# SCIENTIFIC REPORTS

**OPEN**

# A positive feedback loop mediated by Sigma X enhances expression of the streptococcal regulator ComR

Rabia Khan[1], Roger Junges[1], Heidi A. Åmdal[1], Tsute Chen[2], Donald A. Morrison[3] & Fernanda C. Petersen [1]

Natural transformation is used by bacteria to take up DNA from their surroundings and incorporate it into their genomes. Streptococci do so during a transient period of competence, triggered by pheromones that they produce, secrete and sense under conditions influenced by the environment. In *Streptococcus mutans*, *Streptococcus suis*, and species of the bovis, salivarius and pyogenic groups of streptococci, the pheromone XIP is sensed by the intra-cellular regulator ComR, that in turn activates the transcription of *comS*, encoding the XIP precursor, and of *sigX*, encoding the only known alternative sigma factor in streptococci. Although induction of *comR* during competence has been known for more than fifteen years, the mechanism regulating its expression remains unidentified. By a combination of directional RNA-sequencing, optimal competence conditions, stepwise deletions and marker-less genome editing, we found that SigX is the missing link in overproduction of ComR. In the absence of *comR* induction, both *sigX* expression and transformation were significantly reduced. Placing *comR* and *comS* transcripts under the control of different regulators so as to form two interlocked positive feedback circuits may enable *S. mutans* to fine-tune the kinetics and magnitude of the competence response according to their need.

Natural competence for genetic transformation is a survival strategy that enables streptococci to attack their siblings, take up the released DNA or other naked DNA in their surroundings, and incorporate it into their genomes. Most of the effector genes are conserved in streptococci, and are under the control of the alternative sigma factor SigX, the master regulator of competence[1–4]. They comprise a set of 27 to 30 genes grouped into the core SigX regulon, showing a predominance of transformation functions, and accessory genes that depend on SigX for expression in species-specific patterns. Expression of SigX in a streptococcal population is triggered by communication signals that are peptide pheromones, produced as precursor molecules. The peptides are later processed and exported, and then sensed either by surface receptors of two-component regulatory systems or by intra-cellular regulators[5–9].

Surface receptors involved in streptococcal competence that belong to the ComCDE class are found in *S. pneumoniae* and other members of the mitis and anginosus groups of streptococci[5, 10]. In this class, the competence stimulating peptide (CSP) pheromone encoded by *comC* binds to the membrane histidine kinase ComD, resulting in auto-phosphorylation and activation of the cognate response regulator ComE. The modified ComE then activates the promoters of both *sigX* and *comCDE*, inducing their transcription. In this way, the pheromone not only triggers the expression of the master regulator, but creates a positive feedback loop that strongly amplifies the signal.

The second type of competence pheromone system belongs to the ComRS class, found in *Streptococcus mutans*, *Streptococcus suis*, and in the salivarius, bovis, and pyogenic groups of streptococci[1, 6, 7, 11–14]. The pheromone in this class, a peptide encoded by *comS* known as XIP (SigX-inducing peptide), is imported by oligopeptide permeases. Once inside the cells the pheromone binds to and activates the intracellular regulator ComR. The ComRS complex, in analogy to the activated ComE regulator described above, activates the promoters of *sigX* and of the pheromone structural gene, *comS*. However, in contrast to the ComCDE class, which amplifies the signal by inducing the simultaneous expression of the pheromone and its regulator, in most of the ComRS class

[1]Department of Oral Biology, Faculty of Dentistry, University of Oslo, Oslo, Norway. [2]Department of Microbiology, The Forsyth Institute, Cambridge, USA. [3]Department of Biological Sciences, College of Liberal Arts and Sciences, University of Illinois at Chicago, Chicago, USA. Correspondence and requests for materials should be addressed to F.C.P. (email: f.c.petersen@odont.uio.no)
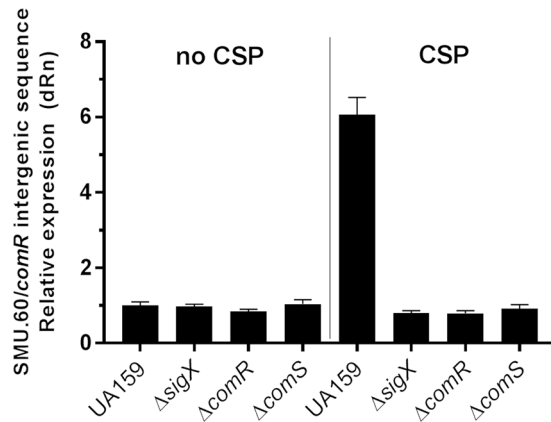
**Figure 1.** Up-regulation of the intergenic region between SMU.60 and *comR* is a late competence response. Relative expression of the region located between SMU.60 and *comR* in the *S. mutans* wild type UA159 and in the isogenic mutants SM061 (Δ*sigX*), SM066 (Δ*comR*), and SM089 (Δ*comS*). The strains were grown in TSB in the presence or absence of CSP. The real-time PCR data was normalized to the expression values in UA159 with no CSP. The results show mean values and standard errors for 3 replicates.

only the pheromone expression is induced. Yet, overexpression of *comR* in *S. mutans* has a strong positive effect on transformation efficiency[7], while in the salivarius group, low levels of *comR* expression are associated with low transformation efficiency in a range of wild-type strains[15]. Thus, although mechanisms for *comR* regulation remain unknown, accumulated evidence indicates that control of *comR* expression would provide an additional way for streptococci to fine tune the competence response.

Variable levels of *comR* induction are actually observed among the genes up-regulated during *S. mutans* competence[16–20]. Interestingly, the 3′ portion of the gene immediately upstream of *comR* was recently found to be up-regulated upon stimulation of competence, possibly by SigX[2]. We hypothesized that this region may have a role in competence, mediated by either a putative peptide annotated as smut_c_1_105 at the Human Oral Microbiome Database (HOMD; http://www.homd.org), or by acting as an intragenic start site for transcription of *comR*. Using a combination of directional RNA-sequencing, optimal competence conditions, stepwise deletions and direct genome editing of the putative peptide gene and of the SigX-box upstream of *comR*, the present study shows that SigX, the master regulator of competence, is the missing link in overproduction of ComR during development of competence in *S. mutans*.

## Results

### Up-regulation of transcription of the intergenic region between SMU.60 and *comR* during competence is abolished in *comS*, *comR*, and *sigX* mutants.

A recent transcription study suggested that an apparent intragenic SigX-box in SMU.60, which is located upstream of *comR*, may drive expression of a transcript that extends to the intergenic region (IGR) between SMU.60 and *comR*[2]. To characterize transcripts extending upstream of *comR* more thoroughly, we performed real-time PCR with primers designed to amplify within the intergenic region between SMU.60 and *comR*. Using cultures optimized for CSP induction of competence[21], the results showed that this region is up-regulated approximately 6-fold in UA159 (Fig. 1). Because ComRS is the complex that induces *sigX* transcription, and a predicted SigX box is located within the SMU.60 ORF, we repeated the analysis for mutants deleted for the *comS*, *comR*, or *sigX* genes; induction of transcription of the IGR was abolished in every case (Fig. 1). The requirement for all three genes for this transcriptional response suggested that SigX may be the proximal regulator of the transcript initiated within SMU.60.

### RNA-seq analysis maps the start of the induced *comR* transcript to an intragenic region in SMU.60 and suggests that it extends to the full length of *comR*.

To obtain a precise map of the transcripts induced by XIP, including the SMU.60/*comR*/*comS* region, we adopted an optimized protocol for *S. mutans* competence development that includes optimal exposure time, dilution rates, and growth in CDM, as previously described[21]. To reduce possible influences from endogenous XIP production, the experiment used a *comS* deletion mutant. The comparisons revealed robust up-regulation of all the 30 core genes of the pan-streptococcal SigX regulon by XIP (Fig. 2 and Fig. 3). The transcriptome map showing the alignment of all transcripts to the reference sequenced genome of UA159 can be visualized at the Microbial Transcriptome Database (MTD; http://bioinformatics.forsyth.org/mtd/?name=RNAseq_smu065&toc=1) and fold changes for the genes in the entire genome are presented in Supplementary Table S1. The transcriptome map obtained with the *comS* mutant allowed us to investigate the up-regulation of the SMU.60-*comR* region at single-base resolution. We found that induced transcription of this locus was indeed initiated within and in the same direction as SMU.60, and that the ~nine-fold elevated transcription included the full length of *comR* (Fig. 4), whereas there was little or no transcription from the complementary strand.

