. lactis</i> phage rlt, NP_695060 (33%)
42	22191	22586	14.5	5.5		ORF33, <i>L. lactis</i> phage rlt, NP_695061 (75%)	<i>S. pyogenes</i> SSI-1, NP_802015 (70%); <i>S. agalactiae</i> 2603V/R, NP_687616 (68%); <i>S. pyogenes</i> MGAS315.2, NP_664915 (62%)
43	22573	22911	12.7	6.6		ORF34, <i>L. lactis</i> phage rlt, NP_695062 (66%)	<i>S. pyogenes</i> M1 GAS, NP_269537 (65%); <i>S. agalactiae</i> 2603V/R, NP_687617 (64%); <i>S. pyogenes</i> SSI-1, NP_802016 (63%)
44	22904	23140	8.6	10.1		<i>S. pyogenes</i> phage315.3, NP_664913 (62%)	<i>S. pyogenes</i> MGAS8232, NP_607544 (62%); <i>S. agalactiae</i> 2603V/R, NP_687618 (62%); Orf 35 <i>L. lactis</i> phage rlt, NP_695063 (57%)
45	23141	23476	12.6	5.2		<i>S. pyogenes</i> M1 GAS, NP_269535 (62%)	<i>S. pyogenes</i> MGAS8232, NP_607543 (61%); Orf 36, <i>L. lactis</i> phage rlt, NP_695064 (61%)
46	23492	24043	19.4	5.0	Major tail protein	Structural protein, <i>S. agalactiae</i> 2603V/R phage LambdaSa1, NP_687620 (62%)	Structural protein, <i>L. lactis</i> phage rlt, NP_695065 (61%); putative structural protein, phage-associated <i>S. pyogenes</i> phage 315.3, NP_664911 (61%); putative structural protein, phage-associated <i>S. pyogenes</i> M1 GAS, NP_269534 (59%)
47	24046	24357	11.6	9.4		<i>S. agalactiae</i> 2603V/R, NP_687621 (50%)	<i>S. pyogenes</i> M1 GAS, NP_269533 (48%); Orf 38, <i>L. lactis</i> phage rlt, NP_695066 (47%)

Continued on following page

TABLE 1—Continued

ORF	Start	Stop	MW	pI	Putative function	Best match ^{a,b}	Other BLAST homologies ^b
48	24372	24752	14.3	4.6		<i>S. pyogenes</i> phage 315.3, NP_664909 (55%)	<i>S. pyogenes</i> M1 GAS, NP_269532 (55%); <i>S. agalactiae</i> 2603V/R, NP_687622 (53%); Orf 39, <i>L. lactis</i> phage rlt, NP_695067 (51%)
49	24752	27775	107.4	4.8	Putative platelet binding protein PblA; tail tape measure protein	<i>S. agalactiae</i> 2603V/R phage LambdaSa1, NP_687623 (57%)	<i>L. lactis</i> phage rlt ORF40, NP_695068 (60%); <i>S. pyogenes</i> phage 315.3, NP_664908 (49%); tape measure protein, <i>L. lactis</i> bacteriophage TP901-1, NP_112708 (33%); putative tape measure protein, <i>L. lactis</i> phage rlt ORF 42, NP_695070 (30%)
50	27775	28470	26.1	8.4		<i>S. pyogenes</i> phage 315.3, NP_664907 (37%)	<i>S. pyogenes</i> MGAS8232, NP_607535 (37%); <i>S. pyogenes</i> M1 GAS, NP_269530 (36%)
51	28467	31655	121.0	8.1	Putative platelet binding protein PblB; tail fiber protein	<i>S. pyogenes</i> MGAS8232, NP_607534 (43%)	<i>S. pyogenes</i> phage 315.3, NP_664906 (43%); <i>S. pyogenes</i> M1 GAS, NP_269529 (43%)
52	31652	31861	7.7	4.9		—	—
53	31851	32498	24.1	4.8		<i>S. pyogenes</i> SSI-1, NP_802386 (30%)	<i>S. pyogenes</i> phage 315.1, NP_664532 (30%); <i>S. pyogenes</i> MGAS8232, NP_606638 (37%)
54	32515	32931	15.6	8.8		<i>S. agalactiae</i> 2603V/R phage LambdaSa1, NP_688829 (37%)	<i>L. casei</i> phage A2, NP_680498 (33%); <i>S. pneumoniae</i> phage Dp-1, CAB07984 (31%)
55	32918	33250	12.3	5.6	Holin	Putative holin, <i>S. pneumoniae</i> phage VO1, CAD35392 (84%)	Putative holin 2, <i>S. pneumoniae</i> phage MM1, NP_150181 (75%); putative holin, <i>S. agalactiae</i> 2603V/R phage LambdaSa2, NP_688828 (65%); putative holin, <i>S. pyogenes</i> phage 315.5, NP_665111 (50%)
56	33253	34140	34.3	4.7	Lysin	<i>N</i> -Acetylmuramoyl-L-alanine amidase, <i>S. pneumoniae</i> phage Dp-1, CAB07986 (72%)	Putative lysin, <i>S. agalactiae</i> 2603V/R phage LambdaSa1, NP_687631 (69%); <i>S. agalactiae</i> 2603V/R phage LambdaSa2, NP_688827 (68%)

