www.sciencemag.org/cgi/content/full/1140818/DC1

Supporting Online Material for

Noise in Gene Expression Determines Cell Fate in *Bacillus subtilis*

Hédia Maamar, Arjun Raj, David Dubnau*

*To whom correspondence should be addressed. E-mail: dubnau@phri.org

Published 14 June 2006 on *Science* Express
DOI: 10.1126/science.1140818

This PDF file includes:

Other Supporting Online Material for this manuscript includes the following:
available at www.sciencemag.org/cgi/content/full/1140818/DC1

Supplementary online material for "Noise in gene expression determines cell fate in *Bacillus subtilis*"

Hédia Maamar[1†], Arjun Raj[2†] and David Dubnau[1]*

[1]Public Health Research Institute Center-New Jersey Medical School, 225 Warren Street, Newark, NJ 07103, USA

[2]Department of Physics, Massachussetts Institute of Technology, 77 Massachussetts Avenue, Cambridge, MA 02139, USA

*To whom correspondence should be addressed; E-mail: dubnau@phri.org

[†] Contributed equally to this study

Materials and Methods

The *B. subtilis* strains used are all derivatives of strain 168 (BD630) and are described in Table 1. Bacteria were grown either in liquid competence minimal medium (CMM) (1) or on Luria-Bertani (LB) agar supplemented with chloramphenicol, erythromycin or kanamycin (5 μg/ml) or spectinomycin (100 μg/ml). *B. subtilis* competent cells were prepared as described previously (1). *E. coli* DH5a (Invitrogen) was used for cloning. DNA manipulation and other molecular biological procedures were performed using standard protocols.

Strain and Plasmid Construction

comK-9aa-cfp fusion*

The *cfp** (codon optimized for *B. subtilis*) gene including the open reading frame was synthesized by polymerase chain reaction (PCR) using Primer 1 and Primer 2 (Table 2) and the plasmid pDR200 (gift from Dr. David Rudner, Harvard Medical School) as template. The product was cut with *Xho*I *and Bss*HII ligated to pED359, previously digested with the same enzymes and transformed into *E. coli* DH5a, with selection for ampicillin resistance. pED359 was derived from pFG10 (a gift from F. Guerrero and R. Losick) and contained a *comk* sequence that bears the *comK* promoter, the ribosome binding site and the first 9 codons of the *comK* open reading frame (ORF) fused to *cfp*. The resulting plasmid was pED1021. pED1021 was used to transform BD630 with selection for kanamycin resistance by Campbell-like transformation, placing P*comK-9aa-cfp** under control of the *comK* promoter and leaving the endogenous copy of *comK*

unaltered, thus producing BD4374. This plasmid was also used to transform BD2955 (*rok*) to produce BD4375.

comK-96aa-cfp fusion*

A *comk* sequence including the promoter region and the first 96 codons of the *comK* ORF was synthesized by PCR using Primer 3 and Primer 4, and BD630 genomic DNA as template. The product was cut with *EcoR*I and *Xho*I and ligated to pED1021 digested with the same restriction enzymes. The resulting plasmid (pED1022) was transferred in BD630 as described above generating BD4376. This plasmid was also used to transform BD2955 (*rok*), producing BD4377.

comK-M2 fusion

A DNA fragment containing 32 head-to-tail tandem repeats of the 50 nucleotide sequence 5'-CAGGAGTTGTGTTTGTGGACGAAGAGCACCAGCCAGCTGATCGACCTCGA-3', was cut from pGEMT-M2 (2,3) using *EcoR*I and *BamH*I, and was cloned into the same sites of the pAC5 plasmid (4), producing the plasmid, pED1025. The *comK* sequence containing the promoter and 9 codons of the ORF was obtained by PCR with Primer 3 and Primer 5, then cut with *EcoR*I and ligated to pED1025 digested with *EcoR*I. The resulting plasmid (pED1026) was used to transform BD630, BD4374*, and* BD4375 with selection for chloramphenical resistance. The transformants were screened on starch plates, and then exposed to I_2, to confirm the insertion of the construct at the *amyE* locus. The resulting strains were BD4378, BD4379, and BD4380, respectively.

comK deletion

A 734 bp fragment upstream of the *comK* ORF was synthesized by PCR using Primers 6 and 7, and then cloned into pUC19 at the *Acc*65I and *Sma*I sites. The resulting plasmid

was digested with *BamH*I and ligated to the *erm* gene cut with *BamH*I from the plasmid pED209 (5), producing pUCComKupERM. A 778 bp fragment downstream from *comK* was produced by PCR with Primers 8 and 9, digested with *Sal*I and *Sph*I and cloned into pUCComKupERM cut with the same enzymes. The resulting plasmid, pED1027, was used to transform BD630 and BD4378 by double crossing over, deleting *comK* totally and replacing it with *erm*, producing BD4381 and BD4382, respectively. The replacement was checked by the loss of transformation ability of the strain and by PCR, using BD630 (control), BD4381 and BD4382 genomic DNAs as templates with Primers 3 and 9.

Construction of the low-noise strain

The initiation codon ATG in the *comK-96aa-cfp* construct was mutated to GTG by PCR mutagenesis using Primers 10 and 11, and the plasmid pED1022 as template. The resulting plasmid was confirmed by sequencing and was used to transform BD2955 with selection for kanamycin resistance by Campbell-like transformation. The genomic DNA of several individual transformants was extracted and used as templates for PCR with Primer 3 and 12. The PCR products were purified and sequenced. Only clones having the mutation in the initiation codon of the full-length *comK* were selected for further characterization. The selected strain (BD4383) was transformed with BD4378 genomic DNA, to introduce *comK-M2*, with selection for chloramphenicol resistance, leading to BD4384.