All the genes of the ComE, ComR, and SigX regulons previously mapped in a comprehensive study of the competence responses in *S. mutans*[2] were identified here (Fig. 2). The surplus of genes that were up-regulated in the present study, except for the locus between SMU.1332 to 1340, were almost all co-transcribed with the genes

| Gene ID | Strand[a] | R, X, or E[c] | Change[d] | Annotation |
|---|---|---|---|---|
| SMU.29 | + | | −4,1 | SAICAR synthase |
| SMU.60[a] | + | (X) | 4,2 | DNA alkylation repair protein |
| SMU.61 | + | | 8,5 | transcriptional regulator; ComR |
| IGS comS | + | R | − | (comS deletion) |
| SMU.63c | − | | −36,9 | Hypothetical protein |
| SMU.64 | + | | 11,3 | Holliday junction DNA helicase RuvB |
| SMU.65 | + | | 9,2 | protein tyrosine-phosphatase |
| SMU.66 | + | | 4,1 | hypothetical protein |
| SMU.109 | + | | 6,2 | permease |
| SMU.150 | + | E | 22,0 | hypothetical protein; NlmA; mutacin IV A |
| SMU.151 | + | | 26,2 | hypothetical protein; NlmB; mutacin IV B |
| SMU.152 | + | | 34,0 | hypothetical protein; mutacin IV immunity |
| SMU.153 | + | | 32,7 | hypothetical protein |
| SMU.166 | + | | 17,4 | hypothetical protein |
| SMU.167 | + | | 18,7 | hypothetical protein |
| SMU.168 | + | | 17,9 | transcriptional regulator |
| SMU.325[b] | + | X | 5,6 | deoxyuridine 5′-triphosphate nucleotidohydrolase |
| SMU.326 | + | | 5,6 | hypothetical protein |
| SMU.327[b] | + | | 5,6 | DNA repair protein RadA |
| SMU.352 | + | X | 16,7 | ribulose-phosphate 3-epimerase |
| SMU.353 | + | | 15,4 | hypothetical protein |
| SMU.354[b] | + | | 15,2 | hypothetical protein; ccs50 |
| SMU.355[b] | + | | 16,5 | CMP-binding factor |
| SMU.356 | + | | 17,4 | pur operon repressor |
| SMU.399 | + | | 5,5 | hypothetical protein |
| SMU.400 | + | | 8,2 | secreted esterase |
| SMU.423 | + | E | 31,4 | hypothetical protein; NlmD, mutacin VI |
| SMU.498[b] | + | X | 527,1 | late competence protein; CflA, ComFA |
| SMU.499[b] | + | | 751,2 | late competence protein; CflB, ComFB |
| SMU.500[b] | + | | 12,5 | ribosome-associated protein; YfIA |
| SMU.505[b] | + | X | 65,9 | adenine-specific DNA methylase |
| SMU.506[b] | + | | 45,9 | type II restriction endonuclease |
| SMU.507 | + | | 28,1 | DeoR family transcriptional regulator |
| SMU.508 | + | | 26,3 | hypothetical protein |
| SMU.530c | − | | 4,2 | hypothetical protein |
| SMU.531 | − | | −8,7 | Chorismate mutase |
| SMU.532 | + | | −13,6 | Anthranilate synthase component I |
| SMU.533 | + | | −12,6 | Anthranilate synthase component II |
| SMU.534 | + | | −15,8 | Anthranilate phosphoribosyltransferase |
| SMU.535 | + | | −27,9 | indole-3-glycerol phosphate synthase |
| SMU.536 | + | | −33,4 | N-(5-phosphoribosyl)anthranilate isomerase |
| SMU.537 | + | | −9,6 | tryptophan synthase subunit beta |
| SMU.538 | + | | −6,5 | tryptophan synthase subunit alpha |
| SMU.539c[b] | − | X | 358,7 | signal peptidase type IV; CilC |
| SMU.625[b] | + | X | 280,0 | competence protein; CilE, ComEA |
| SMU.626[b] | + | | 305,7 | competence protein; CilB, ComEC |
| SMU.627 | + | | 7,2 | hypothetical protein |
| SMU.628 | + | | 5,3 | DNA polymerase III subunit delta |
| SMU.644[b] | + | X | 202,7 | competence protein/transcription factor; CoiA |
| SMU.645[b] | + | | 61,6 | oligopeptidase |
| SMU.646 | + | | 48,0 | phosphatase |
| SMU.647 | + | | 9,5 | methyltransferase |
| SMU.769[b] | + | X | 111,8 | hypothetical protein |
| SMU.770c | − | | −5,6 | Manganese transporter |
| SMU.772 | + | | 22,8 | glucan-binding protein; BglB-like protein |
| SMU.836[b] | + | X | 172,4 | hypothetical protein; CHAP domain; LytF |
| SMU.837 | + | | 82,5 | reductase |
| SMU.838 | + | | 29,3 | glutathione reductase |
| SMU.925 | + | E | 6,0 | hypothetical protein; immB |
| SMU.1001[b] | + | X | 533,7 | DNA processing Smf protein; DprA |
| SMU.1002 | + | | 41,8 | DNA topoisomerase I |
| SMU.1003 | + | | 30,0 | tRNA (uracil-5-)-methyltransferase Gid |
| SMU.1045d | − | | 4,1 | inorganic polyphosphate/ATP-NAD kinase |
| SMU.1046c | − | | 4,8 | GTP pyrophosphokinase |
| SMU.1047c | − | | 5,5 | hypothetical protein |
| (Readthrough) | | | | |
| SMU.1050 | + | | −4,1 | ribose-phosphate pyrophosphokinase |
| SMU.1051 | + | | −4,3 | iron-sulfur cofactor synthesis protein; NifS family |
| SMU.1053 | + | | −6,2 | redox-sensing transcriptional repressor Rex |
| SMU.1054 | + | | −18,8 | glutamine amidotransferase |
| SMU.1055[b] | − | X | 55,1 | hypothetical protein; RadC |
| SMU.1334 | − | | 12,0 | phosphopantetheinyl transferase |
| SMU.1335c | − | | 8,3 | enoyl-ACP reductase |
| SMU.1336 | − | | 11,3 | hypothetical protein |
| SMU.1337c | − | | 17,1 | alpha/beta hydrolase |
| SMU.1338c | − | | 7,4 | permease |
| SMU.1339 | − | | 6,6 | bacitracin synthetase |
| SMU.1340 | − | | 4,2 | surfactin synthetase |
| SMU.1400c | − | X | 10,7 | hypothetical protein |
| SMU.1902c | − | | 5,2 | hypothetical protein; putative bacteriocin; BsmK |
| SMU.1903c | − | | 22,3 | hypothetical protein |
| SMU.1904c | − | | 35,2 | hypothetical protein |
| SMU.1905c | − | | 52,1 | bacteriocin secretion protein; BsmL |
| SMU.1906c | − | E | 37,5 | hypothetical protein; BsmB |
| SMU.1908c | − | | 41,7 | hypothetical protein |
| SMU.1909c | − | | 42,3 | hypothetical protein; immunity protein |
| SMU.1910c | − | | 42,3 | hypothetical protein |
| SMU.1912c | − | | 51,1 | hypothetical protein |
| SMU.1913c | − | | 40,1 | immunity protein, BlpL-like; ImmA |
| SMU.1914c | − | E | 34,7 | hypothetical protein; NlmC; ClpB; mutacin V |
| SMU.1916 | − | | 43,6 | histidine kinase of the competence regulon; ComD |
| SMU.1917 | − | X | 58,4 | response regulator of the competence regulon ComE |
| SMU.1956c | − | | 9,6 | hypothetical protein |
| SMU.1957 | − | | 8,8 | PTS mannose-specific transporter subunit IID |
| SMU.1958c | − | | 10,2 | PTS mannose-specific transporter subunit IIC |
| SMU.1960c | − | | 9,9 | PTS mannose-specific transporter subunit IIB |
| SMU.1961c | − | | 10,7 | PTS sugar-specific transporter subunit IIA |
| SMU.1963c | − | | 20,0 | sugar-binding periplasmic protein |
| SMU.1964c | − | | 27,1 | response regulator |
| SMU.1965c | − | | 32,1 | histidine kinase |
| SMU.1966c | − | | 61,7 | periplasmic sugar-binding protein |
| SMU.1967[b] | − | X | 233,1 | single-stranded DNA-binding protein; SsbB |
| SMU.1975c | − | | 4,9 | hypothetical protein |
| SMU.1976c | − | | 5,4 | hypothetical protein |
| SMU.1977c | − | | 4,9 | transcriptional regulator |
| SMU.1978[b] | − | | 24,4 | acetate kinase |
| SMU.1979c[b] | − | | 209,9 | hypothetical protein |
| SMU.1980c[b] | − | | 473,1 | hypothetical protein; ComGG; CglG |
| SMU.1981c[b] | − | | 469,8 | hypothetical protein; ComGE; CglE |
| SMU.1982c[b] | − | | 304,2 | hypothetical protein; ComGE; CglE |
| SMU.1983[b] | − | | 450,4 | competence protein ComYD; ComGD; CglD |
| SMU.1984[b] | − | | 262,3 | competence protein ComYC; ComGC; CglC |
| SMU.1985[b] | − | | 343,8 | ABC transporter ComYB; ComGB; CglB |
| SMU.1987[b] | − | X | 353,0 | ABC transporter ComYA; ComGA; CglA |
| SMU.1996 | − | | 4,4 | 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase |
| SMU.1997 | − | R | 45,9 | SigX, ComX1, regulator of competence-specific genes |
| SMU.2075c | − | | 4,1 | hypothetical protein |
| SMU.2076c | − | X | 23,7 | hypothetical protein |
| SMU.2085[b] | − | | 14,7 | recombinase A; RecA |
| SMU.2086[b] | − | X | 27,0 | competence damage-inducible protein A; CinA |