^a Dashes indicate no significant homologies.

^b Numbers in parentheses indicate the percent identity.

the integrase and repressor genes of SM1 were clustered within just a few genes of each other, in a manner distinct from that of the prototypic *Siphoviridae* (19). Therefore, this conserved genetic organization, which is also present in SM1, may be a hallmark that distinguishes these phages from the larger superfamily. Additional unusual organizational features present in the genomes of SM1, r1t, and SF370.3 include the distinctive position of a putative structural gene (gp34 in SM1), as well as the one-to-one alignment of 12 morphogenetic genes with those of phages. These features further suggest that these phages are members of a discrete subdivision of *Siphoviridae*.

Classically, morphological features have been used for grouping or subdividing phages (1). However, as phages are characterized at the molecular level, differences in assembly mechanisms are being recognized which may not result in obvious morphological differences. In prototypic phage such as λ , the heads are assembled from the major capsid protein, along with a few accessory proteins, into an icosahedral shell. In most cases, these structures can be disassembled to release monomers of the major capsid protein. However, it was recently shown that the lambdoid coliphage HK97 may assemble its head by a second mechanism in which it is able to form a completely covalently linked shell consisting of a catenated latticework of hexamers and pentamers (32). This form of highly cross-linked capsid is also the mature form of the head

for a number of other phages, including the coliphage HK022 (18), L5 of *Mycobacterium tuberculosis* (13), and D3 of *Pseudomonas aeruginosa* (12). Thus, morphologically similar phages can be subdivided, based on differences in the capsid assembly mechanism.

Although SM1 and r1t resemble other members of the family *Siphoviridae* morphologically, our results with SM1 along with studies of r1t by others (30) indicate that these phages use a third, novel mechanism for head morphogenesis. That is, the major capsid proteins form discrete stable hexameric and pentameric oligomers. In HK97, stable hexamers and pentamers are catenated to form a covalently linked superstructure, as is evident from the relatively small amount of these oligomers present within phage preparations separated by SDS-PAGE (25). However, preparations of SM1 separated by SDS-PAGE (Fig. 4) revealed abundant amounts of pentamers and hexamers, indicating that they are indeed free to dissociate from the entire head structure. This distinct mechanism for capsid assembly may represent another feature that distinguishes SM1 and r1t from other double-stranded DNA icosahedral phages.

We showed previously that SM1 encodes two putative virulence factors, PblA and PblB, that mediate the binding of *S. mitis* to human platelets. Bioinformatic analysis of the proteins encoded in the SM1 genome did not reveal the presence of any other traditional virulence factors, such as toxins. However, the

carboxyl terminus of one gene product showed 43% identity to MafB, a putative neisserial adhesin. gp35 of SM1 appears to have arisen through the acquisition of a small portion of a neisserial chromosomal locus that contains *mafB* followed by a short neisserial ORF of unknown function. Despite the homology to a putative adhesin, further genetic and biochemical analysis will be required to determine whether gp35 is localized to the SF100 surface and whether gp35 contributes to the virulence of *S. mitis*.

