Characterization of *Streptococcus thermophilus* Host Range Phage Mutants

Martin Duplessis,1,2 Céline M. Lévesque,2† and Sylvain Moineau1,2,3*

Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie,1 Groupe de Recherche en Écologie Buccale (GREB), Faculté de Médecine Dentaire;2 and Félix d’Hérelle Reference Center for Bacterial Viruses,3 Université Laval, Quebec City, Quebec, Canada G1K 7P4

Received 25 July 2005/Accepted 12 January 2006

To investigate phage-host interactions in *Streptococcus thermophilus*, a phage-resistant derivative (SMQ-301R) was obtained by challenging a Tn917 library of phage-sensitive strain *S. thermophilus* SMQ-301 with virulent phage DT1. Mutants of phages DT1 and MD2 capable of infecting SMQ-301 and SMQ-301R were isolated at a frequency of 10⁻⁶. Four host range phage mutants were analyzed further and compared to the two wild-type phages. Altogether, three genes (orf15, orf17, and orf18) contained point mutations leading to amino acid substitutions and were responsible for the expanded host range. These three proteins were also identified in both phages by N-terminal sequencing and/or matrix-assisted laser desorption ionization-time-of-flight mass spectrometry. The results suggest that at least three phage structural proteins may be involved in phage-host interactions in *S. thermophilus*.

*Streptococcus thermophilus* is one of the most economically important lactic acid bacteria; it is used to manufacture yogurt and several hard cooked cheeses. *S. thermophilus* bacteriophages have been the subject of ongoing research because of their negative impact on the manufacture of these fermented dairy products (18). Several phage control strategies have been adapted to various dairy fermentation processes to curtail this recurrent problem. One universal method involves the isolation and use of bacteriophage-insensitive mutants (BIMs) derived from phage-sensitive strains (15). Essentially, phage-sensitive bacterial cells are exposed to a high concentration of virulent phages for an extended period, following which a few outgrowing cells, presumably mutated in the phage receptor, are selected for further characterization.

The phage infection process starts with the specific interaction between the receptor binding protein (RBP) located on the distal part of the phage tail and the phage receptor on the host surface. A modification to either of these two structures may influence the phage-host interaction. Phage receptors in *S. thermophilus* have yet to be identified, but some progress has been made in defining the chemical composition of the receptors. Glucosamine, N-acetylg glucosamine, and rhamnose reduce phage adsorption (1), as does treatment of the bacterial cell envelopes with mutanolysin or trichloroacetic acid, suggesting that phage receptors are part of the peptidoglycan layer or a polymer closely linked to it (22). Phage-resistant derivatives have also been obtained by generating *S. thermophilus* mutants using a thermolabile insertion vector. However, none of the mutations prevented phage adsorption (14).

The RBP (ORF18) of virulent *S. thermophilus* phage DT1 has been identified (4). Chimeric DT1 phages, constructed by swapping orf18 of DT1 with the corresponding orf18 of virulent phage MD4, acquire the host range of MD4. Despite this, the DT1 chimeric phages are still able to adsorb to their original host, indicating that other phage proteins are involved in phage adsorption (4). The RBPs of other dairy phages (from *Lactococcus* spp.) have also been characterized using a similar strategy (5, 24). Another approach for studying phage-host interactions is the characterization of phage mutants capable of infecting BIMs (3, 21, 23, 26). This methodology was used to identify the RBP of *Lactobacillus* phage LL-H in host range mutants with a single nucleotide mutation causing an amino acid substitution in the C-terminal portion of gp71 (23).

The aim of the present study was to identify other genetic determinants of the virulent cos-type phage DT1 involved in recognition of its *S. thermophilus* host (SMQ-301) by characterizing host range phage mutants.

**Phage mutant isolation, host range, and adsorption assays.** Despite several challenge experiments, we were unable to isolate BIMs of *S. thermophilus* SMQ-301. It is unclear why we were unable to obtain BIMs with this particular strain and phage, because the same protocol was successful with other *S. thermophilus* phage-host systems (data not shown). To obtain a phage-resistant derivative, we used a different approach by first constructing a Tn917 mutant library of SMQ-301 (7, 12) and then challenging it with phage DT1 at a multiplicity of infection of 5. One phage-insensitive derivative (SMQ-301R) was obtained. Furthermore, the lysate of phage DT1 (GenBank accession number AF085222) contained phage mutants (frequency, 10⁻⁶) still able to infect SMQ-301R. A similar result (frequency, 10⁻⁶) was obtained using a lysate of phage MD2 (GenBank accession number AF348736). Phage MD2 was also tested because of its overlapping host range with DT1 and its different RBP gene (4).

