comYA, a Gene Similar to comGA of Bacillus subtilis, Is Essential for Competence-Factor-Dependent DNA Transformation in Streptococcus gordonii

R. DWAYNE LUNSFORD* AND ARLENE G. ROBLE†

Oral Infection and Immunity Branch, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland

Received 24 December 1996/Accepted 14 March 1997

Tn4001 mutagenesis identified a new competence gene in Streptococcus gordonii Challis designated comYA. A comYA mutant was completely deficient in transformation and exhibited decreased levels of DNA binding and hydrolysis. The deduced 319-amino-acid ComYA protein exhibited 57% similarity and 33% identity to the ComGA transporter protein of Bacillus subtilis and contained the Walker A-box motif conserved in ATP-binding proteins as well as aspartic acid boxes Asp-1 and Asp-2 present in some components of the general secretory pathway of gram-negative bacteria. comYA appeared to be part of a putative operon encompassing a comGB homolog, designated com1B, together with sequences that could encode ComGC- and ComGD-like peptides designated ComYC and ComYD, respectively, as well as other components. The putative ComYC and ComYD peptides had leader sequences similar to the type IV N-methylphenylalanine pilins of gram-negative bacteria, but unlike other examples in this class, including B. subtilis, they contained an alanine at position -1 of the leader instead of the usual glycine residue. Northern analysis identified a single 6.0-kb comYA-containing transcript strictly dependent on exogenous competence factor for expression in ComA1 cells. An identical pattern of expression was seen in wild-type Challis cells grown under conditions of maximal competence but not in cells that were noncompetent.

Oral streptococci play a critical role in the normal microbial ecology of the human oral cavity by functioning as primary colonizers of hard tissue surfaces (29). Members of this group also are the leading cause of subacute bacterial endocarditis and dental caries. Several species are proficient in natural genetic transformation, a competence-phase-specific ability to incorporate free exogenous DNA from the environment. Gram-positive transformable bacteria generally process DNA in a similar fashion. Competent cells bind native DNA in a sequence-independent manner. Following binding, one strand of the helix is nicked, the other is hydrolyzed via an entry nuclease, and the intact single strand is internalized, whereafter it can integrate into the host chromosome or, in the case of a plasmid molecule, establish itself extrachromosomally. Competence in Streptococcus gordonii, as well as other streptococci, is an early- to mid-growth-phase state in which nearly every cell in the population is transformable although exact periods of competence vary between strains and species. In contrast, Bacillus subtilis becomes competent only in late stationary phase and then with only a small subpopulation of cells being able to transform (21). Competence in S. gordonii Challis is modulated by unknown environmental stimuli that trigger the production of a secreted protein competence factor (CF) (13). Amplification of the initial stimulus at a critical cell density by CF, at least under in vitro conditions, assures complete competence induction in all cells of a particular culture via the induction of competence-specific gene expression (26). Whether similar levels of competence expression and kinetics of induction exist in the human oral cavity remains to be determined.

Challis CF was originally described as a trypsin-sensitive heat-resistant cationic peptide of small size (14). More recently, the synthetic operon for CF as well as the receptor sensing pathway of induction exist in the human oral cavity remains to be determined.

MATERIALS AND METHODS

Tn4001 mutagenesis. All strains and plasmids utilized in this study are listed in Table 1. S. gordonii Challis-2 (Ch2) wild-type was mutagenized as previously described by transformation with the Tn4001 delivery vector pU (16). A total of 800 Gm tk Em clones were subjected to a 96-well plate microtransformation assay as follows. Clones were grown for 18 h in 100 μl of Todd-Hewitt broth (Difco Laboratories, Detroit, Mich.) containing 50 μg of gentamicin per ml. The cultures were diluted 1:100 into drug-free transforming broth containing 5 μg of Sm tk DNA from strain Ch5 per ml (16). Following incubation for 6 h, the cultures were diluted 1:100 into 100 μl of brain heart infusion (Difco) containing 500 μg of streptomycin per ml. Absence of or decreased Sm tk growth in each well was scored after 16 h of incubation. Potential mutants exhibiting transformation defects were rescreened in a standard quantitative tube transformation assay (16).

**Mutant characterization.** Clones exhibiting defective transformation properties with genomic DNA were tested for the following phenotypes: (i) the ability...
Currently, IPCR inserts were utilized directly as probes to isolate genomic clones.