Phage evolution appears to involve a great deal of genetic exchange that leads to a mosaic genome (14, 15). Such mosaicism is thought to occur through illegitimate recombination across the entire chromosome (14). Close inspection of the SM1 genome revealed that it contains homologs of over 20 different individual phages or bacterial species. The presence of regions of homology to neisserial sequences within genes 15 and 35 of the functional SM1 genome (Table 1) is of particular interest, since these neisserial coding sequences are likely to represent an abrupt shift for SM1 away from its ancestral subfamily. This high level of genetic plasticity may enhance its ability to disseminate its virulence factors.

Recent reports indicate that tape measure and tail fiber genes, such as *pblA* and *pblB* of SM1, are coding regions that are particularly amenable to genetic exchange. Recombination within the tail fiber genes appears to be important for diversifying the host range of individual phage types (20). Such a high degree of recombination within a gene that is able to undergo frequent mutation may have also facilitated the acquisition of motifs associated with adhesion by PblB. Moreover, the specific adhesive properties associated with PblA are especially intriguing, given the alternative functions proposed for the tape measure proteins of several mycobacteriophages. It has been speculated that these recently identified proteins may serve as signaling molecules (resuscitating factors) (24). Although the secondary functions of PblA and the mycobacteriophage tape measure proteins are entirely different, the possible association of these proteins and virulence suggests that certain phage tail proteins are especially amenable to the acquisition of additional nonphage functions. Further study of these putative virulence gene products will enable us to more precisely characterize these important functions and define how they may be spread among other hosts.