Two host range mutants each of phages DT1 and MD2 (DT1.R1, DT1.R2, MD2.R1, and MD2.R2), which were capable of replicating on SMQ-301R, were selected from isolated plaques, amplified, and immediately used for DNA extraction. The restriction profiles of the host range mutants were the same as those of the wild-type phages (data not shown), and the four phage mutants were able to form plaques on both SMQ-301 and SMQ-301R (efficiency of plaquing, 1.0). Phage adsorption assays showed that the wild-type phages...
DT1 and MD2 still adsorbed to mutant strain SMQ-301R, but at lower levels than to the wild-type strain SMQ-301 (Table 1). The four host range phage mutants isolated using SMQ-301R were also able to adsorb and infect SMQ-301 (Table 1). The adsorption of the host range phage mutants was lower on SMQ-301R than on the wild-type strain (Table 1).

Localization of mutations on the host range phage mutant genomes. To identify the mutation(s) leading to the extended host range, the complete nucleotide sequence of phage mutant DT1.R2 was first determined (34,815 bp). A genomic comparison with the corresponding sequence of wild-type phage DT1 uncovered a total of three nucleotide changes leading to amino acid substitutions in the putative tail proteins ORF15 and ORF17 as well as in the RBP (ORF18) (Table 2). The region from orf15 to orf17 was then sequenced in the genome of DT1.R1, and a total of two missense mutations were detected, one in orf15 and the other in orf17 (Table 2). The mutations in orf15 and orf17 were at positions different from those found in DT1.R2 (Table 2; Fig. 1). The MD2.R1 and MD2.R2 mutants also possessed two missense mutations, one in orf17 and the other in orf18. The mutation in orf18 was the same for both phages (MD2.R1 and MD2.R2), while the mutations differed in orf17 (Table 2; Fig. 1).

ORF15. orf15 of DT1 codes for a putative 1,658-amino-acid (aa) tail protein. The ORF15 proteins of DT1 and MD2 shared 99% amino acid identity. In silico analysis revealed that orf15 likely codes for the tape measure protein (TMP), which is responsible for determining tail length (8). ORF15 also possesses a putative coiled-coil structure (highly α-helical) at its N-terminal portion (aa 1 to 200) and a variable region (VR3) in the central section of the deduced protein sequence (aa 651 to 900) (Fig. 1A). Lucchini et al. (13) previously postulated that the variable region of ORF15 is involved in host specificity.

Interestingly, a CHAP domain (cysteine, histidine-dependent aminohydrolase/peptidase) (aa 1244 to 1363) and a soluble lytic transglycosylase (SLT) motif (aa 1431 to 1553) were found in the C-terminal region of ORF15 (Fig. 1A) using MotifScan (http://myhits.isb-sib.ch/cgi-bin/motif_scan). Both motifs are associated with a murein hydrolase activity, suggesting that ORF15 is probably a bifunctional peptidoglycan lysozyme with both endopeptidase and glycosidase activities (20). Cell wall-degrading activity associated with structural proteins is widespread in phages of both gram-negative and gram-positive bacteria (10, 17). This activity is needed to break down the peptidoglycan layer to allow the ejection of the phage DNA into the bacterial cell (16).

Only the two DT1 host range mutants had mutations in orf15. For DT1.R1, the mutation (I1419V) was located between the CHAP and SLT domains (Table 2; Fig. 1C). The valine at this position is unique to DT1.R1 and is not found in the ORF15 homologues of other S. thermophilus phages (Fig. 1C). The mutation in DT1.R2 (V1254L) was located within the CHAP domain, 10 amino acids upstream from the putative catalytic cysteine (Table 2; Fig. 1B). The leucine at this position is also unique to DT1.R2 (Fig. 1B). Lastly, both exclusive mutations were located in regions that are highly conserved in cos-type S. thermophilus phages.

ORF17. The ORF17 proteins of DT1 and MD2 are also nearly indistinguishable, with 99% amino acid identity. This phage protein is highly conserved in S. thermophilus cos-type phages, but a function and motifs have yet to be assigned to this protein. The ORF17 homologue in pac-type phage 2972 (ORF19) is a hybrid structural protein that links the two S. thermophilus phage groups (11). The N- and C-terminal regions of ORF19 of phage 2972 share homology with pac-type phages, while the central part of the deduced protein is closer to cos-type phages (11). Interestingly, all mutations in the ORF17 proteins of the four host range phage mutants were located in a short stretch of five amino acids (FLYGV) located in the central part of the protein (Fig. 1D), a region named the FLY motif.