Probes for in situ hybridization

The probes used for the in situ hybridization were DNA oligonucleotides synthesized on an Applied Biosystems 394 DNA synthesizer using mild phosphoramidites (Glen Research). The oligonucleotide sequences were

P1: 5'-GGCAATTGRTGCGCCGTRCACTTCARACGATTC–3';

P2: 5'-CGAAGGGRCCACCATGARCGGCGGCTTGRGTGAAAT–3';

P3: 5'-ATAAGAGARCGGCAGCRCCATCGTTTRCCCATTG –3';

P4: 5'-TTTGGTTCRGAGCCACGCRGTTCGGTARACCTGGTT-3';

P5:5'-CCGCTCTRCTTTCGGGRACAGCARAAATTCC-3';

P6:5'-ATACCGTRCCCCGAGCRCACGCAAAARAAAATCA-3',

PM2:5'-RCGAGGTCGARCAGCTGGCTGGRGCTCTTCGRCCACAAACA-3' where P1, P2, P3, P4, P5 and P6 are complementary to *comK* messenger and PM2 is complementary to the repeated sequence in M2. The "R"s represent locations where an amino-dT was introduced in place of a regular dT. The oligonucleotides were synthesized on a controlled pore glass column (Glen Research) that introduced an additional amine group at the 3' end of the oligonucleotide and an additional amine group was added to the 5' in the final synthesis step. The amine groups of P1-P6 were then coupled to tetramethylrhodamine (Molecular Probes) to create the following probes: P1-TMR, P2-TMR, P3-TMR, P4-TMR, P5-TMR, P6-TMR. These six probes were quantified and combined at a concentration of 50 ng/μl each, creating the probe pool C6-TMR. The amine groups of PM2 were coupled to the Alexa 594 fluorophore (Molecular Probes), yielding PM2-Alexa-594. All probes were individually purified on an HPLC column to isolate oligonucleotides displaying the highest degree of coupling of the fluorophore to

the amine groups.

Preparation of cells for microscopy

To measure the number of competent cells throughout stationary phase, the strains were grown in CMM and samples were taken throughout growth for microscopy. Cells were attached to polylysine-coated slides. Cross walls were visualized by staining nucleic acids with propidium iodide (10 μg/ml). Samples were mounted in Slow Fade (Molecular Probes) for imaging.

For time-lapse microscopy, cells were grown in CCM to mid-log, then pelleted and resuspended in conditioned CMM, One or two microliters of cell suspension were layered onto a 20 μl pad of 1% agarose dissolved in conditioned CMM, covered with a cover slip that was then sealed to the slide. Bright field and fluorescence pictures were taken to follow cell growth on the pad and competence development. Temperature was maintained at near 35 °C.

For the *in situ* hybridization experiments, cells were fixed with 10% formaldehyde for 10 min at room temperature, washed with PBS, and then allowed to attach to polylysine-coated multichambered coverglasses (Lab-Tek, Nalgene Nunc). The cells were permeabilized by treatment with lysozyme (2 mg/ml) for 2 min and then washed with PBS. FISH was then performed using combinations of probe pool C6-TMR or PM2-Alexa 594 at a concentration of 1 ng/μl each, following the procedure outlined in Femino et al. (6). The optimal level of formamide used during hybridization and washing for maximum signal to background was empirically determined to be 25% to detect *M2*

messenger and 10% to detect the *comK* messenger.

Microscopy

Routine microscopy was performed with an upright Nikon Eclipse 90i microscope equipped with an Orca-ER Digital Camera (Hamamatsu), and a Nikon TIRF 1.45 NA Plan Neo-Fluor 100 X oil immersion objective. Velocity software (Improvision) and Adobe Photoshop were used for image acquisition and processing. Three images were captured for each field of cells: DIC and fluorescence images for CFP and PI. Appropriate Semrock Optical filter sets were used for each fluorophore.

After *in situ* hybridization, cells were imaged using an Axiovert 200M inverted fluorescence microscope (Zeiss), equipped with a 100X oil-immersion objective, a CoolSNAP HQ camera (Photometrics, Pleasanton, California, United States), cooled to -30 °C; and standard filter sets obtained from Omega Optical. Openlab acquisition software (Improvision) was used to acquire the images. For three-dimensional imaging, randomly chosen fields were imaged by taking adjacent Z-axis optical sections that were 0.3 μm apart.

Image analysis

Custom software written in MATLAB (The Mathworks, Natick, MA) was used to analyze the fluorescence *in situ* hybridization images. Broadly, the method consisted of first locating individual cells from the differential image contrast image using morphological operations to reduce non-uniformities in the background, after which a

simple threshold was used to separate the cells. Incorrectly segmented cells were discarded based on size and shape. For the fluorescence images, individual molecules were isolated by first running a median filter to remove shot noise, after which a custom linear filter loosely based on the discrete Laplacian (3) was applied to enhance particular signals. Using the cell boundaries and locations determined from the DIC image, the number of mRNA in each cell was counted by using a simple threshold to distinguish particles. Since the number of *comK* mRNA molecules was relatively low, overlapping mRNAs did not pose much of a concern, and the cell boundaries were expanded slightly to account for possible frame shifts between the DIC and fluorescent images. To exclude the competent cells from the determination of the mean and variance of the number of mRNA molecules, we manually set a threshold for the CFP and TMR fluorescence; cells above either of those fluorescence levels were deemed competent. Thresholds for both channels were required because when cells initially become competent, the CFP has not yet had time to fold, but *comK* mRNA are in high abundance (>20 per cell), thus allowing one to determine which cells are competent using the TMR channel. Later in stationary phase, though, the mRNA signal vanishes due to an unknown *comK* shutdown mechanism; however, the CFP signal remains, allowing one to still distinguish competent cells from non-competent ones.

