Supporting Information

Intracellular signaling through the *comRS* system in *Streptococcus mutans* genetic competence

Simon A.M. Underhill, Robert C. Shields, Justin R. Kaspar, Momin Haider, Robert A. Burne and Stephen J. Hagen*

List of supporting information, figures and tables:

- Figure S1: Measurement of *comS*, *comX* and *comR* transcript levels
- Figure S2: Response of cocultures is time-independent
- Modeling the fluorescence polarization binding data
- Figure S3: Effect of histidine tag on ComR binding of ComS and XIP
- Deterministic fit: Equation system and calculated parameter values
- Robustness analysis for model
- Table S1: Parameters for gamma distribution fits to single cell P*comX* GFP fluorescence distributions in microfluidic experiments
- Table S2: Fitted values for the 12 parameters of the model and statistical measurement of their robustness from bootstrap process
- Table S3: RT-qPCR primer sequences

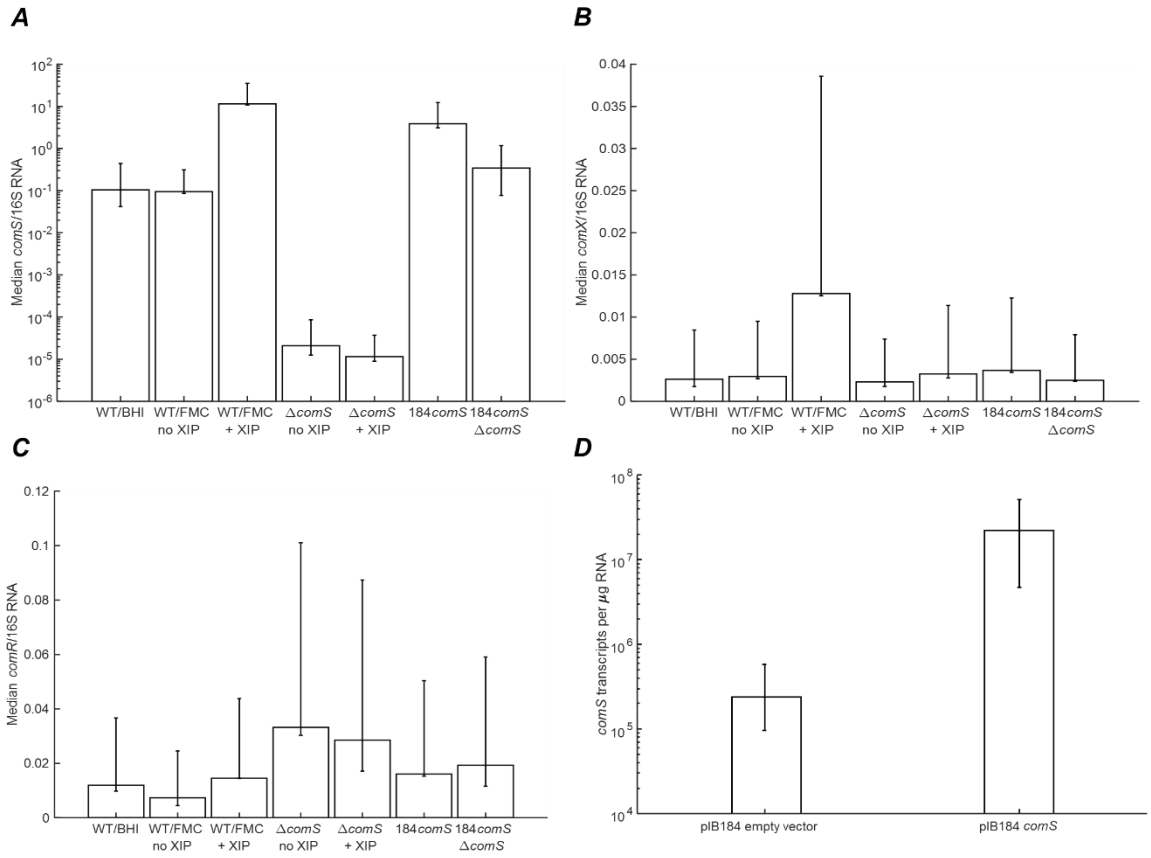

Fig. S1: Measurement of *comS*, *comX* and *comR* transcript levels