REFERENCES

- Ackermann, H. W. 2001. Frequency of morphological phage descriptions in the year 2000. *Arch. Virol.* **146**:843–857.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiler, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. Sonnhammer. 2002. The Pfam protein families database. *Nucleic Acids Res.* **30**:276–280.
- Bensing, B. A., C. E. Rubens, and P. M. Sullam. 2001. Genetic loci of *Streptococcus mitis* that mediate binding to human platelets. *Infect. Immun.* **69**:1373–1380.
- Bensing, B. A., I. R. Siboo, and P. M. Sullam. 2001. Proteins PblA and PblB of *Streptococcus mitis*, which promote binding to human platelets, are encoded within a lysogenic bacteriophage. *Infect. Immun.* **69**:6186–6192.
- Boyd, E. F., and H. Brussow. 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol.* **10**:521–529.
- Bruand, C., M. Farache, S. McGovern, S. D. Ehrlich, and P. Polard. 2001. DnaB, DnaD and DnaI proteins are components of the *Bacillus subtilis* replication restart primosome. *Mol. Microbiol.* **42**:245–255.
- Corpet, F., J. Gouzy, and D. Kahn. 1999. Recent improvements of the ProDom database of protein domain families. *Nucleic Acids Res.* **27**:263–267.
- Davey, M. J., L. Fang, P. McInerney, R. E. Georgescu, and M. O'Donnell. 2002. The DnaC helicase loader is a dual ATP/ADP switch protein. *EMBO J.* **21**:3148–3159.
- Davis, B. M., H. H. Kimsey, A. V. Kane, and M. K. Waldor. 2002. A satellite phage-encoded antirepressor induces repressor aggregation and cholera toxin gene transfer. *EMBO J.* **21**:4240–4249.
- Desiere, F., W. M. McShan, D. van Sinderen, J. J. Ferretti, and H. Brussow. 2001. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic streptococci: evolutionary implications for prophage-host interactions. *Virology* **288**:325–341.
- Gilakjan, Z. A., and A. M. Kropinski. 1999. Cloning and analysis of the capsid morphogenesis genes of *Pseudomonas aeruginosa* bacteriophage D3: another example of protein chain mail? *J. Bacteriol.* **181**:7221–7227.
- Hatfull, G. F., and G. J. Sarkis. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol. Microbiol.* **7**:395–405.
- Hendrix, R. W. 2002. Bacteriophages: evolution of the majority. *Theor. Popul. Biol.* **61**:471–480.
- Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. USA* **96**:2192–2197.
- Hoen, B., F. Alla, C. Selton-Suty, I. Beguinot, A. Bouvet, S. Briancon, J. P. Casalta, N. Danchin, F. Delahaye, J. Etienne, V. Le Moing, C. Lepout, J. L. Mainardi, R. Ruimy, and F. Vandenesch. 2002. Changing profile of infective endocarditis: results of a 1-year survey in France. *JAMA* **288**:75–81.
- Ishigo-Oka, D., N. Ogasawara, and S. Moriya. 2001. DnaD protein of *Bacillus subtilis* interacts with DnaA, the initiator protein of replication. *J. Bacteriol.* **183**:2148–2150.
- Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdaoid bacteriophages. *J. Mol. Biol.* **299**:27–51.
- Lucchini, S., F. Desiere, and H. Brussow. 1999. Similarly organized lysogeny modules in temperate Siphoviridae from low GC content gram-positive bacteria. *Virology* **263**:427–435.
- Lucchini, S., F. Desiere, and H. Brussow. 1998. The structural gene module in *Streptococcus thermophilus* bacteriophage phi Sfi11 shows a hierarchy of relatedness to Siphoviridae from a wide range of bacterial hosts. *Virology* **246**:63–73.
- Lukashin, A. V., and M. Borodovsky. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**:1107–1115.
- Nauta, A., D. van Sinderen, H. Karsens, E. Smit, G. Venema, and J. Kok. 1996. Inducible gene expression mediated by a repressor-operator system isolated from *Lactococcus lactis* bacteriophage r1t. *Mol. Microbiol.* **19**:1331–1341.
- Obregon, V., J. L. Garcia, E. Garcia, R. Lopez, and P. Garcia. 2003. Genome organization and molecular analysis of the temperate bacteriophage MM1 of *Streptococcus pneumoniae*. *J. Bacteriol.* **185**:2362–2368.
- Pedulla, M. L., M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence, W. R. Jacobs, R. W. Hendrix, and G. F. Hatfull. 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**:171–182.
- Popa, M. P., T. A. McKelvey, J. Hempel, and R. W. Hendrix. 1991. Bacteriophage HK97 structure: wholesale covalent cross-linking between the major head shell subunits. *J. Virol.* **65**:3227–3237.
- Riedel, H. D., J. Heinrich, A. Heisig, T. Choli, and H. Schuster. 1993. The antirepressor of phage P1. Isolation and interaction with the C1 repressor of P1 and P7. *FEBS Lett.* **334**:165–169.
- Shearwin, K. E., A. M. Brumby, and J. B. Egan. 1998. The Tum protein of coliphage 186 is an antirepressor. *J. Biol. Chem.* **273**:5708–5715.
- Sullam, P. M., D. G. Payan, P. F. Dazin, and F. H. Valone. 1990. Binding of viridans group streptococci to human platelets: a quantitative analysis. *Infect. Immun.* **58**:3802–3806.
- Sullam, P. M., F. H. Valone, and J. Mills. 1987. Mechanisms of platelet aggregation by viridans group streptococci. *Infect. Immun.* **55**:1743–1750.
- van Sinderen, D., H. Karsens, J. Kok, P. Terpstra, M. H. Rutgers, G. Venema, and A. Nauta. 1996. Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage r1t. *Mol. Microbiol.* **19**:1343–1355.
- Wagner, P. L., and M. K. Waldor. 2002. Bacteriophage control of bacterial virulence. *Infect. Immun.* **70**:3985–3993.
- Wikoff, W. R., L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix, and J. E. Johnson. 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**:2129–2133.