Spectrofluorimetry

The strains carrying the *comK-96aa-cfp** fusion were grown in competence medium in duplicate and samples were taken throughout growth. Cells were pelleted, washed once and resuspended in Spizizen salts (7). The total fluorescence in the samples was

measured using a spectrofluorimeter (Photon Technology International) using 437 nm as the exciting wavelength and 475 nm as the emission wavelength. In each experiment the background fluorescence was measured using BD630 (wild type strain without *cfp*) at different growth stages and subtracted from the values obtained for the strains containing the fusions. All measured values were normalized to the optical densities values, as a measure of cell density. The mean CFP fluorescence values for the wild type, *rok* and low noise strains were 148, 408, 274 at $T_{-0.5}$ and 667, 1083, 856 at T_0, respectively.

Table 1: Strains

Strain	Genotype	Source
BD630	*his leu met*	Laboratory strain
BD2955	*his leu met rok::spc*[a]	Tran et al 2001
BD4374	*his leu met comK-9aa-cfp* (CBL*[b]*, kan*[a]*)*	This study
BD4375	*his leu met rok::spc*[a] *comK-9aa-cfp* (CBL*[b]*, kan*[a]*)*	This study
BD4376	*his leu met comK-96aa-cfp* (CBL*[b]*, kan*[a]*)*	This study
BD4377	*his leu met comK-96aa-cfp* (CBL*[b]*, kan*[a]*) rok::spc*[a]	This study
BD4378	*his leu met amyE:: comK-M2 (cat*[a]*)*	This study
BD4379	*his leu met amyE:: comK-M2 (cat*[a]*) comK-9aa-cfp* (CBL*[b]*, kan*[a]*)*	This study
BD4380	*his leu met amyE:: comK-M2 (cat*[a]*) comK-9aa-cfp* (CBL*[b]*, kan*[a]*) rok::spc*[a]	This study
BD4381	*his leu met ΔcomK::erm*[a]	This study
BD4382	*his leu met amyE:: comK-M2 (cat*[a]*) ΔcomK::erm*[a]	This study
BD4383	*his leu met comK*[c] *comK-96aa-cfp* (CBL*[b]*, kan*[a]*) rok::spc*[a]	This study
BD4384	*his leu met amyE:: comK-M2 comK*[c] *comK-96aa-cfp* (CBL*[b]*, kan*[a]*) rok::spc*[a]	This study

[a]. *kan, cat, erm* and *spc* stand for resistance to kanamycin, chloramphenicol, erythromycin and spectinomycin, respectively.

[b] Inserted by Campbell like integration

[c] a mutated allele of *comK* in which the start codon ATG was switched to GTG

* codon optimized version of CFP to use in *B. subtilis*.

Table 2. Primers

Name	Sequence[a]	Use
Primer 1	5'-CC **CTCGAG** GGA ATG GTT TCA AAA GGC GAA GAA C-3'	*comK-9aa-cfp** fusion
Primer 2	5'-CG **GCGCGC** TTAC TTA TAA AGT TCG TCC-3'	*comK-9aa-cfp** fusion
Primer 3	5'-GAC ATC **GAATTC** TTT TGT T-3'	*comK-96aa-cfp** fusion
Primer 4	5'-TCA G **CTCGAG** CG AAG AAA GTG GGA ATA AAA AG-3'	*comK-96aa-cfp** fusion
Primer 5	5'-G**GAATTC CCCGGG TTCAAG** CTT AGG TGC CTC TGT TTT CTG ACT-3'	*comK-M2*
Primer 6	5'-CG **GGTACC** G CTT CAA TTA CCG CCT ATA AAG-3'	*comK* deletion
Primer 7	CAT AAG ACT TGC CGG TTT ACG	*comK* deletion
Primer 8	5'-GC **GTCGAC** G AGA TGT CAC ATA GCT TGA ATT C	*comK* deletion
Primer 9	5'-CAT **GCATGC** CAA AAG AAC ATG AAG CCT GTC-3'	*comK* deletion
Primer 10	5'-GGC CAT AAT <u>GTG</u> AGT CAG AAA ACA GAC GCA CCT-3'	*comK* mutagenesis
Primer 11	5'-AGG TGC GTC TGT TTT CTG ACT <u>CAC</u> ATT ATG GCC-3'	*comK* mutagenesis
Primer 12	5'-ACAT**GCATGC**ATTGACATCTCAGGTATATGG-3'	Check *comK* deletion

[a] Restriction sites are in bold and mutagenic codons are underlined.

Stochastic model of competence initiation

In this supplement, we describe a model of competence initiation in *B. subtilis*. Our modeling effort had two specific goals: 1) to show that modulating the rate of *comK* transcription can significantly alter the rate of transitions to the competent state, thus explaining the "window of opportunity", and 2) showing that reducing the noise in the number of ComK proteins while keeping the mean the same (by increasing the transcription rate and decreasing the translation rate) can lead to a dramatic decrease in the rate of transitions to the competent state. Our model is able to reproduce both of these findings, thus presenting a simple unified model of competence initiation.

We first summarize the basic features of competence germane to any model of this behavior, after which we formulate the model mathematically and explore its behavior through the use of stochastic simulations.