RT-qPCR measurement of (A) *comS,* (B) *comX* and (C) *comR* transcripts in cultures harvested at OD_{600} = 0.5. Each bar indicates the ratio of the median transcript count to the median 16S rRNA count, as measured in multiple biological and technical replicates (see *Methods*). WT/BHI, 184*comS*, and 184*comS* Δ*comS* samples were grown in BHI medium. Remaining samples were grown in FMC. Synthetic XIP was supplied at 100 nM concentration when used. Error bars indicate the range from second lowest to second highest value among the replicates for each condition. (*D*) Comparison of *comS* mRNA transcripts from *S. mutans* harboring the ComS overexpression plasmid versus the empty plasmid.

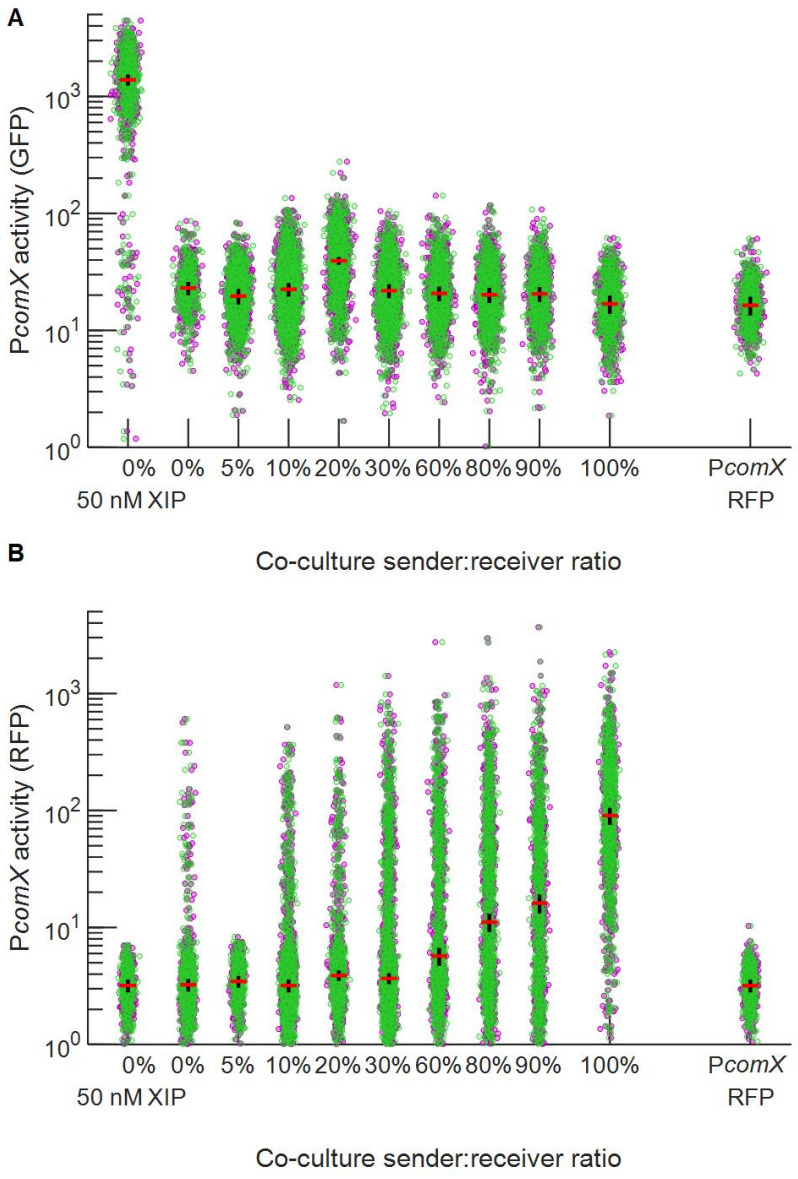

Figure S2 – Response of cocultures is time-independent

(A) GFP (*comX* reporter) and (B) RFP (*comY*) fluorescence of individual cells in co-cultures of sender (184comS P*comX-rfp*) and receiver (P*comX-gfp* Δ*comS*) strains of *S. mutans*. Samples are labeled by percentage by volume of 184*comS* (sender) culture in the initial preparation of the coculture. Fluorescence was measured immediately (0 h, green) after mixing, or 4 h (magenta) after mixing the coculture. The red horizontal bars

show the median fluorescence immediately after mixing (0 h); the black vertical bars show the median at 4 h. Data are from the coculture experiment of Fig. 3.

