The response regulator ComE in \textit{Streptococcus mutans} functions both as a transcription activator of mutacin production and repressor of CSP biosynthesis

Jens Kreth,\textsuperscript{1} David C. I. Hung,\textsuperscript{2} Justin Merritt,\textsuperscript{1} Julie Perry,\textsuperscript{3} Lin Zhu,\textsuperscript{1} Steven D. Goodman,\textsuperscript{2} Dennis G. Cvitkovitch,\textsuperscript{3} Wenyuan Shi\textsuperscript{1} and Fengxia Qi\textsuperscript{1}\

In \textit{Streptococcus pneumoniae}, competence and bacteriocin genes are controlled by two two-component systems, ComED and BlpRH, respectively. In \textit{Streptococcus mutans}, both functions are controlled by the ComED system. Recent studies in \textit{S. mutans} revealed a potential ComE binding site characterized by two 11 bp direct repeats shared by each of the bacteriocin genes responsive to the competence-stimulating peptide (CSP). Interestingly, this sequence was not found in the upstream region of the CSP structural gene \textit{comC}. Since \textit{comC} is suggested to be part of a CSP-responsive and ComE-dependent autoregulatory loop, it was of interest to determine how it was possible that the ComED system could simultaneously regulate bacteriocin expression and natural competence. Using the intergenic region IGS1499, shared by the CSP-responsive bacteriocin \textit{nlmC} and \textit{comC}, it was demonstrated that both genes are likely to be regulated by a bifunctional ComE. In a \textit{comE} null mutant, \textit{comC} gene expression was increased similarly to a fully induced wild-type. In contrast, \textit{nlmC} gene expression was nearly abolished. Deletion of ComD exerted a similar effect on both genes to that observed with the \textit{comE} null mutation. Electrophoretic mobility shift assays (EMSA) with purified ComE revealed specific shift patterns dependent on the presence of one or both direct repeats in the \textit{nlmC–comC} promoter region. The two direct repeats were also required for the promoter activity of both \textit{nlmC} and \textit{comC}. These results suggest that gene regulation of \textit{comC} in \textit{S. mutans} is fundamentally different from that reported for \textit{S. pneumoniae}, which implicates a unique regulatory mechanism that allows the coordination of bacteriocin production with competence development.

\section*{INTRODUCTION}

\textit{Streptococcus mutans} is considered to be the primary pathogen causing dental caries (tooth decay). In addition to the known virulence factors of \textit{S. mutans}, such as biofilm formation, acid production/tolerance and competence development, the ability to produce bacteriocin (mutacin) is thought to play an important role in the establishment of \textit{S. mutans} in the dental biofilm community (Kreth et al., 2005; Qi et al., 2001; Yonezawa & Kuramitsu, 2005). Currently, two types of mutacins have been characterized, the lantibiotics and the non-lantibiotics (Balakrishnan et al., 2000; Chen et al., 1998, 1999; Hillman et al., 1998; Mota-Meira et al., 1997; Qi et al., 1999a, b, 2000, 2001; Yonezawa & Kuramitsu, 2005). Lantibiotics are ribosomally synthesized and post-translationally modified peptides that contain lanthionine or \(\beta\)-methyl-lanthionine as well as dehydrated amino acids (McAuliffe et al., 2001; Twomey et al., 2002), whereas non-lantibiotics consist of either one or two small unmodified peptides (Balakrishnan et al., 2000; Qi et al., 2001).

Production of mutacin is initiated at the onset of stationary phase or under conditions of high cell density, allowing the mutacin producers to kill other competitive bacterial...
species when nutrients become limited. As with other stress responses, mutacin production is stringently regulated. Recent work in our laboratory has demonstrated that production of the lantibiotic mutacin I is regulated by multiple inputs (Tsang et al., 2005, 2006), while production of the non-lantibiotic mutacin IV is controlled by the ComED two-component system in response to the competence-stimulating peptide (CSP) (Kreth et al., 2005). Further studies demonstrated that mutacin IV production and competence development are temporally coordinated, with a 2 h delay in competence development following activation of mutacin IV gene expression by CSP (Kreth et al., 2005).