ORF18. The RBP of DT1 possesses two domains, while the RBP of MD2 has three (4). The domains are delimited by motifs called collagen-like repeats (Fig. 1A), which may be recombinational hot spots for DNA rearrangements or domain shuffling (2, 4, 13). The first domains (aa 1 to 510) of DT1 and MD2 are identical (100% amino acid identity). The second domain (aa 492 to 896) of phage 2972 (approximately 55% amino acid identity). Mutations in ORF18 were observed in three of the four host range phage mutants (DT1.R1, MD2.R1, and MD2.R2). The mutation in DT1.R2 (F186I) was located in the amino-terminal region of the protein (Table 2; Fig. 1A). This region is conserved in cos-type phages of S. thermophilus and may be involved in the binding of the RBP to the phage structure (4). The same mutation (P798H) was found in the ORF18 proteins of both MD2 host range mutants and was located in the third collagen-like repeat (second domain), which is absent in the ORF18 of DT1 (Fig. 1A).

Structural protein identification of DT1 and MD2. CsCl-purified DT1 and MD2 phage particles were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-
FIG. 1. (A) Diagrams of ORF15, ORF17, and ORF18 of phages DT1 and MD2. Putative conserved domains are displayed in white. The FLY motif of ORF17 is also identified. Collagen-like repeats are numbered and are represented by small dark-gray regions within ORF18. VR1, VR2, and VR3 indicate variable regions 1, 2, and 3. (B) Alignment of ORF15 (aa 1236 to 1265) where a mutation occurred in DT1.R2. (C) Alignment of ORF15 (aa 1405 to 1434) where a mutation occurred in DT1.R1. (D) Alignment of ORF17 (aa 313 to 352), including the FLY motif, where mutations were present in all the host range phage mutants. Amino acid substitutions are in bold. Numbers represent the positions on the amino acid sequence of DT1.
to determine whether ORF15, ORF17, and ORF18 were present in their structures. Seven structural proteins were identified in each phage by N-terminal sequencing or mass spectrometry (matrix-assisted laser desorption ionization–time-of-flight [MALDI-TOF] mass spectrometry or liquid chromatography-tandem mass spectrometry [LC-MS-MS]) as described by Lévesque et al. (11). Three DT1 structural proteins have already been identified: ORF8 (major capsid protein), ORF13 (major tail protein), and ORF19 (tail protein) (11, 25). In the present study, four additional proteins were identified in DT1 (Fig. 2): ORF6 (putative portal protein), ORF15 (putative tail protein), ORF17 (putative tail protein), and ORF18 (RBP). The same seven proteins were also found in the structure of phage MD2 (Fig. 2), suggesting that they may be conserved in cos-type phages of *S. thermophilus* (11).

The three proteins that were mutated in the four host range phage mutants were present in the phage structure. ORF8, ORF13, ORF17, ORF18, and ORF19 had molecular weights similar to the predicted molecular weights (Fig. 2). ORF6 of DT1 began at the ninth amino acid and was probably processed to generate the mature protein, as observed with the putative portal protein (gp3) of *Lactobacillus* phage A2 (6). Interestingly, the ORF15 proteins of both DT1 and MD2 were significantly smaller than expected: their estimated molecular sizes were approximately 110 kDa, while the sizes of the deduced proteins were 182 kDa (Fig. 2). The N-terminal sequence of ORF15 starts at the predicted second amino acid, suggesting that the C-terminal region of the protein is processed. To verify this hypothesis, the

<table>
<thead>
<tr>
<th>#</th>
<th>MW (kDa)</th>
<th>N-terminal sequencing or mass spectrometry</th>
<th>Description*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>108</td>
<td>AKIQATMSTE and MALDI-ToF</td>
<td>DT1, ORF15, tail protein</td>
</tr>
<tr>
<td>2</td>
<td>103</td>
<td>MQIWHDKSM and MALDI-ToF</td>
<td>DT1, ORF18, RBP</td>
</tr>
<tr>
<td>3</td>
<td>78</td>
<td>VEFWSNND</td>
<td>DT1, ORF19, tail protein</td>
</tr>
<tr>
<td>4</td>
<td>52</td>
<td>MIGMSVTYDG</td>
<td>DT1, ORF17, tail protein</td>
</tr>
<tr>
<td>5</td>
<td>37</td>
<td>LATEPPIN</td>
<td>DT1, ORF6, portal protein</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
<td>LLDSKTDHSY</td>
<td>DT1, ORF8, major head protein</td>
</tr>
<tr>
<td>7</td>
<td>26</td>
<td>AIVGLKLVLKL</td>
<td>DT1, ORF13, major tail protein</td>
</tr>
<tr>
<td>8</td>
<td>142</td>
<td>LC-MS-MS</td>
<td>MD2, ORF18, RBP</td>
</tr>
<tr>
<td>9</td>
<td>118</td>
<td>LC-MS-MS</td>
<td>MD2, ORF15, tail protein</td>
</tr>
<tr>
<td>10</td>
<td>89</td>
<td>—</td>
<td>MD2, ORF19, tail protein</td>
</tr>
<tr>
<td>11</td>
<td>58</td>
<td>LC-MS-MS</td>
<td>MD2, ORF17, tail protein</td>
</tr>
<tr>
<td>12</td>
<td>41</td>
<td>LC-MS-MS</td>
<td>MD2, ORF6, portal protein</td>
</tr>
<tr>
<td>13</td>
<td>33</td>
<td>LC-MS-MS</td>
<td>MD2, ORF8, major head protein</td>
</tr>
<tr>
<td>14</td>
<td>26</td>
<td>LC-MS-MS</td>
<td>MD2, ORF13, major tail protein</td>
</tr>
</tbody>
</table>