Modeling the fluorescence polarization binding data

We compared the FP data to a two-step binding model in which the peptide ComS or XIP forms a multimeric complex with ComR (with dissociation constant k_1), and then a single copy of this complex binds to the fluorescent DNA probe (with dissociation constant k_2), increasing its fluorescence anisotropy. The model is summarized by

$$P + R \rightleftharpoons C \qquad C + D \rightleftharpoons D^* \qquad C + U \rightleftharpoons U^*$$

Here *P* is the peptide (ComS or XIP), *R* is ComR, *C* is the peptide-ComR multimeric complex, *D* (*U*) is the free labeled (unlabeled) probe, *D** (*U**) is the labeled (unlabeled) probe with complex bound. The order of multimerization of the complex *C* is *n*. We solved the equilibrium equations for the model using the multivariate Newton-Raphson method in Matlab. We performed separate data analyses for the FP data ComS and XIP, respectively. In each analysis we searched for parameter values (k_1, k_2, n) that simultaneously minimized the sum of squares residuals for both the association (Figure 5A) and competition (Figure 5B) experiments for a given peptide *P*.

In general the FP data are compatible with a range of parameter values. If *n* is constrained to be less than 2.5 then optimal values are in the range $k_1 \sim$ 1-6 μM and $k_2 \sim$ 1-30 nM and $n \simeq$ 2-2.5 for for XIP interacting with ComR, and $k_1 \sim$ 3-20 μM and $k_2 \sim$ 30-200 nM and $n \simeq$ 1.6-2.5 for ComS interacting with ComR.