During our studies of ComED regulation of mutacin IV, we found that in addition to mutacin IV (nlmA), a group of mutacin or mutacin-like genes are also regulated by the ComED system (Kreth et al., 2006). Similar findings have also been reported previously (van der Ploeg, 2005). These CSP-responsive genes include nlmC (Smu1738/SMU. 1914C), encoding the newly identified mutacin V (Hale et al., 2005). More interestingly, the nlmC gene shares the intergenic region with the divergently transcribed comC. This finding raised a puzzling question: since the putative ComE binding site sequence (Kreth et al., 2006; van der Ploeg, 2005). More interestingly, the nlmC gene shares the intergenic region with the divergently transcribed comC. This finding raised a puzzling question: since the putative ComE binding site sequence (Kreth et al., 2006; van der Ploeg, 2005). More interestingly, the nlmC gene shares the intergenic region with the divergently transcribed comC. This finding raised a puzzling question: since the putative ComE binding site sequence (Kreth et al., 2006; van der Ploeg, 2005). More interestingly, the nlmC gene shares the intergenic region with the divergently transcribed comC. This finding raised a puzzling question: since the putative ComE binding site sequence (Kreth et al., 2006; van der Ploeg, 2005). More interestingly, the nlmC gene shares the intergenic region with the divergently transcribed comC. This finding raised a puzzling question: since the putative ComE binding site sequence (Kreth et al., 2006; van der Ploeg, 2005). More interestingly, the nlmC gene shares the intergenic region with the divergently transcribed comC. This finding raised a puzzling question: since the putative ComE binding site sequence (Kreth et al., 2006; van der Ploeg, 2005). More interestingly, the nlmC gene shares the intergenic region with the divergently transcribed comC. This finding raised a puzzling question: since the putative ComE binding site sequence (Kreth et al., 2006; van der Ploeg, 2005). More interestingly, the nlmC gene shares the intergenic region with the divergently transcribed comC. This finding raised a puzzling question: since the putative ComE binding site sequence (Kreth et al., 2006; van der Ploeg, 2005). More interestingly, the nlmC gene shares the intergenic region with the divergently transcribed comC. This finding raised a puzzling question: since the putative ComE binding site sequence (Kreth et al., 2006; van der Ploeg, 2005).

**METHODS**

**Bacterial strains and culture conditions.** Bacterial strains used in this study and their relevant characteristics and genotypes are listed in Table 1. All S. mutans derivatives were grown in Brain Heart Infusion (BHI, Difco) or on BHI agar plates. For the selection of antibiotic resistant clones, BHI plates were supplemented with either spectinomycin (Spc; 800 μg ml⁻¹) or erythromycin (Erm; 15 μg ml⁻¹). All S. mutans strains were grown anaerobically (90 % N₂/5 % CO₂/5 % H₂) at 37 °C. For cloning purposes and plasmid proliferation Escherichia coli cells were used. Cells were grown in LB medium supplemented with spectinomycin (100 μg ml⁻¹) or ampicillin (100 μg ml⁻¹) aerobically at 37 °C.

**DNA manipulation.** Standard procedures were used for plasmid isolation, DNA digestion and analysis and cloning. PCR was performed with a MyCycler thermocycler (Bio-Rad) using a protocol from the supplier.

**Construction of nlmC–luc reporter gene fusion.** The backbone vector for the construction of all reporter gene fusions was pFW5–luc, which contains a spectinomycin resistance marker (aad9) that works in both Gram-negative and Gram-positive bacteria (Podbielski et al., 1996). The promoter region of nlmC was amplified by PCR from the chromosomal DNA of strain UA159. Primers used for the amplification incorporated restriction enzyme sites for BanHI as well as Sall and are listed in Table 2. The promoter region was then cloned into pFW5–luc to generate the plasmid pFW5 : : Φ(nlmCp–luc). The plasmid was then transformed into S. mutans strain UA159 and integrated into the chromosome via single crossover recombination. Transformants were selected on BHI+Spc plates and confirmed by PCR and reporter gene activity.

**Construction of comE and comD mutants.** The comE and comD defective strains were constructed by transferring chromosomal DNA from previously constructed mutant strains (SMCE1 ComE⁻ ErmR and SMCD1 ComD⁻ ErmR (Li et al., 2001) into the nlmC–luc reporter strain. Transformants were confirmed by PCR.