Summary of competence behavior

The core of the network that controls the onset of competence in *B. subtilis* is shown in Figure 1A. We will define time T_x as amount of time x from the entry into stationary phase. Essentially, at time T_0, the concentration of ComS proteins in the cells is dramatically increased, largely due to the accumulation of the pheromone ComX in the growth medium (8). Since both ComK and ComS are degraded by the MecA/ClpC/ClpP complex, the increased numbers of ComS proteins effectively reduces the degradation rate of ComK through competition. The increased overall amount of ComK protein allows for stochastic transitions from the non-competent state to the competent state (Figure 1C). This is due to positive feedback in the form of ComK stimulating its own production by binding to its promoter in a cooperative manner (9); the cooperativity is required for the establishment of a bistable system (10). These transitions continue for around 2 hours, at which point the transitions stop occurring and cells that are not yet competent are no longer able to become competent

while those cells which are competent remain competent.

Once a cell transitions to the competent state, it produces a very large amount of ComK, ultimately on the order of 50,000 protein dimers (11). This huge amount of protein keeps the cell in the competent state until the cell is diluted into fresh medium. At that time, ComK is degraded at a constant rate before the cell may resume exponential growth (11). This linear decrease indicates that the MecA/ClpC/ClpP complex is saturated, corroborated by estimates of the abundance of MecA/ClpC/ClpP protease at only around 300 per cell (11).

Since cells are unable to transition from the competent state to the non-competent state but the final percentage of competent cells is well below 100%, it is necessary to assume that there is a "window of opportunity" during which cells can transition into the competent state; otherwise, cells would continue to transition to competence until all the cells are competent. This window of opportunity appears to coincide with a decrease in the mean number of *comK* mRNAs in non-competent cells as the cells continue into stationary phase (Figure 2D). A natural inference is that the cells are producing too few mRNAs to allow for the transitions to occur, as not enough ComK protein is around to induce transitions to the competent state. We address this assertion quantitatively later in this supplement.

Although we do not directly address the issue of the exit from competence in this work, there are a few points about ComK dynamics after competence is initiated that are relevant for our model, in particular concerning the possibility of transitions from the competent state to the non-competent state. As stated previously, under our experimental conditions, cells are unable to transition from the competent state to the non-competent state, at least during the time-scales we observe. Most evidence shows that competent cells only leave competence upon dilution into fresh medium, at which point ComK is degraded to non-competent levels over a period of around 1.5 hours, after which the cells resume exponential growth.

Furthermore, we found that *comK* transcription was turned off once the cells had been competent for a certain length of time. This indicates that *comK* expression is inactivated once ComK reaches very high levels. The mechanism by which this occurs may either be the

same as the mechanism that reduces the transcription of comK in the non-competent cells observed in Figure 2D, or it may be some other mechanism specific to *comK* regulation in competent cells. It is also important to note that the levels of ComK are extremely high at this point (∼50,000 dimers per cell). This is likely due to the fact that ComK production continues unchecked, as the MecA/ClpC/ClpP degradation mechanism quickly saturates at the higher level of ComK production instigated upon activation of the positive feedback loop. Of course, the cell is not capable of producing infinitely much ComK protein, so another mechanism must eventually shut down production of ComK. Although the mechanism responsible for this shut down is unknown, we found some evidence that it acts by shutting down transcription of the *comK* gene. When comparing the level of *comK* mRNA and the amount of fluorescence emitted by CFP under control of the *comK* promoter at different times during stationary phase, we found that at the start of stationary phase, some cells had large levels of mRNA fluorescence without yet having much CFP fluorescence, indicating that the cells had already triggered the positive feedback loop but had not yet produced appreciable amounts of fluorescent CFP, with the lag presumably due to the folding time of the fluorescent protein. However, as time progressed in stationary phase, we found that the cells with large amounts of CFP had little fluorescent mRNA signal. This indicates that competent cells eventually turn off transcription of the *comK* gene once they have a large number of ComK protein molecules[1], although this is not to say that the high levels of ComK proteins are necessarily responsible for the shutdown in *comK* transcription. We make no attempt to include this mechanism in our model, since we are primarily concerned with the initiation of competence rather than the behavior of the cell once competence is underway. This mechanism is, however, relevant in the sense that we assume that this unknown mechanism prevents runaway ComK production, so our model does not have to include the ComK degradation mechanism stabilizing the "on" state as is required in most

[1]This is a different mechanism than that proposed by Suel et al. (12), in which *comK* transcription is shut down by deactivation of the *comK* positive feedback loop, which is only possible upon degradation of ComK to non-competent levels.

models of bistable gene expression (10). This allows for a much broader set of parameters admitting random competence initiation, since the parameters governing the degradation rate in the non-competent cells (through the MecA/ClpC/ClpP degradation complex) only need to stabilize the "off" state.

Model

Here, we formalize our stochastic model for competence initiation in *B. subtilis*. Since even the most basic model of competence must necessarily involve a large number of unknown parameters, we focused our efforts on creating the simplest model that recreated the experimental behavior observed and then sought to alter the network in such as way as to find novel qualitative behavior that could be tested experimentally.

The key features we incorporated were the MecA/ClpC/ClpP degradation mechanism and the ComK positive feedback loop. The MecA/ClpC/ClpP degradation mechanism was modeled by allowing both ComK and ComS to bind to MecA/ClpC/ClpP, followed by degradation of the protein. ComS is also assumed to have another MecA/ClpC/ClpP independent degradation mechanism (12).