**RESULTS**

Generation of transposon mutants and phenotype characterization. Ch2 was mutagenized with Tn4001 to yield 500 Gm3 Emr clones that were screened in a microtransformation assay. Of this population, 16 were found to be defective in competence. Southern analysis of genomic DNA with an IS256-specific probe indicated that 14 members of this population contained a common insertion site in the comD locus represented by m3D3 (reference 11 and data not shown). We previously have shown that Tn4001 insertion is fairly random after screening transformants solely for Gm3 (16). However, in the context of competence mutants at least, there is clearly a hot spot for insertion of this element into comD. The two remaining mutants contained unique insertion sites, with one, m8G2, being the focus of this study. The remaining mutant appeared to involve a regulatory locus and will be described elsewhere. m8G2 was screened for the ability to transform with both genomic and plasmid DNA, production of and stimulation by CF, sensitivity to UV light, and 3H-DNA uptake and solubilization. Cells were found to be completely defective for transformation with either genomic or plasmid DNA and were resistant to UV light, also suggesting that the defect did not involve conventional recombination pathways (27). m8G2 was proficient in CF production but was clearly defective in DNA uptake and processing.

**Insertion site characterization.** Sequence information flanking transposon insertion sites was generated by IPCR. Genomic DNA from each mutant was digested with MboI and self-ligated under dilution conditions that enriched the DNA for intramolecular circularization. This material was used as template for PCRs with outward-oriented primers based on the IS256 sequence (5). Reaction products were cloned directly into pΩ to generate targeting vectors for insertion-duplication mutagenesis with which to confirm that the original mutants were indeed due to Tn4001 insertions (16). As expected, pCC93 regenerated the original transformation-defective phenotype of m8G2. The insertion site was found to reside within containing wild-type loci from a fosmid library (20), m8G2 yielded pCC93 and f1D5. Also isolated was the m3D3-specific fosmid f1H7.

**DNA sequence analysis.** IPCR products and fosmid clones were utilized directly as sequencing templates (20). The sequence was assembled with AssemblyLIGN and analyzed with MacVector (International Bioinformatics, Inc., New Haven, Conn.). Database searches were conducted with the BLAST network system (National Center for Biotechnology Information-National Library of Medicine, Bethesda, Md.). Protein alignments were generated with GAP (Wisconsin Package, version 8.0 for VMS-VAX; Genetics Computer Group, Madison, Wis.).

**CF synthesis.** Ch2 CF propeptide was synthesized on a Synergy 432A personal peptide synthesizer (Applied Biosystems, Foster City, Calif.) with fast-atom-bomb chemistry and a sequence based on the ComC theoretical peptide. Sequencing of the leader sequence occurred just C-terminal to the double glycines at positions 30 and 31 (deduced sequence from U80077). The product was purified by C18 hydrophobic-affinity high-pressure liquid chromatography, assayed on ComA1, and stored desiccated at 4°C. Aliquots were prepared in phosphate-buffered saline immediately before use. Activity for competence induction was maximal at 1 ng/ml.

**Northern hybridization analysis.** Ten-microgram samples of total RNA extracted from 4- or 16-h Ch2 cultures and ComA1 cells induced with either pure CF (100 ng/ml) or buffer controls in competence medium were isolated and subjected to Northern analysis as described elsewhere (17). A ComA probe was synthesized as a PCR product and corresponded to bp 2268 to 3025 of U81957 synthesized with 5’-caggagacctggcgtgatgatcgtgag3’ and 5’-cgatggtgatgatatagctctggc5’. The fragment was isolated by gel electrophoresis in low-melting-point agarose with NACs affinity resin (GIBCO-BRL), labeled with 32P by nick translation (Promega Corp., Madison, Wis.), and utilized at 107 dpm/ml in Quick-Hyb solution (Stratagene) at 65°C. Washes were at high stringency (0.1 X SSC) and 0.15 M NaCl plus 0.015 M sodium citrate-0.1% sodium dodecyl sulfate, 60°C.

**Nucleotide sequence accession number.** The sequences were submitted to GenBank under accession no. U81957 (comD region of f1D5) and U80077 (comD region of f1H7).