**Construction of direct repeat I (DR I) and direct repeat II (DR II) deletions.** Deletion of the DR I and DR II regions of the nlmC promoter was performed by inverse PCR using pFW5 : : Φ(nlmCp–luc) as the template and primers Delta-BS-F and Delta-BS-R. Inverse PCR with Delta-BS-F and Delta-BS-R would

**Table 1. Bacterial strains used in this study**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant characteristics</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli BL21(DE3)pLys</td>
<td>F⁻ ompT gal hsdSB (rB mB) dcm lon λDE3 pLysS</td>
<td>Studier et al. (1990)</td>
</tr>
<tr>
<td>S. mutans UA159</td>
<td>Wild-type SpcE ErmR</td>
<td>Ađić et al. (2002)</td>
</tr>
<tr>
<td>UA159E</td>
<td>comE ErmR</td>
<td>This work</td>
</tr>
<tr>
<td>UA159D</td>
<td>comD ErmR</td>
<td>This work</td>
</tr>
<tr>
<td>SMCD1</td>
<td>comD ErmR</td>
<td>Li et al. (2001)</td>
</tr>
<tr>
<td>SMCE-L1</td>
<td>comE ErmR</td>
<td>Li et al. (2001)</td>
</tr>
<tr>
<td>JKI738-159</td>
<td>UA159 : : Φ(nlmCp–luc) SpcR</td>
<td>This work</td>
</tr>
<tr>
<td>JKI738-159D</td>
<td>JKI738-159 comD ErmR</td>
<td>This work</td>
</tr>
<tr>
<td>JKC159</td>
<td>UA159 : : Φ(comCP–luc) SpcR</td>
<td>This work</td>
</tr>
<tr>
<td>JK-DR1738</td>
<td>UA159 : : Φ(nlmCcDrB+II–luc) SpcR deletion of DRI+II in nlmCp</td>
<td>This work</td>
</tr>
<tr>
<td>JK-DRC</td>
<td>UA159 : : Φ(comCPccDrB+II–luc) SpcR deletion of DRI+II in comCp</td>
<td>This work</td>
</tr>
</tbody>
</table>
create a deletion of DR II, and with Delta-BS-F and Delta-BS-R1 would create a deletion of both DR I and DR II (Fig. 1a). The PCR fragments were then religated and transformed into E. coli. The correct plasmid constructs [pFW5 : : Φ(nlmC<sub>DR II</sub>luc)159 and pFW5 : : Φ(nlmC<sub>DR I+II</sub>luc)159] were confirmed by restriction enzyme digestion followed by DNA sequencing. The plasmids were then transformed into UA159 and integrated into the nlmC promoter region via single crossover recombination.

### Table 2. Primers used in this study

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
<th>Gene targeted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1738BamHI</td>
<td>5'-GGATCCCACCTTCAACAGCTGAAAGTGC-3'</td>
<td>nlmC promoter region</td>
</tr>
<tr>
<td>1738SalI</td>
<td>5'-GTCGACTAAAAACTCTGGTAAAACCGCCG-3'</td>
<td>nlmC promoter region</td>
</tr>
<tr>
<td>1738-RT-F</td>
<td>5'-ATGGAATATGAACGCTTACGC-3'</td>
<td>RT-PCR nlmC</td>
</tr>
<tr>
<td>1738-RT-R</td>
<td>5'-ATAAATTTCGCCAGCAACCTTA-3'</td>
<td>RT-PCR nlmC</td>
</tr>
<tr>
<td>comC-RT-F</td>
<td>5'-GACTGATGAATTAGAGATATCTACCTTG-3'</td>
<td>RT-PCR comC</td>
</tr>
<tr>
<td>comC-RT-R</td>
<td>5'-TTTCCGAAACACTGTTAAAACCTT-3'</td>
<td>RT-PCR comC</td>
</tr>
<tr>
<td>ComE inv</td>
<td>5'-GGTGGATTGCACAGGCTG-3'</td>
<td>comE primer</td>
</tr>
<tr>
<td>Delta-BS-F</td>
<td>5'-ATATTTCTGCCATTGAAATAATTTG-3'</td>
<td>DR deletion</td>
</tr>
<tr>
<td>Delta-BS-R</td>
<td>5'-GATATTTCTGCTAATGTATATTACG-3'</td>
<td>DR deletion</td>
</tr>
<tr>
<td>Delta-BS-R1</td>
<td>5'-GATATTTCTGCTAATGTATATTACG-3'</td>
<td>DR deletion</td>
</tr>
<tr>
<td>1738-comC-B</td>
<td>5'-GGATCCTAAAACCTCCTGTTAAACAGCCCG-3'</td>
<td>promoter flip nlmC to comC</td>
</tr>
<tr>
<td>1738-comC-S</td>
<td>5'-GTCGACCACCTCCTAAGAGGCTAAG-3'</td>
<td>promoter flip nlmC to comC</td>
</tr>
</tbody>
</table>