**Figure 2.** Genes differentially regulated in response to XIP in the *S. mutans comS* deletion mutant (>4 fold; $p < 0.05$; > 8000 mean reads). Genes belonging to the same transcript are grouped together. Induced genes by XIP that are specific to the early (red) and late (blue) responses to CSP are indicated by the corresponding colors[2]. Downregulated genes (green). [a]SMU.60 was up-regulated only in its 3′-terminal region preceding *comR*. [b]Pan-streptococcal genes belonging to the SigX regulon. [c]First genes in the transcripts regulated by ComR (R), SigX (X), and ComE (E). [d]Mean fold change in transcript expression from 3 independent biological experiments comparing 120 min XIP treated to untreated parallel cultures. [e]Plus and minus strand for the ORF are presented, since some of the annotated genes in UA159 (accession AE014133) in the complementary strand (- strand) do not have the "c" designation after the gene number.
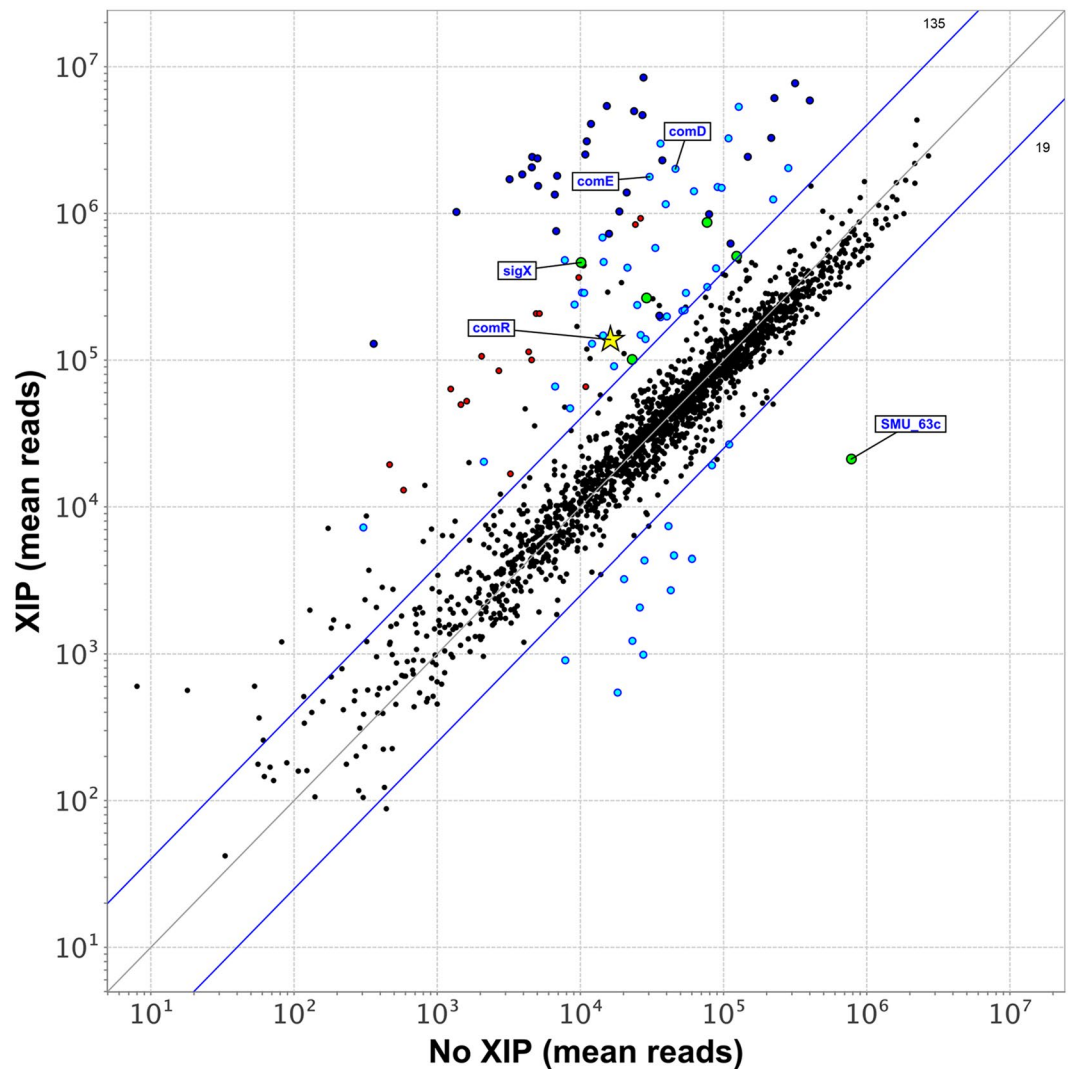
**Figure 3.** Transcriptome scatter-plot comparing *S. mutans* grown in the presence or absence of the XIP pheromone in CDM. Normalized mean reads (log scale) for all annotated *S. mutans* open reading frames (mean values are from three independent experiments). ORFs of the ComR (green), ComED (red), and SigX (light and dark blue) regulons, as shown in Fig. 2. The ORFs of the SigX regulon comprising the pan-streptococcal SigX regulon are in dark blue. The yellow star corresponds to the *comR* gene. Four-fold change marker (blue line; 135 ORFs > 4, 19 ORFs < 4).

of one of the three regulons (Fig. 2 and Fig. 3). Transcripts induced in the anti-sense direction are presented in the maps depicted in Supplementary Figure S1. Overall, the transcriptome obtained here by RNA-seq and XIP as the inducing pheromone showed induction of basically the same sets of transcripts as those found in the previous comprehensive study using a tiling array and CSP, with few exceptions.

**Stepwise deletions link the transcription of the intragenic region upstream of *comR* to expression of *sigX*, and point to a second predicted promoter upstream of *comR*.** To investigate the involvement of the 3′ portion of SMU.60 in regulation of competence, we used a stepwise deletion strategy to construct (1) a mutant that lacked the 5′ part of the SMU.60, but in which the intragenic SigX box was preserved (mutant SM181), (2) a mutant that lacked the entire SMU.60, including the intragenic SigX box (mutant SM183), and (3) a mutant that lacked both the entire gene SMU.60, the SigX-box, and in addition a *comR* upstream region with a palindromic sequence close to the start of the *comR* ORF (SM185) (Fig. 5). All the mutants were constructed in a *comS* deletion background that included a SigX-luciferase reporter.

The expression of *sigX* upon stimulation with XIP in mutants SM181, SM183, and SM185 was compared to that in the SM091 control using the luminescence assay. As expected, without XIP no *sigX* expression was observed in any of the strains (for simplicity, the results are only shown for the SM091 in Fig. 5). In the presence of XIP, induction of *sigX* was reduced by approximately 40% in SM183, and by 100% in SM185, while no significant differences were observed between SM181 and the control SM091 strain (Fig. 5). The results thus indicate that an intact DNA alkylation repair protein encoded by SMU.60 is not required for *sigX* induction by XIP. In contrast,
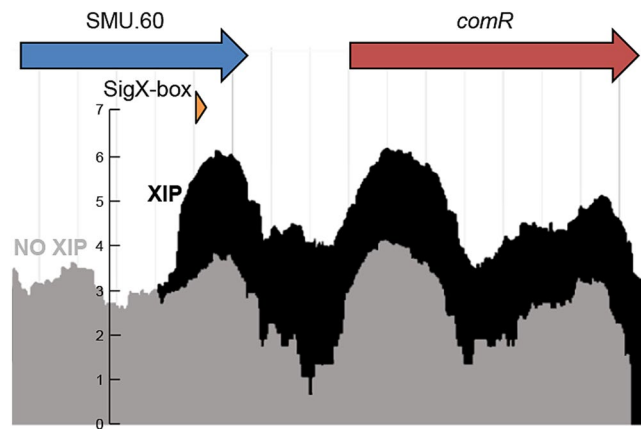
**Figure 4.** Transcriptome map showing expression of *comR* and its upstream region in *S. mutans* Δ*comS* grown in the presence (black) or absence (light gray) of XIP in chemically defined medium. The images are superimposed to illustrate the differences in expression between the two conditions, starting next to the predicted intragenic SigX box. The axis on the left represent the log$_2$ average of read counts of normalized individual duplicated samples for each condition, as detailed in material and methods.
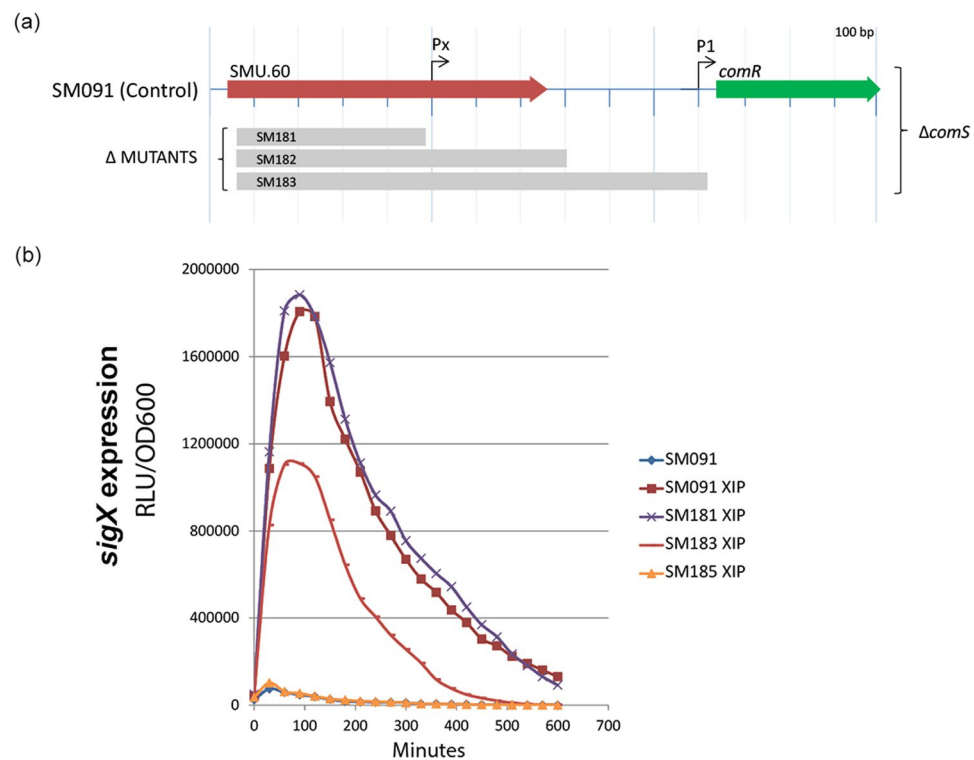