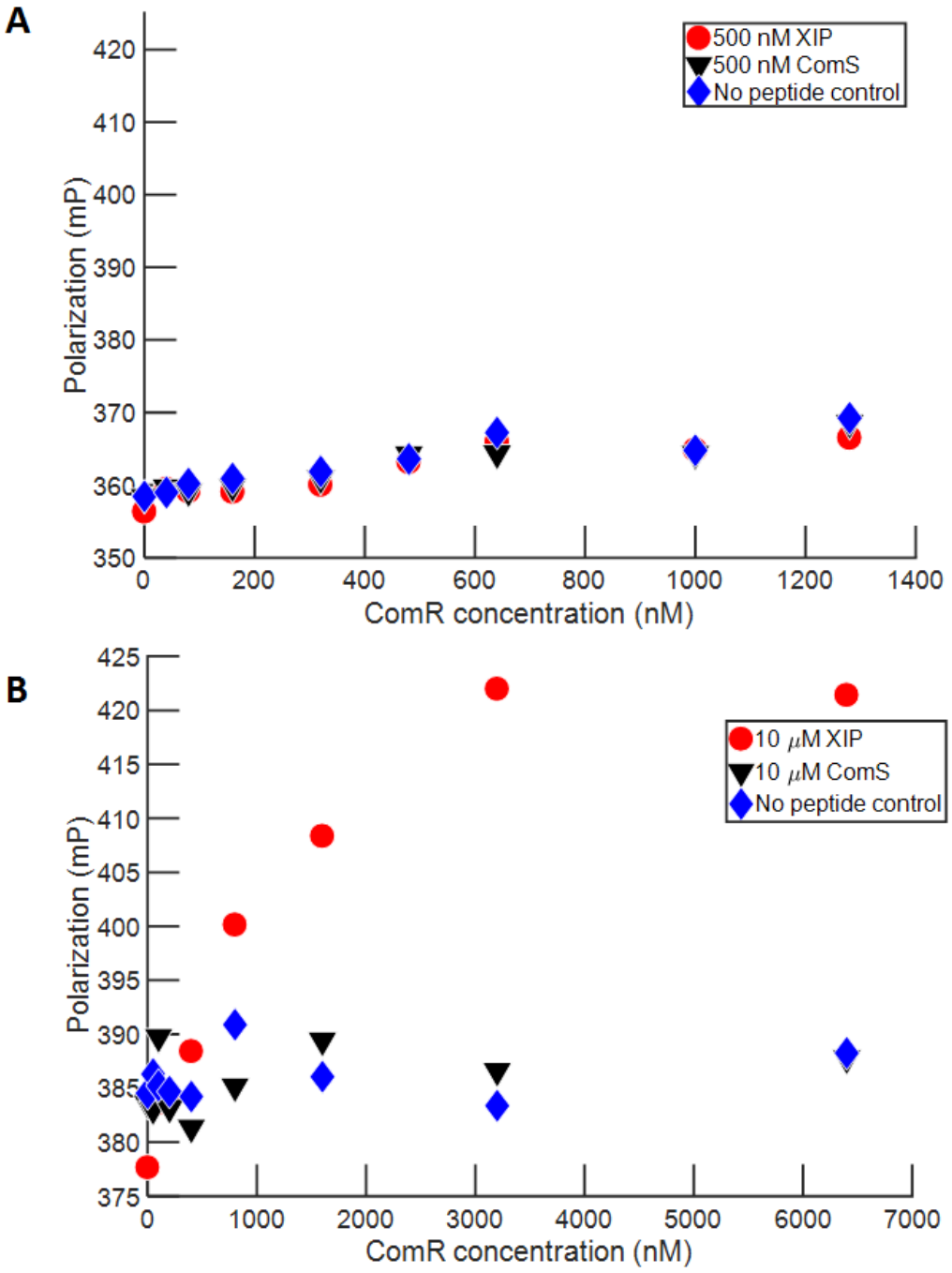

Figure S3: Effect of histidine tag on ComR binding of ComS and XIP

Fluorescence polarization study of ComS and XIP interaction with ComR that was N- or C-terminally tagged with 6X-histidine. Polarization is plotted versus [ComR] for (A) C-terminally tagged ComR and (B) N-terminally tagged ComR. In each case 1 nM

fluorescent DNA and 0.05 mg ml^{-1} salmon DNA were present along with no signal peptides (black), XIP (blue) or ComS (green). The *comR* gene (SMu.61) was amplified using gene-specific primers (forward, AAAGAATCCTATGTTAAAAGA; reverse, CACCCTAGGAGACCCATCAAA) and cloned into the pET45b(+) vector bearing an N-terminal 6xHis tag. The resulting vector, pET45b(+)his-comR_{UA159}, was transformed into *E. coli* 10-beta. After sequencing confirmed the correct insertion (using T7 promoter and T7 terminator primers), the vector was transformed into *E. coli* BL21(DE3) prior to protein purification.

Deterministic fit: Equation system and calculated parameter values

A mathematical model (below) was constructed for the ComRS activation of *comX* with intracellular feedback, in the presence of extracellular XIP, and this model was used to fit the *comX* activation data in Fig. 1. The same offset and multiplicative factor were used to map calculated [ComX] onto the GFP fluorescence curves for both strains, as this is a property of the GFP, not the gene circuit. To model the *comS*-deficient strain, parameters representing *comS* feedback and constitutive production were set to zero. ComR was assumed to be present at around 15 copies per cell, as only modest changes in its expression result from early competence inducing factors (54). Exogenous XIP was taken to be a non-depleting reservoir. The system of ODEs used to fit microfluidic data is given below. X represents ComX, Z the internal XIP concentration, S the internal ComS concentration, R the (constant) ComR concentration and Exo the exogenous XIP level. All units are in nM and seconds where appropriate. Other symbols are parameters describing the reaction kinetics. A star indicates one of the *V*

parameters contributing to feedback, while unstarred *V*s indicate a maximum rate of production of ComX. Hill kinetics corresponding to inferred cooperativity from FP assays were used. Calculated parameters are given in Table S2. A 200-iteration bootstrap analysis of the data was performed in order to estimate parameter robustness, with the 10th and 90th percentiles of parameter values reported. These percentile values demonstrate preservation of the relative order of magnitude between dissociation constants for XIP-ComR and ComS-ComR.