![Fig. 1. Intergenic region IGS1499. (a) Nucleotide sequence of the intergenic region (IGS1499) between nlmC (Smu1738) and comC, including the ATG codon for each gene. The top strand is the coding strand for comC, and the bottom strand is the coding strand for nlmC. The putative extended −10 and −35 regions for nlmC are highlighted with boxes. The two direct repeats (DR I and DR II) and their directions are indicated by arrows. Brackets denote the regions that were deleted in the DR II and the DR I+II deletion mutants. (b) Putative ComE binding site in the promoter region of comED (S. mutans) and putative ComE and BlpR binding sites in the promoter regions of comCDE and blpS (S. pneumoniae) and their relative arrangement on the chromosome [modified after (Martin et al., 2006)]. The imperfect direct repeats are underlined in all three sequences and separated by a 10 bp spacer (11 bp in comCDE in S. pneumoniae). The light grey box indicates the location of the putative ComE or BlpR binding sites. Sm, S. mutans; Sp, S. pneumoniae.](image-url)
To create comC-luc fusion lacking the same direct repeats, the promoter fragment in pFW5 : : \( \Phi (nlicerC_{ADRI}+_{11}-luc)159 \) was amplified with the primers 1738comC-B and 1738comC-S, which generated the BamHI and SalI restriction sites for reversal of the promoter region, thus creating the plasmid pFW5 : : \( \Phi (comC_{ADRI}+_{11}-luc)159 \). Plasmid pFW5 : : \( \Phi (comC_{ADRI}+_{11}-luc)159 \) was subsequently integrated into the chromosome of UA159 to create a comC-luc reporter strain lacking the two direct repeats in the upstream region.

**Luciferase assay.** Luciferase assays were performed using a method similar to one described previously (Loimaranta et al., 1998). Briefly, 25 μl 1 mM d-luciferin (Sigma) suspended in 100 mM citrate buffer, pH 6, was added to 100 μl cell culture. To ensure sufficient levels of intracellular ATP, cells were recharged with 1 % glucose for 10 min prior to luciferin addition. Luciferase activity was measured using a TD 20/20 luminometer (Turner Biosystems). Usually, three parallel cultures were measured at each time point and the mean value was taken. Each experiment was repeated at least two times.

**RNA extraction and real-time RT-PCR.** Overnight cultures of UA159 and derivatives were diluted 1 : 40 in 200 ml BHI (no antibiotics) and grown for 2–3 h until OD\textsubscript{600} reached approx. 0.1. If CSP induction was required, synthetic CSP (1 μg ml\textsuperscript{-1}) was added to the culture (Kreth et al., 2005). The cells were further incubated for 2 h until OD\textsubscript{600} was between 0.5 and 0.6, a time point where comC induction is at the maximum. Cells were harvested by centrifugation and stored at \(-80^\circ\text{C}\) until needed. For RNA extraction, cDNA synthesis and real-time RT-PCR a procedure described previously (Merritt et al., 2005) was followed, using the iCycler (Bio-Rad) real-time PCR System. Total cDNA abundance between test samples was normalized using the 16S RNA gene as a housekeeping control.