**Figure 5.** Effect of stepwise deletions upstream of *comR* on *sigX* expression. (**a**) Map of the SMU.60 intragenic region exhibiting a putative promoter with a σx-box (Px; AACGAATA), a predicted *comR* promoter close to the start of the gene (P1; palindrome sequence TAGTAAATTTACTA), and the regions deleted in the SigX-luciferase reporter mutants SM181, SM183, and SM185, respectively. All the strains were constructed in a Δ*comS* background. (**b**) *sigX* expression in response to XIP in the wild type background (SM091), and in the mutants SM181, SM183, and SM186. Luciferase activity was measured in the presence of luciferin, and is presented as relative light units (RLU) per OD600. The results are mean values of 3 experimental replicates, and are representative of 4 independent experiments.

induced expression of *sigX* was reduced two-fold by deleting the entire SMU.60 orf, and was completely abolished by extending the deletion nearly to *comR*. This pattern is consistent with expression of *comR* being absolutely dependent on a proximal promoter (P1) and partially dependent on an additional promoter (Px) located within SMU.60 near the apparent SigX-box.
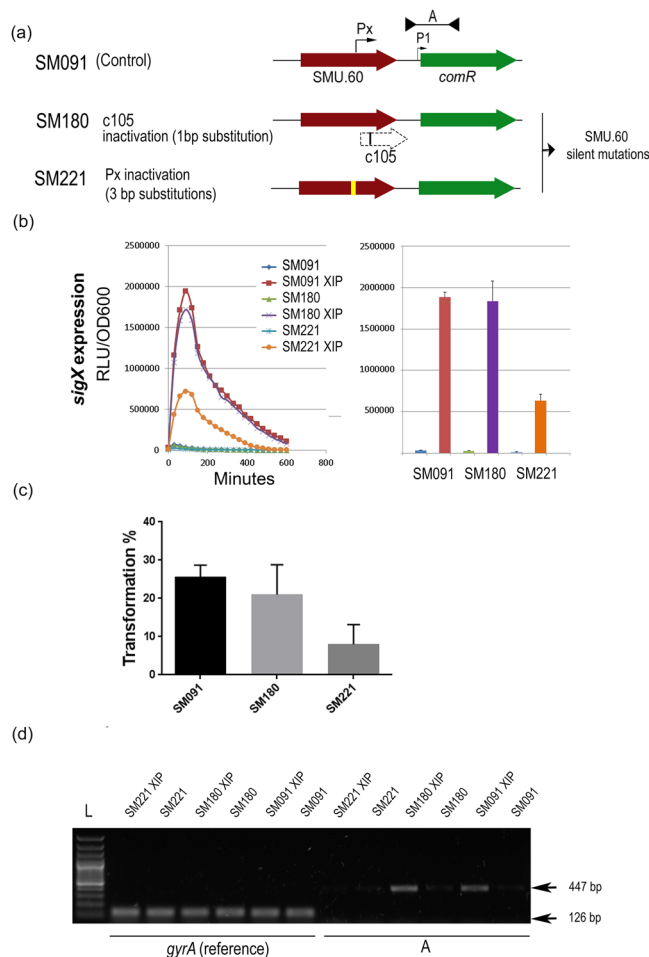
**Figure 6.** Effect of single base mutations upstream of *comR* on *sigX* expression, competence, and *comR* transcript length. (**a**) Map of the SMU.60 intragenic region with a putative promoter containing a σx-box (Px) and the downstream predicted *comR* promoter closest to the start of the *comR* gene (P1). The A region corresponds to the PCR fragment amplified by the primer pair FP785 and FP1200. The *S. mutans* strains used were all Δ*comS*, SigX-luciferase reporters: SM091, wild type background; SM221, mutant with an inactivated σx-box (yellow vertical box); SM180, mutant with a stop codon (black vertical line) within a possible alternative open reading frame (ORF; dashed arrow) overlapping with part of the SMU.60 gene. (**b**) *sigX* expression with and without addition of XIP in CDM. Luciferase activity measured in the presence of luciferin is presented as relative light units (RLU) per OD600. Left panel: Kinetic results representative of three independent experiments. Right panel: Average from three independent experiments at 100 min. Bars represent standard deviation. (**c**) Transformation efficiency using the 6,3 Kb amplicon aRJ02 as donor DNA. The results are averages of 3 independent experiments. Bars represent standard deviations. (**d**) Gel image showing cDNA amplification in the A region of the WT and SM180 in the presence of XIP. The A transcript was not detected in SM221. Primers FP785 and FP1200 were used to amplify A (447 bp), and FP299 and FP300 to amplify *gyrA* (126 bp; reference). L; 1Kb Plus DNA ladder (Invitrogen).

**Inactivation of the intragenic SigX box by direct genome editing reduces *sigX* expression and induced expression of *comR*, and reduces transformation in response to XIP.** To investigate the intra-SMU.60 SigX-box site more precisely, we chose to inactivate the SigX box in SMU.60 without affecting its expression, by using primers for overlapping PCR in which bases A, C, and G in positions 2, 3, and 8 in the SigX box (T**AC**GAAA**A**) were replaced by G, A, and G (T**GA**GAAA**G**). The three single base substitutions introduced only silent mutations in SMU.60 (SM221 mutant). Both expression of *sigX* and transformation were reduced in SM221 by 50%. We then compared the expression of the transcript initiated at the putative SigX box in SM221 with the expression in the control strain SM091. Induction of the transcript by XIP in this mutant, as measured by reverse-trancriptase PCR, was severely reduced (Fig. 6). We conclude that the SigX box is essential for the increase in *comR* expression observed both here and in previous studies[16, 19, 20, 22].

**Introduction of a stop codon in a putative open reading frame downstream of the intragenic SigX box has no significant effect on the expression of *sigX* and *comR*, or on competence.** Annotation of the genome at HOMD (http://www.homd.org/) indicates the presence of a short ORF

**(a)**

```
S. mutans UA159
SMU.0061          GTGGCAATTGACCACCAGCTTTTACGAAAA -(574)-ATG
SMU.1001          CTCTATCAATTGATAGAGTTTTTTCGAATA - (22)-ATG
SMU.0325          AAATGAACTTTTGCTTTTGCTTTGCGAATA -(119)-ATG

S. pyogenes NZ131
Spy49_0032        TCCCTTCTTTTTTATTTAAAAAAACGAATA - (80)-ATG

Consensus                                   TACGAATA
```
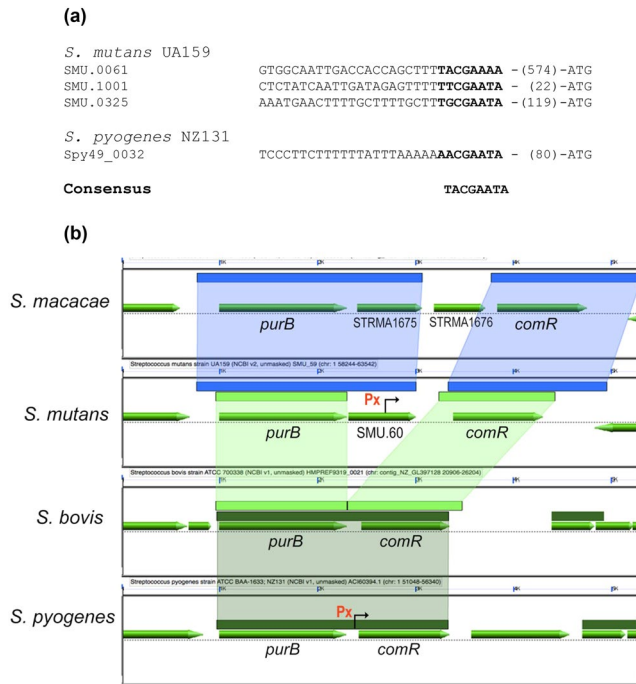
**(b)**



**Figure 7.** Alignment of DNA sequences upstream of *comR* in *S. mutans* UA159, *S. pyogenes* NZ131, *S. bovis* ATCC700338, and *S. macacae* NCTC11558. (**a**) In bold are the SigX-boxes in the Px promoters for *comR* in *S. mutans* and *S. pyogenes*. No apparent SigX-box was identified in *S. macacae* or *S. bovis*. The SigX-box regions of the most induced locus, *dprA* (SMU.1001), and of the least induced locus, *dut/radA* (SMU.0325) are also shown. (**b**) Synteny in the *comR* region as determined by CoGe:GEvo, using the conserved *purB* gene as a reference (https://genomevolution.org/coge). Mutans group; Synteny in the *comR* region was found only in *S. macacae*, as determined by CoGe:GEvo, using the conserved *purB* gene as a reference.