$$\frac{dX}{dt} = \frac{V_1 R^2 Z^2}{R^2 Z^2 + K_X^4} + \frac{V_2 RS}{RS + K_S^2} - \beta_X X \quad (1)$$

$$\frac{dS}{dt} = \alpha_0 - \beta_S S - \gamma S + \frac{V_1^* R^2 Z^2}{R^2 Z^2 + K_X^4} + \frac{V_2^* RS}{RS + K_S^2} - \frac{V_2 RS}{RS + K_S^2} \quad (2)$$

$$\frac{dZ}{dt} = J(Exo - Z) - \beta_Z Z + \gamma S - \frac{V_1 R^2 Z^2}{R^2 Z^2 + K_X^4} - \frac{V_1^* R^2 Z^2}{R^2 Z^2 + K_X^4} \quad (3)$$

Robustness analysis for model

Robustness of fit was tested through the bootstrap method, using the 90th and 10th percentile behavior of parameters to examine whether the transcriptional efficiency difference hypothesized was preserved in this range. Dependence on the initial parameter guess was checked by 50 iterations of adding a Gaussian-distributed random number with a mean of the best fit parameter and standard deviation half the best fit parameter to the start guess vector components used to find the best fit. New sets of fit parameters for each of these were then generated. It was found that the ComS-ComR complex elicited higher *comX* transcription in 100% of cases than did the XIP-ComR complex, and higher *comS* feedback stimulation (V* parameters) in 78% of cases. Thus

while numerous solutions to the system exist, the *comX* transcriptional efficiency discrepancy hypothesized is a generic property of the fit.

Table S1: Parameters for gamma distribution fits to single cell P*comX* GFP fluorescence distributions in microfluidic experiments.

[XIP]	*a*	*b*
Wild type	-	-
0	4.09	5.51
280	5.50	172
700	9.62	153
1840	11.0	184
3250	12.5	168
5230	12.4	174
6000	11.7	172
Δ*comS*	-	-
0	6.14	2.29
30	7.10	2.44
850	4.24	166
940	5.11	148
3000	7.44	160
4020	7.61	172
6000	7.45	180

Table S2: Fitted values for the 12 parameters of the model and statistical measurement of their robustness from bootstrap process.

Parameter	Best fit value – used for Fig. 7	10th percentile from bootstrap	90th percentile from bootstrap	Units
α_0	7.85	1.78	19.1	nM s^{-1}
β_S	7.17 x 10^{-3}	7.97 x 10^{-4}	1.32 x 10^{-2}	s^{-1}
γ	0.452	0.227	2.74	s^{-1}
V_1^*	1.07 x 10^4	3.8 x 10^3	2.46 x 10^4	nM s^{-1}
V_2^*	1.33 x 10^4	3.22 x 10^3	2.65 x 10^4	nM s^{-1}
K_x	148	42	210	nM
K_S	2740	849	3000	nM
V_1	3.55	0.988	558	nM s^{-1}
V_2	778	309	2190	s^{-1}
J	9.39	6.14 x 10^{-2}	17.6	s^{-1}
β_Z	1.28	0.264	3.07	nM s^{-1}
β_X	10.5	1.71	16.2	s^{-1}

Table S3: RT-qPCR primer sequences

Gene and direction	Primer sequence
comX forward	5’-CGTCAGCAAGAAAGTCAGAAA C-3’
comX reverse	5’-ATACCGCCACTTGACAAACAG-3’
comS forward	5’-TCAAAAAGAAAGGAGAATAACA-3’
comS reverse	5’-TCATCTGAGATAAGGGCTGT-3’
comR forward	5’-TATTACGAAGGCCAACCTAT-3’
comR reverse	5’-TTCTTCTTCAGGCAAATGAT-3’
16S rRNA forward	5’-CACACCGCCCGTCACACC-3’
16S rRNA reverse	5’-CAGCCGCACCTTCCGATACG-3’