**ComE purification and electrophoretic mobility shift assays (EMSA).** The comE coding region was PCR amplified from S. mutans strain UA159, and cloned into pCFT7 (NT Topo vector, Invitrogen) with a 6-histidine tag at the N terminus. The vector was transformed into E. coli strain BL21(DE3)pLys. ComE protein was overexpressed by induction with IPTG at a final concentration of 1 mM. Cells were harvested and lysed with lysosome (8 mg ml\textsuperscript{-1}), followed by sonication (Branson sonifier 450, WVR). Cell lysates were centrifuged at 7800 g at 4 °C for 30 min. For ComE protein purification, cleared cell lysate was mixed with Pro-Buffer resin (Invitrogen) and shaken at 4 °C for 1 h. The resin was allowed to settle by gravity to separate the bound ComE and the non-specific proteins in the buffer. The non-specific proteins were secured for later analysis by PAGE. The column was washed four times with 8 ml native washing buffer [native binding buffer (50 mM NaH\textsubscript{2}PO\textsubscript{4}, 500 mM NaCl, 10 mM Tris/HCl, pH 18, 10 %, v/v, glycerol, 5 mM dithiothreitol)] + 20 mM imidazole]. Purified protein was eluted with 10 ml native elution buffer (native binding buffer + 250 mM imidazole). Eluted fractions (1 ml each) were collected. Each eluted fraction had glycerol added to 20 % (v/v) and was quick-frozen in dry ice/methanol before storage at \(-80^\circ\text{C}\). All the fractions were saved and analysed; by running 4–20 % Tris-glycine gel (Invitrogen) at 150 V, 1.5 h, for ComE purity. The ComE concentration was determined from a standard curve with BSA standards. Each gel was stained with colloidal blue staining kit (Invitrogen).

For radioactive probe labelling, 1 μM oligonucleotide probe oSG316 (CCCATTTTTAGTTTTTGTCTG) was labelled at the 5'-end with 0.85 μM \( [\text{32P}] \text{ATP} \) (10 μCi ml\textsuperscript{-1}, 370 MBq ml\textsuperscript{-1}; New Life Science Products) by using T4 polynucleotide kinase (Promega). The labelled primer was used with primer oSG317 (GAAAATACATGGAATTTTGCTTG) in a PCR to amplify promoter regions of comC wild-type (204 bp) \( \text{comC}_{ADRI} \) (192 bp) and \( \text{comC}_{ADRI} \) + II (166 bp). For EMSA, ComE (10 nM) was incubated at room temperature for 30 min in ComE dilution buffer (50 mM HEPES, pH 6.5, 1 mM EDTA, 10 % glycerol), 5 × EMSA buffer (45 % glycerol, 250 mM HEPES, pH 6.5, 250 μg BSA ml\textsuperscript{-1}), with varying concentrations of salmon sperm DNA and 1 nM isotopically labelled DNA substrate. Following incubation, EMSA reactions were analysed on 6 % non-denaturing polyacrylamide gels in electrophoresis buffer (0.5 % TBE) that was pre-run for at least 2 h at 150 V constant voltage. Nucleoprotein complexes were separated by PAGE at 150 V for 3 h and subsequently dried and scanned with the Bio-Rad imaging system.

**RESULTS**

**ComE is required for nlmC transcription but represses comC gene expression**

As a first step towards understanding how the ComED system co-ordinates the temporal expression of nlmC with competence development, we analysed the comE-nlmC intergenic region, IGS1499, and found that it contained two 11 bp direct repeats (DR) separated by a 10 bp spacer (Fig. 1a). This region is localized immediately upstream of the predicted −35 region of nlmC and, like the ComE binding site for S. pneumoniae, is a strong match to the AlrG/AgrA/LyrR family binding sites (Fig. 1b) (Nikolskaya & Galperin, 2002). Since this sequence feature exists only on the DNA strand encoding the promoter of nlmC and not that of comC, we were curious as to whether ComED also regulated comC expression. To test this, real-time RT-PCR was performed in wild-type UA159. As shown in Fig. 2(a), comC gene expression increased ~5-fold in response to CSP. Similar results were also obtained with two other wild-type strains NG8 and UA140 (data not shown). These results were surprising given that the putative ComE binding sequence was not located in the comC promoter.

To further investigate, we performed a similar experiment to measure comC gene expression in a comE mutant background in the presence and absence of CSP. To our surprise, comC expression increased approximately fivefold in the comE mutant background as compared to the uninduced wild-type (Fig. 2a). This was the same level of comC induction observed in response to CSP addition in the wild-type. In contrast, the expression of nlmC was nearly abolished in the comE mutant background (Fig. 2b). This result suggested that in the absence of CSP induction ComE acts as a represor of comC gene expression, whereas ComE is required for nlmC transcription. As a further verification of the repressor function of comC, we measured nlmC and comC expression at growth stages in which comE expression was either high or low. The expression pattern of comC was measured by real-time RT-PCR in competence development medium in the absence of exogenously added CSP. As shown in Fig. 2(c), comC gene expression was higher in mid-exponential phase than in late exponential phase, whereas nlmC gene expression was much higher in late exponential phase than in mid-exponential phase. The most dramatic difference between the two promoters was observed in overnight cultures. At
this stage, *comC* gene expression was three times that achieved in mid-exponential phase; in contrast, *nlmC* gene expression dropped to nearly undetectable levels (Fig. 2d). These results further supported a bifunctional role for ComE that could simultaneously activate *nlmC* and repress *comC*.