(c105) immediately downstream of the SMU.60 intragenic SigX-box. To investigate whether expression of a protein encoded by this ORF could be responsible for the linkage of competence to the presence of the SigX-box, we constructed a mutant with a single-base substitution that created both a silent mutation in SMU.60 and a stop codon at the start of the putative ORF (SM180). This mutant was not affected in *sigX* expression, transformation, or the formation of the extended transcript (Fig. 6). This indicates that if this protein is produced at all, it has no effect on competence.

### Conservation of the SMU.60 intragenic SigX-box in *S. mutans*.

Analysis of all sequenced genomes of *S. mutans* using BLAST at COGE (https://genomevolution.org/coge) revealed that the SMU.60 intragenic SigX-box and the position of the *comR* gene is similar in all of them, indicating that the regulation of *comR* by SigX is widely conserved in *S. mutans*. Other members of the mutans group of streptococci are *Streptococcus sobrinus*, *Streptococcus criceti*, *Streptococcus downei*, and *Streptococcus macacae*, but *comRS* genes among these members were found only in *S. macacae*. The arrangement of the locus in *S. macacae* differs, however, from that in *S. mutans*, as there is an additional gene positioned between SMU.60 and *comR*, and that there is no apparent SigX-box in the SMU.60 intra-genic region. Among the 4 other streptococcal groups that use ComRS for competence regulation, including salivarius, bovis, *S. suis*, and *S. pyogenes*, only species in the bovis and pyogenic groups had a gene arrangement that resembled that in *S. mutans*, showing a *purB* gene upstream of *comR*. However, SMU.60 homologues are not found in their genomes. Despite the lack of SMU.60, there is a SigX-box signature upstream of *comR* in *S. pyogenes*, but not in *S. bovis*. The presence of the SigX-box only 80 bp upstream of the *comR* start codon suggests that *S. pyogenes* may regulate *comR* expression via SigX as well (Fig. 7).

### Discussion

In bacteria that develop competence transiently, the coordinated up-regulation of dozens of genes is supported by one or more positive feedback loops. In *S. pneumoniae*, for instance, the exported CSP pheromone stimulates its own production via the ComDE two-component signal transduction system. Another example is the ComK transcription activator of the DNA-uptake and recombination machinery in *Bacillus subtilis*, which binds to its promoter to activate its own expression. In *S. mutans*, one positive feedback loop connects SigX to ComE and bacteriocin expression and a second connects XIP to *comS* expression. Both feedback loops are involved in the generation of population heterogeneity sometimes observed during competence in *S. mutans*. Yet, the bi-stability behavior cannot be explained by the two loops alone[23]. The involvement of an additional positive loop involving ComR has been put forward as a possible explanation[24]. However, no mechanism for the regulation of the central
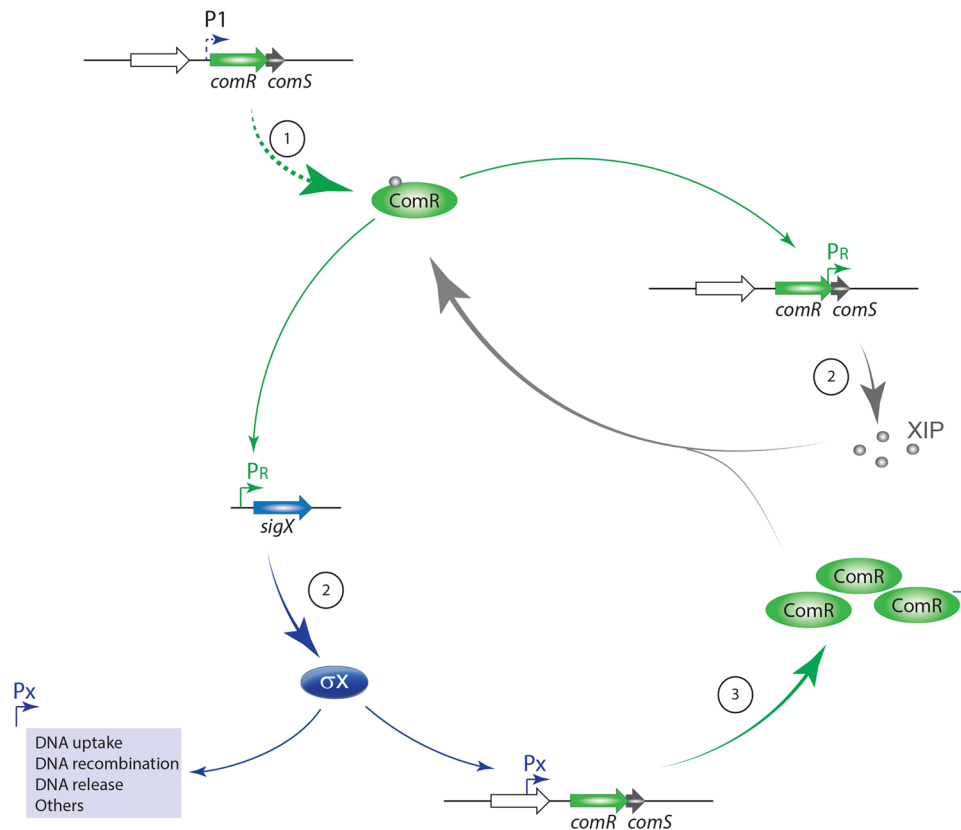
**Figure 8.** Model of *comR* regulation. (1) Basal levels of ComR and the XIP precursor ComS are kept by transcription of *comRS* from the promoter P1. Imported XIP binds to ComR, forming the ComRS complex. (2) ComRS induces the expression of *sigX* and *comS*, by binding to the PR promoter. This positive feedback loop generates increased amounts of XIP. (3) A second positive feedback loop is mediated by SigX, which, upon recognizing an intragenic SigX-box in the *comR* neighbor gene SMU.60, up-regulates the expression of *comR*. Thus, the system couples expression of *comS*, *comR*, and *sigX* to maximize amounts of the ComR regulator and the XIP cognate pheromone during competence.

competence regulator ComR is yet known, and although increased expression of *comR* during *S. mutans* competence has been reported, the described changes are small[16–20].

The proximity of the intragenic SMU.60 SigX box to *comR* and *comS* suggested that this region could play a previously unknown role in competence, perhaps in the regulation of *comR*. Using optimized protocols to achieve high levels of competence[21, 25], the induction of *comR* became more prominent. Nine-fold induction of *comR* was detected in CDM, in contrast to previous reports with results ranging from no induction to 1.4 and 4-fold induction[2, 16–20, 26]. Evaluation of transformation and *sigX* expression in mutants representing stepwise deletions within this segment revealed the regulatory mechanism involved. Marker-less single base substitutions to inactivate the SigX box resulted in a mutant that reproduced the partial competence inhibition found in the SigX box deletion mutant, whereas a base substitution that introduced a stop codon in the putative c105 orf encoding a peptide transcribed from the SigX-box had no effect on competence. Finally, the RNA-seq and RT-PCR results indicating that the transcript initiated at the SigX-box is co-transcribed with *comR*, led to the conclusion that SigX regulates *comR* expression. The discovered mechanism for *comR* regulation is apparently the missing explanation for previous reports on *comR* induction during competence. It may also explain why complementation of a *comR* deletion mutant with a plasmid that incorporates only the proximal promoter of *comR*, and not the SigX-box, does not completely rescue the transformation phenotype observed in the wild type[7], while expressing the same plasmid in the wild type results in a ten-fold increase in transformation efficiency.