FIG. 2. Protein profiles of phages DT1 and MD2 on Coomassie blue-stained SDS-PAGE gels. MW, molecular weight markers. The asterisk indicates that the protein functions are putative except for ORF18 of DT1 (protein 2).
FIG. 3. Amino acid sequences of the ORF15 proteins of phages DT1 and MD2. The light gray region represents the protein identified in the phage structure by MALDI-TOF or LC-MS-MS. Underlined amino acid sequences, peptides identified by mass spectrometry. Variable region, the region that differs from the homologue proteins of other cos-type S. thermophilus phages. Dark-gray regions, CHAP and SLT domains. Boldfaced amino acids in DT1 indicate the amino acids that were mutated in phages DT1.R1 and DT1.R2.
ORF15 proteins of DT1 and MD2 were analyzed by mass spectrometry (11). All the peptides that matched the protein sequence were located in the N-terminal region of the protein (Fig. 3). These results clearly demonstrated that ORF15 was indeed processed and that the C-terminal part of the protein (ORF15-C) was removed. ORF15-C was not found in the phage structure by mass spectrometry using this methodology.

As indicated above, ORF15 is presumably the TMP. Several TMPs are processed during the intracellular development of phages. The TMP (pGpH) of coliphage lambda (92.3 kDa) is also processed at the C-terminal end to give pGpH* (80 kDa) (8). The TMP of lactococcal phage TP901-1 is trimmed from 100 kDa to 70 kDa (19). The TMP (gp12) of Lactobacillus phage A2 is also processed, with the mature protein lacking the first 172 aa. The N-terminal section that is cut off is not present in the phage virion (6). An internal portion of approximately 270 aa of the putative TMP (gp12) of Listeria phage PSA is not present in mature virions (27).

Interestingly, the mutations in the ORF15 proteins of the two DT1 mutants were located in the region that is potentially cleaved off and that contained the CHAP and SLT motifs found in murein hydrolase enzymes. Kenny et al. (9) characterized the Tat3500 protein, which has cell wall-degrading activity, at the tip of the tail of lactococcal phage TUC2009. It is tempting to speculate that ORF15-C has a similar function and that perhaps each phage contains very few molecules of ORF15-C.

Frequency of host range mutants. The four host range phage mutants were obtained at a frequency of 10^-4. Host range mutants of Lactobacillus phage LL-H, which were also isolated at a frequency of 10^-4 to 10^-7, contain only a single mutation in gp71 (23). Similarly, host range phage mutants of Sinorhizobium meliloti were the only common denominator between the four host range mutants of S. thermophilus. While mutations in ORF17 were the only common denominator between the four host range mutants isolated, it is possible that a single amino acid change is sufficient to expand the host range of the two wild-type phages. Alternatively, other compensatory mutations may be needed to efficiently expand the host range.

Conclusions. Four host range mutants isolated from the ly- sates of S. thermophilus phages DT1 and MD2 were characterized. Host range data indicated that the specificity of the phage mutants was expanded, since they were still able to infect the wild-type strain. The mutations that expanded the host range were located in putative tail proteins, including RBP. However, ORF17 was the only structural protein mutated in all four phage mutants. These results indicate that at least three phage structural proteins may be involved in the phage-host interactions in S. thermophilus.

We are grateful to Denise Tremblay for technical assistance in the late stages of the study.

M.D. is the recipient of a graduate student scholarship from the Fonds Québécois de la Recherche sur la Nature et les Technologies (FRQNT). This work was funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

REFERENCES