**Deletion of comD also has an opposing effect on nlmC and comC gene expression**

According to the model of how the ComED two-component system operates, binding of CSP to ComD triggers a phosphorylation cascade comprising autophosphorylation of ComD then phosphorylation of ComE. Phosphorylated ComE then activates or represses its target genes (Alloing et al., 1998; Morrison & Lee, 2000). Based on this model, we predicted that in a *comD* mutant background *comC* expression would be unresponsive to addition of CSP, while the wild-type would exhibit the expected increase in *comC* expression. A similar expression pattern was predicted for *nlmC*. As shown in Fig. 3(a), *comC* expression in the *comD* mutant was similar to that in the wild-type in the absence of CSP, but was reduced >2-fold in the presence of CSP. We speculate that the slightly higher *comC* expression in the *comD* mutant was similar to that in the wild-type in the absence of CSP, but was reduced >2-fold in the presence of CSP. We speculate that the slightly higher *comC* expression in the *comD* mutant was similar to that in the wild-type in the absence of CSP, but was reduced >2-fold in the presence of CSP. We speculate that the slightly higher *comC* expression in the *comD* mutant was similar to that in the wild-type in the absence of CSP, but was reduced >2-fold in the presence of CSP. We speculate that the slightly higher *comC* expression in the *comD* mutant was similar to that in the wild-type in the absence of CSP, but was reduced >2-fold in the presence of CSP. We speculate that the slightly higher *comC* expression in the *comD* mutant was similar to that in the wild-type in the absence of CSP, but was reduced >2-fold in the presence of CSP. We speculate that the slightly higher *comC* expression in the *comD* mutant was similar to that in the wild-type in the absence of CSP, but was reduced >2-fold in the presence of CSP. We speculate that the slightly higher *comC* expression in the *comD* mutant was similar to that in the wild-type in the absence of CSP, but was reduced >2-fold in the presence of CSP.
in Fig. 3(b), nlmC expression in the comD mutant was reduced >40-fold compared with the wild-type in the presence of CSP. In the comD mutant, the >4-fold reduction of nlmC gene expression relative to the uninduced wild-type was probably caused by the same decrease in comE gene expression mentioned above (Fig. 3b). Taken together, these results further support the notion that ComE is a repressor of comC expression but is required for nlmC gene expression.

The direct repeats in the nlmC–comC promoter region are required for ComE binding

Based on our initial results, we hypothesized that the two direct repeats, DR I and DRII (Fig. 1), in the nlmC–comC intergenic region would be required for ComE binding in S. mutans. To test this, the ComE protein was purified from E. coli and tested for binding to the nlmC–comC intergenic region in an EMSA. A 204 bp DNA fragment encompassing a region 17 bp upstream of the nlmC and 46 bp upstream of the comC initiation codons was used as the substrate for ComE binding. As shown in Fig. 4, ComE bound to the wild-type DNA in a concentration-dependent manner. At a molar ratio of 50 : 1 ComE/DNA (5 nM : 0.1 nM), there were two shifted bands, with the lower band more abundant than the upper band (lane 2). When the ComE concentration was increased to 15 nM, there was a conspicuous increase in the abundance of the upper band (lane 3). These results suggested that the two direct repeats may serve as two binding sites for ComE. To further test this, we constructed two deletion mutations of the direct repeat region; one contained a complete deletion of DR II (ΔDRII, see Fig. 1), and the other contained a complete deletion of both DRI and DRII (ΔDRI+II, see Fig. 1). DNA was PCR amplified from both mutants, and EMSA was performed under the same conditions as for the wild-type. As shown in Fig. 4, deletion of DRII resulted in the generation of only the lower shifted band with both concentrations of ComE (lanes 5 and 6). This result suggested that this deletion mutant provided only one binding site. As expected, deletion of both direct repeats abolished ComE binding completely, as shown by the unshifted bands (lanes 7–9). These results demonstrate that the two direct repeats in the nlmC–comC promoter region are required for ComE binding, although determining the molecular mechanism of how ComE binds to the two direct repeats will require further experiments.