Distal to *comS*, there is an additional set of genes consistently reported as up-regulated during competence that extends from SMU.64 to SMU.66, and occasionally to SMU.68[16–20]. Between *comS* and this set of genes is SMU.63c, which codes for a recently identified amyloid protein induced by both XIP and CSP[27]. The reported induction in gene expression contrasts with a recent microarray study using directly labeled RNA, showing that induction of SMU.63c was restricted to the non-coding strand[2]. The present directional RNA-seq results confirmed this observation, and showed that in the sense-direction, the expression of SMU.63c was actually reduced by more than 35-fold. While the most reasonable explanation for the increase in SMU.63c expression reported in previous studies is the use of methodologies that are not strand specific, it does not account for a recent transcriptome study using a directional approach that showed an increase of SMU.63c in the sense direction[26]. In the absence of a candidate regulator for SMU.63.c, it is reasonable to infer that SMU.63c transcription is induced from

| Name | Description | Source |
|------|-------------|--------|
| UA159 | Wild-type | ATCC 700610 |
| SM061 | UA159, but Δ*sigX:: kan*; Kan[R] | This study |
| SM065 | UA159, but Δ*comS::spc*; Spc[R] (from strain MW02 Mashburn-Warren *et al.* (2010) | (Khan *et al.*, 2012) |
| SM066 | UA159, but Δ*comR::spc*; Spc[R] (from strain MW05 Mashburn-Warren *et al.* (2010) | This study |
| SM068 | UA159, $p_{sigX}$-luc; Spc[R] | (Khan *et al.*, 2012) |
| SM089 | UA159, but Δ*comS::ery*; Erm[R] | (Khan *et al.*, 2012) |
| SM091 | SM068, but Δ*comS::ery*; Spc[R] Erm[R] | (Khan *et al.*, 2012) |
| SM166 | UA159, but Δ*SMU.60::kan*; Kan[R] | This study |
| SM168 | UA159, but Δ*SMU.60_c105::kan*; Kan[R] | This study |
| SM170 | UA159, but Δ*SMU.60_pcomR::kan*; Kan[R] | This study |
| SM180 | SM188, but Δ*comS::ery*; Spc[R] Erm[R] | This study |
| SM181 | SM091, but Δ*SMU.60::kan*; Kan[R] Spc[R] Erm[R] | This study |
| SM183 | SM091, but Δ*SMU.60_c105::kan*; Kan[R] Spc[R] Erm[R] | This study |
| SM185 | SM091, but Δ*SMU.60_pcomR::kan*; Kan[R] Spc[R] Erm[R] | This study |
| SM188 | SM068, but c_105 1-bp substitution; Spc[R] | (Morrison *et al.*, 2015) |
| SM218 | UA159, but mutated *cinbox* upstream of c_105; marker-less | This study |
| SM221 | SM091, but mutated *cinbox* upstream of c_105; Spc[R] Erm[R] | This study |

**Table 1.** Strains used in the study.

the non-coding strand as part of a single transcript that extends all the way to SMU.66 or beyond, depending on the magnitude of competence induction.

Streptococci have a core of 27 to 30 genes that are under the control of the alternative sigma factor SigX[2]. All of them were induced by XIP in the present study, confirming the conservation of the regulon. Two other studies have also investigated the effect of XIP on *S. mutans* competence[20, 26]. In the first, five of the SigX core genes were not induced during competence (*dut*, *radA*, *dpnA*, *pilC*, and *radC*)[20], but in this study exposure to XIP was for a short period, and under conditions that do not prevent endogenous XIP production. In the other, only *dut* and *radA* were not found to be up-regulated[26]. The *dut* and *radA* genes are transcribed as a single operon that is among the least induced transcripts of the SigX regulon, which may explain why their induction was not detected in the two studies. The fold changes for all genes in this study, as well as in the previous comprehensive study defining the core set of panstreptococcal genes of the SigX regulon[2], and the recently reported XIP transcriptome[26] are presented in Supplementary Table S1. All the accessory genes of the SigX regulon, as well as the genes of the ComRS and ComE regulons identified in the comprehensive study mentioned above were also up-regulated in our study. The surpluses of genes induced in the present study were, except for a transcript initiated at SMU.1340, extension of transcripts belonging to one of the three regulons (Fig. 2).

A notable feature of the regulons induced during competence in streptococci is that core genes lie close to the promoters recognized by the main regulators of competence[2]. The majority has a function directly linked to natural transformation. The link to stress responses, on the other hand, is not well understood, and in most cases is related to accessory genes located downstream of the core genes. Among these is the *relQ* ppGpp synthase (SMU.1046c) and *levDEFG* (SMU1957 to SMU.1961), both associated with nutritional stress, and induced by XIP[26]. The present results confirmed their up-regulation and further revealed that they belong to the class of accessory genes which transcription initiates from upstream core genes of the competence system (Supplementary Figure 1). It remains to be determined whether *S. mutans* may have co-opted this arrangement of SigX regulons close to genes that may slow down their growth as part of a stress response, or whether what we observe is an indirect effect caused by variation in terminator presence and efficacy, as previously suggested[2].

Combining the present results with previous literature, leads to a model in which *comR* is expressed at a basal level under non-competence conditions, using a P1 neighbor promoter near a distal palindromic sequence (Fig. 8). During development of competence, XIP or intracellular ComS binds to ComR, initiating the competence response. This leads to elevated transcription of *sigX* and *comS*, and to increased levels of SigX and XIP. SigX in turn drives increased transcription of *comR* from a secondary promoter (Px), to assure that sufficient amounts of the ComR regulator are produced to mount an appropriate response. Placing *comR* and *comS* transcripts under the control of different regulators so as to form two interlocked positive feedback circuits may help *S. mutans* to better adjust the kinetics and magnitude of the competence response. This tight control is particularly important given the fitness cost of competence, which should be high judging by the significant autolysis observed in part of the competent *S. mutans* population[28], and the dramatic changes in the array of proteins synthesized during competence[29].

## Materials and Methods
**Bacterial strains and growth conditions.** *S. mutans* strain UA159 and isogenic derivatives used in the study are listed in Table 1. Most mutations were inserted into a $p_{sigX}$-luc and Δ*comS* background. For preparation of pre-cultures and assays, *S. mutans* were cultivated in 5% $CO_2$ at 37 °C in tryptic soy broth (TSB) (Oxoid) or

| ID | Sequence | Restriction site | Description |
|---|---|---|---|
| FP460 | CATTCCCTCTTGTTGCCAAT | — | Construction of SM061 |
| FP461 | aggcgcgccTGCCGAACACAGCAGTTAAG | AscI | |
| FP462 | aggccggccCTTGGTAGCAGGAGAGCAC | FseI | |
| FP463 | AAAGCACAGCCTGCTTCAAT | — | *Kan marker in the FW strand* |
| FP001 | aggcgcgccGTTTGATTTTTAATG | AscI | |
| FP068 | aggccggccTAGGTACTAAAACAATTCATCCAGTA | FseI | |
| FP869 | ACGCCATCATTCGTAAGGAC | — | Construction of SM166 and SM181 |
| FP870 | aggccggccAGCATAAAAGGCAGCCCATC | FseI | |
| FP871 | aggcgcgccGCAATTGACCACCAGCTTTT | AscI | |
| FP872 | CTTTGTTGGTCGCCATAGGT | — | *KanT marker in the RC strand* |
| FP001 | aggcgcgccGTTTGATTTTTAATG | AscI | |
| FP002 | aggccggccATCGATACAAATTCCTC | FseI | |
| FP869 | ACGCCATCATTCGTAAGGAC | — | Construction of SM168 and SM183 |
| FP870 | aggccggccAGCATAAAAGGCAGCCCATC | FseI | |
| FP859 | aggcgcgccTGAAATTTTCTAGCCAGTCGTTT | AscI | |
| FP860 | TTTCAATTCTAAGAGGAGTCCAAAA | — | *KanT marker in the RC strand* |
| FP001 | aggcgcgccGTTTGATTTTTAATG | AscI | |
| FP002 | aggccggccATCGATACAAATTCCTC | FseI | |
| FP869 | ACGCCATCATTCGTAAGGAC | — | Construction of SM170 and SM185 |
| FP870 | aggccggccAGCATAAAAGGCAGCCCATC | FseI | |
| FP863 | aggcgcgccTAAAAGTTTAGACAGAGATTATAGGAAAAGG | AscI | |
| FP864 | TGATGCTTCCATGTCAGAAGA | — | *KanT marker in the RC strand* |
| FP001 | aggcgcgccGTTTGATTTTTAATG | AscI | |
| FP002 | aggccggccATCGATACAAATTCCTC | FseI | |
| FP894 | TCCGGATGCAGAAGGTATTC | — | Construction of SM180 |
| FP895 | CAATAAAAGTTCTC ACCCAATCTGGA | — | |
| FP896 | TCCAGATTGGGTG AGAACTTTTATTG | — | |
| FP897 | CATCCTGCCGTTCCTATCAT | — | |
| FP898 | GGTTGATTGGGTTTTTGTGG | — | Screening primers for SM180 specific to UA159 |
| FP899 | TTTTTATGCTTTTCAATAAAAGTTCTA | — | |
| FP902 | CTCTAAGACTAATCCAGATTGGGTT | — | |
| FP903 | GCGAGTTTCAAAAAGGAAGC | — | |
| FP900 | CGGATTGGATTGGGAGACTA | — | Screening primers for SM180 specific to the mutant |
| FP901 | TTTTTATGCTTTTCAATAAAAGTTCTC | — | |
| FP904 | CTCTAAGACTAATCCAGATTGGGTG | — | |
| FP905 | GGCAGACAGCTTCTTTGGTC | — | |
| FP1552 | GCGAATTGAGGAAATGGTTG | — | Construction of SM218 |
| FP1553 | TGAGTCAGCCGATCAAAGTG | — | |
| FP299 | CCATGACCATCAACCAACAT | — | RT/PCR |
| FP300 | ATCAGCGCGTATTACAGGTG | — | *gyrA* |
| FP867 | AATCAGATGGCCAATTTATCG | — | RT/PCR |
| FP868 | GCGGTTGTTTTTCGTGTGAT | — | *transcript start* |
| FP785 | TCGTTTGCTGCAAGACTACG | — | RT/PCR |
| FP786 | GGGAACACATTCAGCGAGTT | — | *transcript mid* |
| FP1199 | AACGGACGGAGAAAATGTTG | — | RT/PCR |
| FP1200 | CTTTGTTGGTCGCCATAGGT | — | *transcript end (comR)* |
| FP299 | CCATGACCATCAACCAACAT | — | Gel electrophoresis |
| FP300 | ATCAGCGCGTATTACAGGTG | — | *gyrA* |
| FP785 | TCGTTTGCTGCAAGACTACG | — | Gel electrophoresis |
| FP1200 | CTTTGTTGGTCGCCATAGGT | — | *upstream of comR* |

**Table 2.** Primers used in the study.

chemically defined medium (CDM)[30]. For plates, 1.5% Bacto agar (Becton, Dickinson and Company) was added to the media. When needed, antibiotics were used at the following concentrations: kanamycin (Kan), 500 μg/mL; erythromycin (Erm), 10 μg/mL; spectinomycin (Spc), 500 μg/mL.