For ComK, we first assumed that ComK dimerization was essentially instantaneous, allowing us to assume that all ComK proteins were in the dimeric state. Since this is the functional form of the ComK protein, we henceforth use the term ComK to refer to the ComK dimer. To model the positive feedback loop, we utilized the known fact that ComK binds at two sites to its own promoter. We incorporated cooperativity in this binding by assuming that the binding of the first ComK molecule greatly facilitated the binding of the second ComK molecule, an assertion backed up by gel-shift experiments indicating a high degree of cooperativity in the interaction of ComK with its own promoter (9).

The reactions are as follows:

$$
\begin{aligned}
\mathrm{mK} &\xrightarrow{\delta_{\mathrm{mK}}} \\
\mathrm{mK} &\xrightarrow{\mu_{\mathrm{K}}} \mathrm{mK} + \mathrm{K} \\
\mathrm{M} + \mathrm{K} &\xrightarrow{k_{\mathrm{MK}}} \mathrm{MK} \\
\mathrm{MK} &\xrightarrow{\delta_{\mathrm{MK}}} \mathrm{M} \\
\mathrm{M} + \mathrm{S} &\xrightarrow{k_{\mathrm{MS}}} \mathrm{MS} \\
\mathrm{MS} &\xrightarrow{\delta_{\mathrm{MS}}} \mathrm{M} \\
\mathrm{P}_{comK} + \mathrm{K} &\xrightarrow{\lambda_{\mathrm{K}}} \mathrm{P}_{comK,\mathrm{K}} \\
\mathrm{P}_{comK,\mathrm{K}} &\xrightarrow{\gamma_{\mathrm{K}}} \mathrm{P}_{comK} + \mathrm{K} \\
\mathrm{P}_{comK,\mathrm{K}} + \mathrm{K} &\xrightarrow{\lambda_{\mathrm{KK}}} \mathrm{P}_{comK,\mathrm{KK}} \\
\mathrm{P}_{comK,\mathrm{KK}} &\xrightarrow{\gamma_{\mathrm{KK}}} \mathrm{P}_{comK,\mathrm{K}} + \mathrm{K} \\
\mathrm{P}_{comK} &\xrightarrow{\mu_{\mathrm{mK}}} \mathrm{P}_{comK} + \mathrm{mK} \\
\mathrm{P}_{comK,\mathrm{K}} &\xrightarrow{\mu_{\mathrm{mK,K}}} \mathrm{P}_{comK,\mathrm{K}} + \mathrm{mK} \\
\mathrm{P}_{comK,\mathrm{KK}} &\xrightarrow{\mu_{\mathrm{mK,KK}}} \mathrm{P}_{comK,\mathrm{KK}} + \mathrm{mK} \\
&\xrightarrow{\mu_{\mathrm{mS}}} \mathrm{mS} \\
\mathrm{mS} &\xrightarrow{\delta_{\mathrm{mS}}} \\
\mathrm{mS} &\xrightarrow{\mu_{\mathrm{S}}} \mathrm{mS} + \mathrm{S} \\
\mathrm{S} &\xrightarrow{\delta_{\mathrm{S}}}
\end{aligned}
$$

The symbols are defined in the following way: K is ComK protein (in the dimer form), mK is *comK* mRNA, M is the MecA/ClpC/ClpP protease, MK is the MecA/ClpC/ClpP complex when bound to a molecule of ComK, mS is *comS* mRNA, S is ComS protein, MS is the MecA/ClpC/ClpP complex when bound to a molecule of ComS, P_{comK} is the promoter of ComK without any ComK molecules bound, $\mathrm{P}_{comK,\mathrm{K}}$ is the promoter of ComK with one of

the ComK binding sites occupied, and $P_{comK,KK}$ is the promoter of ComK with both ComK binding sites occupied.

The rates are defined in Table 3. When possible, rates were chosen to be close to physical estimates. The rate of mRNA degradation was set to correspond to a half-life on the order of 1.5 minutes. The rate of *comK* transcription was then varied to give something close to the experimentally determined mean number of mRNA per non-competent cell (for instance, a basal transcription rate of 0.008 molecules^{-1} sec^{-1} would lead to a mean of 1 mRNA per cell). The rate of translation was chosen to yield an average of $\mu_{\mathrm{K}}/\delta_{\mathrm{mK}} = 125$ ComK proteins translated per mRNA, in line with what might be expected from the strong Shine-Dalgarno sequence present in the *comK* gene. There is little data on the rates of the protein-protein interactions, but we were able to make an estimate of the rate at which the MecA/ClpC/ClpP complex is able to degrade ComK (δ_{MK}), since previous work has shown it takes roughly 1.5 hours for the 300 molecules of the MecA/ClpC/ClpP complex to degrade 50,000 ComK dimers. The other protein-protein and protein-promoter interactions were chosen to be fast relative to the rate of transcription, thereby reducing their stochastic effects.

Given this set of reactions, one can determine the threshold of ComK dimers beyond which cells are likely to transition to the competent state. by examining the production and degradation rates of ComK. To do so, we analyze the deterministic version of the model, computing both the production and degradation rate of K (the number of ComK dimers) as a function of K. A straightforward computation shows that the production of ComK is given by

$$\mu_{\mathrm{K}}^{\mathrm{total}} = \frac{\mu_{\mathrm{K}}}{\delta_{\mathrm{mK}}} \left(\frac{\mu_{\mathrm{mK}} + \mu_{\mathrm{mK,K}} K_1 K + \mu_{\mathrm{mK,KK}} K_1 K_2 K^2}{1 + K_1 K + K_1 K_2 K^2} \right) \tag{1}$$

where μ_{K} is the total rate of protein production of K, $K_1 \equiv \lambda_{\mathrm{K}}/\gamma_{\mathrm{K}}$ is the affinity constant for the first K binding to the *comK* promoter and $K_2 \equiv \lambda_{\mathrm{KK}}/\gamma_{\mathrm{KK}}$ is the affinity constant for the second K binding.