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Genotype and/or phenotype of strain or plasmid</th>
<th>Reference or supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spp. and strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ch2</td>
<td>Chalis Rf*, Xfo*</td>
<td>20</td>
</tr>
<tr>
<td>Ch5</td>
<td>Chali Sm*, Xfo*</td>
<td>20</td>
</tr>
<tr>
<td>Chali</td>
<td>Chali Rec*</td>
<td>27</td>
</tr>
<tr>
<td>ComA1</td>
<td>Chali comAc:FAM Xfo: CF*</td>
<td>20</td>
</tr>
<tr>
<td>m3D3</td>
<td>comDRTTn4001</td>
<td></td>
</tr>
<tr>
<td>m8G2</td>
<td>comAYTn4001</td>
<td></td>
</tr>
<tr>
<td><strong>E. coli DH10B</strong></td>
<td>F’ mecA Δ(mrr hsdM Rmr mcrBC) endA recA</td>
<td>16</td>
</tr>
</tbody>
</table>

**Abbreviations:** Rf*, Sm*, Em*, Ap*, and Sp* resistant to rifampin, streptomycin, erythromycin, ampicillin, and spectinomycin, respectively; Xfo*, transformation positive; Xfo-, transformation negative. Entries lacking reference or source designations were either laboratory stocks or strains or plasmids constructed in this study. Antibiotic concentrations used: for S. gordonii, 10 μg of erythromycin, 500 μg of streptomycin, 100 μg of rifampin, and 800 μg of spectinomycin per ml; for E. coli, 300 μg of erythromycin, 50 μg of spectinomycin, and 100 μg of ampicillin per ml. 

*TABLE 1. Strains and plasmids*
an open reading frame (ORF) with similarity to the comGA gene of B. subtilis (1). Insertion was precisely between nucleotides 2620 and 2621 of our sequence. We have designated this putative gene in S. gordonii comYA. A total of 4,566 bp of DNA sequence was determined in the region, representing all or portions of six major ORFs (GenBank accession no. U81957). The first, ΔpocD, represented the far C-terminal portion of the β′ subunit of RNA polymerase core enzyme (2). Although the region exhibited significant similarity to several bacterial and eucaryotic DNA-directed RNA polymerases, the highest match was to the B. subtilis enzyme, with 82% similarity and 73% identity. Spontaneous rifampin resistance generally maps as point mutations to the β subunit of RNA polymerase. In B. subtilis, the β subunit ORF lies just upstream of β′. A similar structure probably exists in S. gordonii, since we could demonstrate a high rate of cotransformation between Rf′ and Em′ by using genomic DNA from Ch2 (Rf′) clones transformed with pCC93 (13.5% cotransformation linkage compared to 0.4% linkage for negative control DNA). orfE could potentially encode a 122-amino-acid peptide with some similarity to DNA-binding proteins, specifically a 40-amino-acid domain of the subunit of RNA polymerase. Em′ allowed by a hydrophobic stretch of 20 to 30 residues that probably specify membrane localization. In fact, leader cleavage is required for subsequent membrane insertion (7). Sequence similarity then diverges distal of these conserved N-terminal domains. As expected, alignments of ComYC and ΔComYD with ComGC and ComGD indicated that the N-terminal regions were generally similar (Fig. 4). The streptococcal leaders were longer, however (11 and 15 amino acids compared to 5 and 10 for the B. subtilis proteins, respectively), but contained the conserved lysine residue at -2. Most surprising was the presence of an alanine residue at position -1 in both ComYC and ΔComYD. The glycine residue normally present at position -1 produces a mature pilin-like protein with an N-terminal phenylalanine at position +1 followed by a hydrophobic stretch of 20 to 30 residues that probably specify membrane localization. In fact, leader cleavage is required for subsequent membrane insertion (7). Sequence similarity then diverges distal of these conserved N-terminal domains. As expected, alignments of ComYC and ΔComYD with ComGC and ComGD indicated that the N-terminal regions were generally similar (Fig. 4). The streptococcal leaders were longer, however (11 and 15 amino acids compared to 5 and 10 for the B. subtilis proteins, respectively), but contained the conserved lysine residue at -2. Most surprising was the presence of an alanine residue at position -1 in both ComYC and ΔComYD. The glycine residue normally present at position -1 is conserved in every known example of a prepeplin or prepilin-like protein previously examined (12).

RNA analysis. Total RNA was isolated from cultures of the CF-defective mutant ComA1 and wild-type strain Ch2 that had
either been induced with pure CF or grown for periods of time in which competence is either maximal or nonexistent in our culture system. Results of Northern analysis of the samples are shown in Fig. 5. Each panel represents hybridization with a comYA-specific probe. Figure 5A contains RNA from ComA1 cells pulsed with CF (lane 1) or buffer control (lane 2). Figure 5B contains RNA from Ch2 cells at maximum competence (lane 3) or noncompetent 18-h cells (lane 4). Clearly evident is the strict dependence on exogenous CF for the expression of a single 6.0-kb transcript containing comYA sequences. Furthermore, temporal expression of this transcript followed the competent state of the cultures. Four-hour cells were maximally competent, with the production of a transcript, and 18-h cells were noncompetent, with essentially no detectable transcription.