**Synthetic peptides.** The *sigX*-inducing peptide (XIP) (NH2-GLDWWSL-COOH; 98% purity; GenScript) was reconstituted with 20 μl dimethyl sulfoxide (DMSO) (Sigma-Aldrich), to which 1 mL distilled water was added to give a final concentration of 10 mM. The mature competence-stimulating peptide (CSP18)[31, 32] (NH2-SGSLSTFFRLFNRSFTQA-COOH; 98% purity; GenScript) was reconstituted in water to a final concentration of 1 mM. Both peptides were stored in aliquots at −20 °C.

**Construction of mutants with antibiotic resistance markers.** A PCR-ligation mutagenesis strategy was used for the construction of four deletion mutants carrying antibiotic markers in the wild-type *S. mutans* strain UA159 (SM061, SM166, SM168 and SM170)[33]. Briefly, flanking regions of ~ 1-kb up- and down-stream of the deleted region were amplified from genomic template DNA, and the kanamycin resistance cassette was amplified from a plasmid. Amplicons carried AscI or FseI restriction sites which were digested with specific restriction enzymes, and ligated with T4 DNA ligase (New England Biolabs) for the creation of the final ligation product. The kanamycin resistance cassette in SM166, SM168 and SM170 was placed in the reverse strand to avoid transcriptional read-through from its promoter into downstream genes. Transformation into *S. mutans* was performed by culturing cells with CSP18 as described previously[21]. Briefly, stock cultures at OD600 0.5 were diluted 1:10 in TSB and an aliquot of a 1 mL was transferred to a microtube. Donor DNA and 18-CSP (50 nM) were added, and the culture was incubated at 37 °C in air for 3 h. At the end of the incubation period, cultures were plated onto selective THB-agar plates. Primers used for the construction of amplicons are listed in Table 2. The mutants were verified for gene deletion by PCR and gel electrophoresis. Genomic DNA from SM166, SM168 and SM170 was transformed into the *ΔcomS* background to create SM181, SM183 and SM185 respectively. SM066 was created by transformation with DNA from previously described strains[7].

**Genome editing.** A marker-less mutation in strain SM218 was constructed using a high-efficiency strategy for genome editing[25]. Overlapping primers were designed with three single base substitutions in the SigX box of SMU.60 present upstream of *comR*. The introduction of three base substitutions inactivated the SigX box with creation of silent mutation in gene SMU.60. The resulting amplicon containing the desired substitutions were transformed in *S. mutans* UA159. Colonies were selected and screened by PCR to identify a clone containing the mutated SigX box, which was designated as SM218. The amplicon was also transformed into SM091 to place the mutation in a *comS* defective background (SM221).

**Transformation.** Competence for natural transformation in *S. mutans* can be achieved either indirectly by use of CSP or directly with XIP; hence, two different protocols previously described were used for transformation in this study[21, 25]. A traditional method uses CSP in complex media such as THY and does not require chemically-defined media, while a recently described method using XIP in CDM results in higher efficiency of transformation. The traditional method was used for transformation of mutations carrying antibiotic markers[21], since maximal transformation levels were not required. Briefly, pre-culture stocks at OD$_{600}$ 0.5 were diluted 1:10, and CSP was added to a final concentration of 50 nM. Purified PCR products were used as donor DNA. Cultures were then incubated at 37 °C in air for 3 h. The more recent strategy using XIP in CDM was employed for construction of marker-less mutations[25].

**Luciferase reporter assays.** For luciferase assays, frozen pre-cultures of reporter strains were prepared by growth in CDM to an optical density at 600 nm (OD$_{600}$) of 0.5, supplementation to 15% glycerol, and storage at −80 °C. For the experiments, pre-cultures were thawed, diluted 10-fold in CDM in the presence or absence of 1 μM XIP, and distributed as 200 μL aliquots in wells of 96-well microtiter plates (Thermo Fisher Scientific). Blank wells contained only CDM and luciferin. A volume of 20 μL of a 1 mM aqueous D-luciferin solution (Synchem, Felsberg-Altenberg, Germany) was added to each well, followed by sealing of the plates with Top Seal (PerkinElmer), and incubation at 37 °C. Relative luminescence units (RLU) and OD600 were measured at various time intervals during growth by reading the plates in a multi-detection microplate reader (Synergy HT and Cytation 3; BioTek).

**Real time PCR.** For RNA extraction, the High Pure RNA isolation kit (Roche) was used, following the protocol provided by the producer, except that we added one step to improve cell lysis: cells were incubated at 37 °C for 20 min in 200 μL of lysis buffer containing 10 mM Tris (pH 8), 20 mg per mL of lysozyme, and 100 U of mutanolysin per mL. DNase I was used during and after RNA extraction to ensure complete removal of DNA. Templates of cDNA templates were prepared from RNA with the First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). In all assays, controls without reverse transcriptase were included. Assays were carried out with quantitative PCR master mix SYBR green/ROX (Thermo Fisher Scientific), according to the manufacturer's protocol. For data comparisons, ROX was used as a reference, and the constitutive gene *gyrA* as a normalizer. Real-time PCR was performed (Agilent Technologies Stratagene Mx3005 P), and data were collected and analyzed with the software and graphics program MxPro (Stratagene). Table 2 shows the list of primers used.

**Gel electrophoresis.** DNA and cDNA amplified by True start Hot start Taq DNA Polymerase (Thermo Fisher Scientific). Gel electrophoresis was performed by preparing 1% Gel (SeaKem LE Agarose; Lonza) in TAE Buffer with SYBR safe DNA gel stain (Thermo Fisher Scientific) 1:10000. Sample and loading dye were mixed 5:1 and loaded into the wells of agarose gel. Gels were protected from light during 25–50 min at 65 V. Gene ruler 1 kb ladder was used as a size standard, and images were captured in G:Box (Syngene) with GeneSnap software v.7.12 (Syngene).

**RNA-sequencing analysis.** An overnight culture of SM065 was diluted 1:10 in CDM to a final volume of 180 mL, corresponding to an optical density at 600 nm of 0.2. Total volume of diluted culture was equally divided into two tubes and grown in the presence or absence of 1 μM XIP for 2 hours at 37 °C in 5% $CO_2$ atmosphere ($OD_{600}$ of approximately 0,4); cells were then harvested for RNA extraction. Initial RNA extraction was performed with the mirVana miRNA Isolation Kit (Life Technologies), followed by DNase treatment with TURBO DNase (Life Technologies) and a second RNA extraction. RNA quality was verified by running the samples on Bioanalyzer (Agilent 2100 Expert) prior to enrichment with MICROBExpress (Thermo Fisher Scientific). Briefly, purified RNA was incubated with the Capture Oligonucleotide Mix in Binding Buffer for 15 minutes. Magnetic beads, derivatized with an oligonucleotide that hybridizes to the capture oligonucleotide, were then added to the mixture and allowed to hybridize for 15 minutes. The magnetic beads, with 16 S and 23 S rRNAs attached, were pulled to the side of the tube with a magnet. The enriched RNA in the supernatant was transferred to a new microtube. Magnetic beads were briefly washed, and finally the RNA was precipitated with ethanol. More than 90% removal of 16 S and 23 S rRNA was verified using Bioanalyzer. The RNA samples were then treated with NEXTFlex Directional RNA-seq Kit (dUTP based) for preparation of the DNA library for sequencing using Illumina® HiSeq, following the intructions provided by the producer. A FASTQ file was generated for each sample. Bioinformatic analyses were performed as described previously[34, 35]. Briefly, using the software Bowtie, nucleotide reads were mapped to the annotated genome of strain UA159 (GenBank accession: NC_004350.2). SAMtools software was used to divide the reads in each according strand. For visualization, JBrowse viewer was employed[36]. Differential expression analyses of samples with and without XIP were determined based on the mapped read profiles using a Perl script. The read counts were subjected to "DESeq" to evaluate the differential levels when comparing samples. The results are derived as means from three independently conducted biological experiments. The average and standard deviation for the total number of paired-end sequenced fragments was 7 456 931 (SD 27 663) for the control, and 7 271 703 (SD 365 331) for the XIP-treated samples.