The computation of the degradation rate is made somewhat more complex by the active degradation mechanism employed by the cell to degrade ComK. To find the degradation

Rate parameter	Value	Description
δ_{mK}	0.008 molecules^{-1} sec^{-1}	*comK* mRNA half-life of 1.5 min
μ_{K}	1 molecules^{-1} sec^{-1}	ComK translation rate
k_{MK}	0.1 molecules^{-2} sec^{-1}	ComK binding to protease
δ_{MK}	0.03 molecules^{-1} sec^{-1}	ComK "degradation" rate
k_{MS}	0.1 molecules^{-2} sec^{-1}	ComS binding to protease
δ_{MS}	0.03 molecules^{-1} sec^{-1}	ComS "degradation" rate
λ_{K}	0.1 molecules^{-2} sec^{-1}	First ComK binding to promoter
γ_{K}	226.42 molecules^{-1} sec^{-1}	First ComK dissociating from promoter
λ_{KK}	0.1 molecules^{-2} sec^{-1}	Second ComK binding to promoter
γ_{KK}	2.2642 molecules^{-1} sec^{-1}	Second ComK dissociating from promoter
μ_{mK}	0.008 molecules^{-1} sec^{-1}	Basal *comK* transcription (varies)
$\mu_{\mathrm{mK,K}}$	0.008 molecules^{-1} sec^{-1}	Basal *comK* transcription (varies)
$\mu_{\mathrm{mK,KK}}$	0.8 molecules^{-1} sec^{-1}	Activated *comK* transcription (varies)
μ_{mS}	0.33 molecules^{-1} sec^{-1}	*comS* transcription (varies)
δ_{mS}	0.008 molecules^{-1} sec^{-1}	*comS* mRNA half-life of 1.5 min
μ_{S}	0.3 molecules^{-1} sec^{-1}	ComS translation rate
δ_{S}	0.18 molecules^{-1} sec^{-1}	ComS degradation rate

Table 3: Parameters used in stochastic simulations of the competence circuit in *B. subtilis*. Parameters labeled "varies" are time-dependent as described in the text.

rate, we first write down the following system of ODEs representing the system's dynamics:

$$\frac{dS}{dt} = \mu_{\mathrm{S}}^{\mathrm{total}} - k_{\mathrm{MS}}MS - S\gamma_{\mathrm{S}} \quad (2)$$

$$\frac{dK}{dt} = \mu_{\mathrm{K}}^{\mathrm{total}} - k_{\mathrm{MK}}MK \quad (3)$$

$$\frac{M_{\mathrm{K}}}{dt} = k_{\mathrm{MK}}MK - \delta_{\mathrm{MK}}M_{\mathrm{K}} \quad (4)$$

$$\frac{M_{\mathrm{S}}}{dt} = k_{\mathrm{MS}}MS - \delta_{\mathrm{MS}}M_{\mathrm{S}} \quad (5)$$

where $\mu_{\mathrm{S}}^{\mathrm{total}} \equiv \frac{\mu_{mS}\mu_S}{\delta_{mS}}$ is the total ComS protein production per unit time, M_{K} is the number of MK complexes, and M_{S} is the number of MS complexes. The number of free degradation complexes is given by M, which can be obtained from the conservation relation $M_0 = M + M_{\mathrm{K}} + M_{\mathrm{S}}$.

Since we are interested in the degradation of K in the steady state, we must find the time-independent solution to this system as a function of K while ignoring the dynamic

equation for K's evolution. The degradation rate will then be given by $k_{\mathrm{MK}}MK$. The task then becomes to find out the equilibrium value of M. Setting the last two dynamic equations to zero and using the conservation relation yields

$$M = \frac{M_0}{1 + \frac{k_{\mathrm{MK}}}{\delta_{\mathrm{MK}}}K + \frac{k_{\mathrm{MS}}}{\delta_{\mathrm{MS}}}S} \tag{6}$$

Setting the first dynamic equation to zero and solving for S yields

$$S = \frac{1}{2}\left(\frac{\mu_{\mathrm{S}}^{\mathrm{total}}}{\gamma_{\mathrm{S}}} - \frac{\delta_{\mathrm{S}}M_0}{\gamma_{\mathrm{S}}} - \frac{\delta_{\mathrm{S}}}{k_{\mathrm{MS}}}\tilde{K}\right) + \sqrt{\frac{1}{4}\left(\frac{\mu_{\mathrm{S}}^{\mathrm{total}}}{\gamma_{\mathrm{S}}} - \frac{\delta_{\mathrm{S}}M_0}{\gamma_{\mathrm{S}}} - \frac{\delta_{\mathrm{S}}}{k_{\mathrm{MS}}}\tilde{K}\right)^2 + \frac{\delta_{\mathrm{S}}\mu_{\mathrm{S}}^{\mathrm{total}}}{\gamma_{\mathrm{S}}k_{\mathrm{MS}}}\tilde{K}} \tag{7}$$

where

$$\tilde{K} = 1 + \frac{k_{\mathrm{MK}}}{\delta_{\mathrm{MK}}}K \tag{8}$$

Using these expressions, one can compute the degradation rate of K as a function of K. A graph showing both the rate of production and degradation of K as a function of K (computed with parameters at T_0) is shown in Figure S2 for a few different *comK* transcription rates. The second intersection point, representing an unstable fixed point in the dynamics, is then the threshold.