Fig. 3. Effect of a comD mutation on comC and nlmC gene expression. (a) Expression of comC in wild-type and comD mutant backgrounds with and without CSP. Transcript levels were measured by real-time RT-PCR using 16S RNA as a housekeeping control. cDNA abundance was normalized against the 16S cDNA and expressed as relative cDNA abundance. The wild-type level without CSP was arbitrarily set as 1. (b) nlmC expression in wild-type and comD mutant backgrounds with and without CSP. Transcript levels were measured using the nlmC–luc reporter and normalized against cell density. Values and standard deviation are calculated from at least two independent experiments performed in triplicate.

Fig. 4. EMSA of the ComE protein with comC wild-type, comCADRIII and comCADRI+II DNA fragments. EMSA analysis was carried out using purified ComE protein with 1 nM radiolabelled DNA probe. The reactions were run on a 6% non-denaturing polyacrylamide gel and the image was scanned with a Bio-Rad imaging machine. The experiments were repeated three times with similar results. A representative gel image is shown.
The direct repeats in the intergenic region are also required for \textit{nlmC} and \textit{comC} gene expression

The \textit{comE} mutation data presented in Fig. 2 demonstrated that an intact ComE is a repressor for \textit{comC}, but an activator for \textit{nlmC} expression. The data presented in Fig. 4 demonstrated that the two direct repeats in the \textit{nlmC}–\textit{comC} promoter region are required for ComE binding. Based on these findings, we predicted that deletion of the two direct repeats would diminish \textit{nlmC} gene expression but derepress \textit{comC} expression. To test this, luciferase reporter fusions lacking the two direct repeats were constructed for both \textit{nlmC} and \textit{comC}. Luciferase activity was measured using the wild-type as a control. Cells were grown to late exponential phase, when expression levels of both promoters can be measured and compared in the absence of CSP. As shown in Fig. 5, deletion of DR I+II reduced \textit{nlmC} gene expression ~20-fold as was expected (Fig. 5a). Surprisingly, deletion of the two direct repeats nearly completely abolished \textit{comC} gene expression (Fig. 5b). These results were rather puzzling since the two direct repeats are located 130 bp upstream of the initiation codon for ComC. To find a possible explanation for this unexpected result, we tried to localize the \textit{comC} transcriptional start site by RACE PCR and RT-PCR. Despite multiple attempts, these experiments have so far yielded ambiguous results. It appeared that none of the predicted promoter sequences served as classical promoters for \textit{comC}; rather, the \textit{comC} transcript appeared to start further upstream, even beyond the promoter region for \textit{nlmC} (J. Merritt & F. Qi, unpublished results). Therefore, a mechanistic explanation of how ComE regulates \textit{comC} gene expression would require additional research.

\textbf{DISCUSSION}

The objective of this study was to determine how ComE co-ordinates the temporal gene expression of both \textit{nlmC} and \textit{comC}. Both luciferase gene fusion and real-time RT-PCR were used to characterize promoter activities of \textit{nlmC} and \textit{comC} in different mutant backgrounds. EMSAs were used to analyse ComE binding to the wild-type and DR deletion mutants of the \textit{nlmC}–\textit{comC} intergenic region. We demonstrated that ComE is required for both activation of \textit{nlmC} transcription and repression of \textit{comC} gene expression. Furthermore, we showed that the two 11 bp direct repeats in the upstream regions of \textit{nlmC} and \textit{comC} are required for ComE binding to the promoter region, suggesting that the two direct repeats may serve as the binding sites for ComE.