## References

1. Fontaine, L. *et al.* A novel pheromone quorum-sensing system controls the development of natural competence in *Streptococcus thermophilus* and *Streptococcus salivarius*. *J Bacteriol* **192**, 1444–1454, doi:10.1128/JB.01251-09 (2010).
2. Khan, R. *et al.* Comprehensive Transcriptome Profiles of *Streptococcus mutans* UA159 Map Core Streptococcal Competence Genes. *mSystems* **1**, doi:10.1128/mSystems.00038-15 (2016).
3. Lee, M. S. & Morrison, D. A. Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. *J Bacteriol* **181**, 5004–5016 (1999).
4. Peterson, S. N. *et al.* Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol Microbiol* **51**, 1051–1070 (2004).
5. Håvarstein, L. S., Gaustad, P., Nes, I. F. & Morrison, D. A. Identification of the streptococcal competence-pheromone receptor. *Mol Microbiol* **21**, 863–869 (1996).
6. Khan, R. *et al.* Extracellular identification of a processed type II ComR/ComS pheromone of *Streptococcus mutans*. *J Bacteriol* **194**, 3781–3788, doi:10.1128/JB.00624-12 (2012).
7. Mashburn-Warren, L., Morrison, D. A. & Federle, M. J. A novel double-tryptophan peptide pheromone controls competence in Streptococcus spp. via an Rgg regulator. *Mol Microbiol* **78**, 589–606, doi:10.1111/j.1365-2958.2010.07361.x (2010).
8. Monnet, V., Juillard, V. & Gardan, R. Peptide conversations in Gram-positive bacteria. *Crit Rev Microbiol* **42**, 339–351, doi:10.3109/1040841X.2014.948804 (2016).
9. Pestova, E. V., Håvarstein, L. S. & Morrison, D. A. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol Microbiol* **21**, 853–862 (1996).
10. Johnston, C., Martin, B., Fichant, G., Polard, P. & Claverys, J. P. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* **12**, 181–196, doi:10.1038/nrmicro3199 (2014).
11. Morrison, D. A., Guedon, E. & Renault, P. Competence for natural genetic transformation in the *Streptococcus bovis* group streptococci *S. infantarius* and *S. macedonicus*. *J Bacteriol* **195**, 2612–2620, doi:10.1128/JB.00230-13 (2013).
12. Shanker, E. *et al.* Pheromone Recognition and Selectivity by ComR Proteins among Streptococcus Species. *PLoS Pathog* **12**, e1005979, doi:10.1371/journal.ppat.1005979 (2016).
13. Zaccaria, E. *et al.* Control of competence for DNA transformation in *Streptococcus suis* by genetically transferable pherotypes. *PLoS One* **9**, e99394, doi:10.1371/journal.pone.0099394 (2014).
14. Fontaine, L. *et al.* Mechanism of competence activation by the ComRS signalling system in streptococci. *Mol Microbiol* **87**, 1113–1132, doi:10.1111/mmi.12157 (2013).
15. Haustenne, L., Bastin, G., Hols, P. & Fontaine, L. Modeling of the ComRS Signaling Pathway Reveals the Limiting Factors Controlling Competence in *Streptococcus thermophilus*. *Front Microbiol* **6**, 1413, doi:10.3389/fmicb.2015.01413 (2015).
16. Dufour, D., Cordova, M., Cvitkovitch, D. G. & Levesque, C. M. Regulation of the competence pathway as a novel role associated with a streptococcal bacteriocin. *J Bacteriol* **193**, 6552–6559, doi:10.1128/JB.05968-11 (2011).
17. Lemme, A., Grobe, L., Reck, M., Tomasch, J. & Wagner-Dobler, I. Subpopulation-Specific Transcriptome Analysis of Competence-Stimulating-Peptide-Induced *Streptococcus mutans*. *J Bacteriol* **193**, 1863–1877, doi:10.1128/Jb.01363-10 (2011).
18. Okinaga, T., Niu, G., Xie, Z., Qi, F. & Merritt, J. The *hdrRM* operon of *Streptococcus mutans* encodes a novel regulatory system for coordinated competence development and bacteriocin production. *J Bacteriol* **192**, 1844–1852, doi:10.1128/JB.01667-09 (2010).
19. Perry, J. A., Jones, M. B., Peterson, S. N., Cvitkovitch, D. G. & Levesque, C. M. Peptide alarmone signalling triggers an auto-active bacteriocin necessary for genetic competence. *Mol Microbiol* **72**, 905–917, doi:10.1111/j.1365-2958.2009.06693.x (2009).
20. Reck, M., Tomasch, J. & Wagner-Dobler, I. The alternative sigma factor SigX controls bacteriocin synthesis and competence, the two quorum sensing regulated traits in *Streptococcus mutans*. *PLoS Genet* **11**, e1005353, doi:10.1371/journal.pgen.1005353 (2015).
21. Salvadori, G. *et al.* Natural Transformation of Oral Streptococci by Use of Synthetic Pheromones. *Methods Mol Biol* **1537**, 219–232, doi:10.1007/978-1-4939-6685-1_13 (2017).
22. Okinaga, T., Xie, Z., Niu, G., Qi, F. & Merritt, J. Examination of the *hdrRM* regulon yields insight into the competence system of *Streptococcus mutans*. *Mol Oral Microbiol* **25**, 165–177, doi:10.1111/j.2041-1014.2010.00574.x (2010).
23. Hagen, S. J. & Son, M. Origins of heterogeneity in *Streptococcus mutans* competence: interpreting an environment-sensitive signaling pathway. *Phys Biol* **14**, 015001, doi:10.1088/1478-3975/aa546c (2017).
24. Son, M., Ahn, S. J., Guo, Q., Burne, R. A. & Hagen, S. J. Microfluidic study of competence regulation in *Streptococcus mutans*: environmental inputs modulate bimodal and unimodal expression of comX. *Mol Microbiol* **86**, 258–272, doi:10.1111/j.1365-2958.2012.08187.x (2012).

25. Morrison, D. A., Khan, R., Junges, R., Amdal, H. A. & Petersen, F. C. Genome editing by natural genetic transformation in *Streptococcus mutans*. *J Microbiol Methods* **119**, 134–141, doi:10.1016/j.mimet.2015.09.023 (2015).
26. Wenderska, I. B. *et al.* Transcriptional Profiling of the Oral Pathogen *Streptococcus mutans* in Response to Competence Signaling Peptide XIP. *mSystems* **2**, doi:10.1128/mSystems.00102-16 (2017).
27. Besingi, R. N. *et al.* Functional Amyloids in Streptococcus mutans, their use as Targets of Biofilm Inhibition and Initial Characterization of SMU_63c. *Microbiology*, doi:10.1099/mic.0.000443 (2017).
28. Dufour, D. & Levesque, C. M. Cell death of *Streptococcus mutans* induced by a quorum-sensing peptide occurs via a conserved streptococcal autolysin. *J Bacteriol* **195**, 105–114, doi:10.1128/JB.00926-12 (2013).
29. Morrison, D. A. & Baker, M. F. Competence for genetic transformation in pneumococcus depends on synthesis of a small set of proteins. *Nature* **282**, 215–217 (1979).
30. Chang, J. C., LaSarre, B., Jimenez, J. C., Aggarwal, C. & Federle, M. J. Two group A streptococcal peptide pheromones act through opposing Rgg regulators to control biofilm development. *PLoS Pathog* **7**, e1002190, doi:10.1371/journal.ppat.1002190 (2011).
31. Petersen, F. C., Fimland, G. & Scheie, A. A. Purification and functional studies of a potent modified quorum-sensing peptide and a two-peptide bacteriocin in *Streptococcus mutans*. *Mol Microbiol* **61**, 1322–1334, doi:10.1111/j.1365-2958.2006.05312.x (2006).
32. Hossain, M. S. & Biswas, I. An extracelluar protease, SepM, generates functional competence-stimulating peptide in *Streptococcus mutans* UA159. *J Bacteriol* **194**, 5886–5896, doi:10.1128/JB.01381-12 (2012).
33. Lau, P. C., Sung, C. K., Lee, J. H., Morrison, D. A. & Cvitkovitch, D. G. PCR ligation mutagenesis in transformable streptococci: application and efficiency. *J Microbiol Methods* **49**, 193–205 (2002).
34. Miller, C. L. *et al.* RsmW, *Pseudomonas aeruginosa* small non-coding RsmA-binding RNA upregulated in biofilm versus planktonic growth conditions. *BMC Microbiol* **16**, 155, doi:10.1186/s12866-016-0771-y (2016).
35. Yu, W. H., Hovik, H. & Chen, T. A hidden Markov support vector machine framework incorporating profile geometry learning for identifying microbial RNA in tiling array data. *Bioinformatics* **26**, 1423–1430, doi:10.1093/bioinformatics/btq162 (2010).
36. Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J. & Holmes, I. H. JBrowse: a next-generation genome browser. *Genome Res* **19**, 1630–1638, doi:10.1101/gr.094607.109 (2009).

## Acknowledgements

## Author Contributions

R.K., R.J., D.A.M. and F.C.P. designed the experiments. R.K., R.J., H.Å., D.A.M. and F.C.P. performed the experiments. T.C., data curation. R.K., R.J., T.C., D.A.M. and F.C.P. performed data analysis. All authors participated in writing and reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-04768-5

**Competing Interests:** The authors declare that they have no competing interests.

**Accession number for the RNA-sequencing data:** Bioproject PRJNA378429 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA378429

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.