Stochastic simulations

In the absence of stochastic fluctuations, it would be impossible for a cell to transition to the competent state, since the above analysis shows that such a state is stable. However, in the presence of stochastic gene expression, fluctuations in the level of K can lead to transitions as occasional fluctuations push the number of Ks past the threshold.

We explored the stochastic properties of our model by using the Gillespie algorithm for generating exact trajectories of the stochastic system (13). The simulations incorporated a time-dependent rate of *comK* transcription based on the mean mRNA measurements obtained from the wild-type FISH experiments. The simulations were started 8 hours before

T_0 and *comS* transcription is turned on 15 minutes before T_0[2], and the total number of MecA/ClpC/ClpP complexes was set to 300. We found that, given the rates chosen, only a fraction of the randomly generated trajectories would transition to competence, whereas the rest of the trajectories would maintain low basal rates of ComK production. A few sample realizations for a particular set of parameters are given in Figure S3. It is worth pointing out that once a cell transitions to competence in our model, the production of ComK continues unchecked. This is because, as mentioned previously, we do not explicitly model the unknown shutdown of the *comK* gene once the cells become competent; such behavior is not relevant for our purposes, however, as our model is mostly concerned with the initiation of competence[3].

In Figure 4, we show the distribution of *comK* mRNA and ComK proteins in non-competent cells at T_0 as computed by our model. Cells were scored as non-competent by choosing cells with mRNA counts less than or equal to 9 mRNA per cell. The protein threshold given by the vertical red line was computed as outlined in the previous section. The Fano factor for the mRNA distributions in the wild-type strain was 1.29 and 1.12 for the low-noise strain, in agreement with the experimental data show in Figure S6.

Given that our experiments showed that the rate of competence transition seemed to depend heavily on changes in the transcriptional efficiency of *comK* (Figure 2), we then proceeded to examine how our model responded to changes in the rate of *comK* transcription by simulating a large ensemble of "cells" to determine when transitions to competence were most likely. It is important to note that both the basal and activated transcription rates of *comK* are modulated during stationary phase. This is supported by experimental evidence showing that *comK* transcription is modulated by repressors both when ComK is bound to

[2]There is data indicating that *comS* transcription begins up to 45 minutes before T_0. However, ComS only begins to accumulate to appreciable amounts at around T_0. This is likely due to the gradual accumulation of ComX, which is the upstream activator of *ComS* transcription. For modeling purposes, we ignore the effects of ComX and instead simply begin the transcription of *comS* at a time closer to T_0.

[3]It should be noted, however, that removing the requirement of a stable competent state admits a much larger range of parameters under which one may observe transitions, since the system only has to transition from the non-competent state to the competent state and not vice-versa.

the promoter and when it is not bound (14). It is also possible that the modulation of *comK* transcription is due to a global decrease in transcriptional activity, but the interpretation does not change in that case, since any global decrease (caused by, say, a decrease in the number of RNAP enzymes) would affect the promoter regardless of whether or not ComK is bound.

A histogram of transition times is shown in Figure S4A along with the cumulative number of competent cells. Clearly, there is a dramatic reduction in the number of cells that become competent after around T_2, showing that relatively small changes in the rate of transcription can completely inhibit transitions to competence. Furthermore, the fact that transitions essentially stop after the level of *comK* transcription decreases indicates that noise in the level of ComS does not play a significant role in making cells transition to competence.

The rate of transition to competence in our model is, however, sensitive to the mean level of ComS produced during stationary phase. Upon increasing or decreasing the rate of *comS* transcription during stationary phase, the number of competent cells would change markedly (Figure S4B), as observed experimentally (15). This strong parameter sensitivity is typical of the behavior of noise-induced transitions between metastable states (16).

Modulation of noise in ComK levels and its effect on competence initiation

In our model, it is the noise in gene expression of *comK* that leads to transitions to the competent state; the occasional random production of a large number of ComK molecules is what triggers activation of the positive feedback loop. Thus, it is reasonable to assume that different levels of noise in the expression of *comK* could lead to different frequencies of competence initiation. We examine this possibility by first using a simple model of stochastic gene expression to gain intuition about the behavior of noise and then proceed to apply the intuitive results to our full model of competence initiation to see whether reducing noise indeed results in a lowered rate of competence initiation.

To gain some intuition for stochastic gene expression, we noted recent work (17; 18)

showing that the mean level of gene expression and noise in gene expression (defined by the standard deviation divided by the mean), is dependent on the rates of transcription and translation as given by the following formulae:

$$\langle p \rangle = \frac{\mu_m \mu_p}{\delta_m \delta_p}, \quad \frac{\mathrm{var}(p)}{\langle p \rangle^2} = \frac{\delta_m \delta_p}{\mu_m \left(\delta_m + \delta_p\right)} + \frac{1}{\langle p \rangle} \tag{9}$$

where p is the number of proteins, μ_m and μ_p are the mRNA transcription rate and protein translation rates, and δ_m and δ_p are the mRNA and protein degradation rates, respectively. Note that the rates given here do not relate to those given earlier for the full model; here, we are presenting simpler results to gain an intuition for the function of the complete system, which we will examine shortly.

We should mention here that our expression for the "noise" is rather different than that given in Thattai and van Oudenaarden (17). In that article, the noise was defined to be the variance divided by the mean, also known as the Fano factor, which allows one to measure deviations from a Poisson distribution (in which case the Fano factor is 1). However, the Fano factor is not a particularly useful definition for measuring the size of variations relative to the mean. That quantity, given by the standard deviation divided by the mean (or, equivalently, the variance divided by the square of the mean), is more relevant and the expressions we have given reflect that choice.