The strongest evidence supporting the bifunctionality of ComE in regulating \textit{nlmC} and \textit{comC} gene expression came from the mutagenesis studies of \textit{comE}. In a \textit{comE} null mutant background, \textit{nlmC} gene expression was nearly abolished, whilst \textit{comC} gene expression was increased to the same level as in CSP-induced wild-type cells (Fig. 2). Further support for ComE bifunctionality came from the expression pattern analyses of \textit{nlmC} and \textit{comC}. In overnight cultures, when ComE was presumably absent, \textit{comC} gene expression displayed its strongest induction; in contrast, \textit{nlmC} gene expression was reduced to below detectable levels. In addition, the differential response of \textit{nlmC} and \textit{comC} to CSP induction could also add further support to the opposing regulatory mechanisms for \textit{nlmC} and \textit{comC} gene expression via ComE. For example, the maximal induction for \textit{nlmC} is usually between 50- and 100-fold (Fig. 2b), whereas the maximal induction for \textit{comC} is usually between 3- and 5-fold (Fig. 2a). In addition to the potentially dissimilar promoter strengths, an additional factor in the different levels of CSP induction could result from the positive feedback loop regulation of \textit{nlmC} gene expression vs the derepressive regulation of \textit{comC} gene expression. Unlike \textit{nlmC} induction, the derepression of \textit{comC} in response to CSP could not be further enhanced by autoinduction of the \textit{comED} operon.

As a further support for the direct role of ComE in \textit{nlmC} and \textit{comC} gene regulation, purified ComE was shown to bind to the \textit{nlmC}–\textit{comC} intergenic region. This binding was also dependent upon the presence of two direct repeats, which are reminiscent of the \textit{S. pneumoniae} ComE binding site. In the presence of both direct repeats, there are two shifted bands, presumably reflecting binding of ComE to either one or both direct repeats (Fig. 4). Deletion of one direct repeat resulted in only one shifted band, whilst deletion of both direct repeats abolished ComE binding completely. These results suggest that ComE can bind to either direct repeat; however, whether the presence of both direct repeats promotes cooperative binding is not known.

The direct involvement of the two direct repeats in gene expression of \textit{nlmC} and \textit{comC} was tested by deletion

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Effects of the DRI+II deletion on \textit{nlmC} and \textit{comC} gene expression. (a) Luciferase gene expression driven by the wild-type and the DR I+II-deleted \textit{nlmC} promoter. Cells were harvested at late exponential phase. Note the log scale of the y-axis. (b) Luciferase gene expression driven by the wild-type and the DR I+II-deleted \textit{comC} promoter. Cells were harvested at the late exponential phase. All values are the mean of at least two independent experiments done in triplicate.}
\end{figure}
mutagenesis. Deletion of the two direct repeats diminished nlmC gene expression (Fig. 5). This is consistent with previous findings (van der Ploeg, 2005), which showed that point mutations made in either of the direct repeats severely impaired nlmA gene expression [nlmA has a near-identical promoter region to nlmC (Kreth et al., 2006)]. Unexpectedly, the same direct repeat deletion also abolished comC gene expression. This result is rather puzzling as the two direct repeats exist only on the DNA strand encoding the promoter of nlmC, not that encoding the comC promoter. Two explanations may be entertained at this point. (1) The promoter region of comC may overlap the direct repeats region; thus deletion of the direct repeats may have abolished RNA polymerase binding to the comC promoter. (2) The comC transcript may start further upstream of the direct repeat region and the direct repeat region is required for RNA stability of the comC transcript. Since real-time RT-PCR measures RNA at steady state, rapid degradation of the comC transcript as a result of the DR deletion may have resulted in the nearly undetectable levels of comC–luc gene expression as observed in Fig. 5. Although the detailed mechanisms of how ComE regulates comC gene expression await further research, our data appear to suggest that the mechanism of comC transcription is more complicated than is currently accepted.

Taken together, this study demonstrates that ComE plays a very different regulatory role in S. mutans from that of its homologue in S. pneumoniae. Interestingly, both of the C-termini of the ComE proteins of S. pneumoniae and S. mutans share high homology to the DNA binding domains of AlgR/AgrA/LytR family members. In addition, both proteins bind to direct repeat sequences that match extremely well to the predicted binding sites for the AlgR/AgrA/LytR family members (Nikolskaya & Galperin, 2002). However, despite this similarity, S. mutans has evolved a repressive mode of regulation for comC and a delayed competence response to the addition of CSP. This suggests that the mechanism of competence regulation in S. mutans may be quite distinct from that in S. pneumoniae.

ACKNOWLEDGEMENTS

This work was supported in part by NIH grant R01-DE014757 to F. Q., NIDCR T32 training grant DE007296 to J. M. and a Delta Dental grant WDS78956 to W. S.

REFERENCES


---

Edited by: R. J. Lamont