**DISCUSSION**

The only transformation mutant previously described for *S. gordonii* is the chemically derived recombinant-defective strain CipA11 (27). This strain, while being sensitive to UV light and methylating agents, is impaired for chromosomal transformation but still capable of reduced levels of transformation with plasmid DNA. For this reason, conventional transposon mutagenesis was utilized to generate a collection of transformation-defective strains with which to identify new genes involved in the development of competence. Traditionally, Tn916 has been utilized for mutagenesis in the viridans streptococci. Tn916 is an *Enterococcus faecalis* conjugal transposon capable of encoding both transposition and DNA mobilization functions (8). We decided to utilize the *Staphylococcus aureus* composite-type transposon Tn4001 instead because of its small size, stability, and lack of any DNA mobilization machinery that could complicate analysis of transformation functions in host cells. Tn4001 is a 4.5-kbp element composed of a central resistance gene encoding a single protein with both aminoglycoside acetyltransferase and phosphotransferase activities flanked by dual copies of IS256 (16). Our present Tn4001 insertion collection represents several new classes of mutants defective in competence regulation, CF production, and DNA processing. The comYA mutant described here was unique in that it involved the first streptococcal competence locus with significant structural and sequence similarity to a *B. subtilis* operon essential for transformation.

The putative comY operon currently consists of four reading frames with significant similarity to the *B. subtilis* comG locus (1). Additional comG homologs are expected in Challis as DNA sequence analysis progresses in this region. The Tn4001 insertion in m8G2 was found to reside within comYA. ComYA and ComYB are similar to ComGA and ComGB. ComYA possesses signature domains indicative of a protein involved with secretion. A single 6.0-kb transcript containing comYA sequences is clearly induced by CF in the nonproducing mutant ComA1. Similar inducibility was noted in wild-type cultures grown under natural conditions of either competence or non-competence. We also have noticed strict CF-dependent synthesis of a 2.4-kb comD-containing transcript in ComA1 cells as well as in maximally competent Ch2. No obvious message was detected in buffer-pulsed ComA1 or in noncompetent Ch2 cells (data not shown). These observations for *S. gordonii* are in sharp contrast to studies with *Streptococcus pneumoniae* indicating either partial inducibility (23) or constitutive expression (6) of the comCDE locus.

The ComGC, ComGD, and ComGE proteins of *B. subtilis* represent small peptides of 98, 143, and 115 amino acids, respectively, that contain leader sequences of high similarity to those present on several gram-negative type IV prepilins and components of protein secretion systems (12). ComYC and ComYD are similar to ComGA and ComGB. ComYB are similar to ComGA and ComGB. ComYD are unique in that a ComC-like peptidase activity. ComYC and ΔComYD are unique in that a conserved glycine residue present at position −1 of the deduced leaders are found in boldface with the downward arrows indicating the probable point of cleavage.
these small peptides combine to form a pseudopilus-like structure that functions as an aqueous pore or tube spanning the cell wall and membrane to function as a direct route for DNA entry (12). In fact, ComGC mutants of \textit{B. subtilis} are completely defective in competence and fail to bind DNA above background levels seen with noncompetent cells (3). However, the situation appears to be different in streptococci. We have previously described Mg\textsuperscript{2+}-independent DNA nicking and binding activities in surface extracts of both competent and noncompetent Challis cultures (19). Levels of these activities appear to be the same regardless of the competence state of the cells. Southwestern blot analysis of the preparations identified a single 56-kDa DNA-binding protein present in both competent and noncompetent cells that could function as a primary DNA receptor. In \textit{S. pneumoniae}, the 30-kDa cell wall endonuclease EndA has been shown to be essential for competence (25). EndA also was found to be constitutively expressed. It would appear, therefore, that at least some components of the DNA recognition and uptake machinery in streptococci are not directly inducible by CF and that these components could associate with inducible subunits to produce a complete DNA uptake apparatus. It is tempting to speculate that the ComY structure may have a dual function: first, to provide the actual pathway for DNA transport by supplying a channel capable of spanning both the cell membrane and the relatively thick streptococcal cell wall; and second, to act as a scaffolding structure able to sequester constitutively expressed components such as the primary DNA receptor and nickase from the membrane and project them through the cell wall toward the extracellular milieu. Studies to determine the exact composition and topology of the \textit{S. gordonii} DNA uptake system are in progress.

**ACKNOWLEDGMENTS**

Frank Robey is gratefully acknowledged for his assistance with high-pressure liquid chromatography purification of the CF peptide. Myra Jacobs and Jack London provided helpful comments on the manuscript.

**REFERENCES**