One can simplify the expression for the noise by considering reasonable values for the parameters in Equation (9). If we assume a mean number of ComK proteins of around 100, that there is roughly 1 mRNA per cell and that the translation rate is roughly 50 times that of the mRNA degradation rate (i.e., 50 proteins per mRNA), then the protein degradation rate must be roughly half that of the mRNA degradation rate. Using this information, one can show that the first term in the noise expression is around 0.5, while the $1/\langle p \rangle$ term is 0.01. This indicates that in the relevant parameter regime, one can ignore the $1/\langle p \rangle$ term,

thus leading to the following expression for the noise:

$$\frac{\mathrm{var}(p)}{\langle p \rangle^2} = \frac{\delta_m \delta_p}{\mu_m \left(\delta_m + \delta_p\right)} \tag{10}$$

This expression shows that the noise in gene expression is inversely proportional to the rate of transcription while not depending significantly upon the rate of translation. Thus, one can reduce the noise while keeping the mean protein level the same by increasing the rate of transcription while reducing the rate of translation[4]. It seems reasonable to assume that this reduction in the amount of noise should then lead to fewer transitions to competence, since the number of cells exhibiting a large enough amount of ComK to initiate a transition should be reduced.

We then proceeded to check whether increasing the rate of transcription while reducing the rate of translation in the complete stochastic model would in fact result in a lowered rate of competence initiation. As before, we ran an ensemble of simulations of the stochastic model, but now with the translation rate reduced by a factor r_{comK}, while simultaneously increasing the transcription rate by the same factor. Importantly, the transcription rates in *all* the promoter states are increased by this factor; if only the basal rates of transcription (i.e., when ComK is not bound to the promoter) are changed without changing the rate of transcription when ComK is fully bound to the promoter, the threshold beyond which competence occurs is also altered. In other words, if the transcription rate of the activated *comK* gene were the same in the low-noise and wild-type strains, then the effective strength of the positive feedback would be lower in the low-noise strain due to the lowered translation efficiency. Recent experimental examinations of the mechanisms by which transcription factors act at the *comK* promoter, however, show that transcription when both ComK and Rok are bound to the promoter is significantly weaker than when ComK alone is bound (14), meaning that the strength of the feedback is likely the same in the low-noise and wild-type

[4]While it is true that one may also reduce the noise about the same mean by varying the rates of mRNA and protein degradation, these manipulations are difficult to perform experimentally and so we do not consider this case any further.

strains, justifying our model. The results of the simulations for a few values of r_{comK} are shown in Figure S5. Clearly, the rate of competence initiation strongly decreases as the noise is decreased, validating the intuitive picture outlined above. The experiments detailed in the main text confirmed this prediction, providing a powerful corroboration of the model.

Measurements of mRNA noise

In the model, we assume that the *comK* promoter transitions are all fast relative to the rate of transcription from the *comK* promoter. In the absence of feedback, this would lead to a Poisson distribution of mRNA numbers. However, we found that the variance was consistently higher than the mean in all the strains we examined, as measured by the Fano factor (variance/mean) shown in Figure S6 This is likely due to either the small extrinsic component of the noise or to some mild bursting.

In particular, it is interesting to note that the wild-type strain Fano factor was consistently somewhat higher than that of the *comK* strain. This is likely due to some small amount of positive feedback of ComK at its own promoter, causing some mild bursting. Similar statistics were observed in the *rok* strain, consistent with the model. The increase in the Fano factor at later times in stationary phase is likely due to the occasional inclusion of a competent cell in the analysis at those time points.

Propagation of extrinsic transcriptional fluctuations to protein noise

In the main text, we made the assumption that if two genes driven by the *comK* promoter produced relatively uncorrelated numbers of mRNA, then the noise in ComK protein expression would also be primarily intrinsic. However, since the number of mRNA molecules we have observed is very low, the intrinsic contribution to the variations is necessarily large. Un-

der certain conditions, one could imagine a case where small extrinsic variations in mRNA numbers could become significant at the protein level. This could happen if the protein degradation rate is very slow; in that case, translation effectively acts as an averaging filter, smoothing out the random fluctuations in mRNA levels[5]. If the extrinsic fluctuations in mRNA are sufficiently slow, the averaging effects of translation would result in both genes following the extrinsic variations in transcription while smoothing over the uncorrelated random fluctuations in mRNA production and degradation.

This argument, however, rests on the protein degradation rate being sufficiently low. To show that the lack of correlation we observe in the mRNA counts of the two genes should result in a lack of correlation between the two proteins in our case, we performed stochastic simulations in which the transcription rate varied in a coordinated fashion, thus simulating extrinsic transcriptional noise. In the stochastic simulations, we had two copies of a gene, each with the same mRNA degradation rate (0.008 molecules^{-1} sec^{-1}), translation rate (1 sec^{-1}) and protein degradation rate (0.2 molecules^{-1} sec^{-1}). The burst parameter (i.e., number of proteins per mRNA, or the translation rate divided by the mRNA degradation rate) was chosen to be large, as might be expected for a gene with a strong Shine-Dalgarno sequence, such as *comK*, and the degradation rate was chosen to yield 5 protein molecules per cell, on average, given a mRNA mean of 1 molecule per cell, which corresponds with the results of the stochastic simulations outlined previously. The transcription rate was chosen to vary periodically on the order of a cell cycle between 0.003 and 0.013 molecules sec^{-1} as shown in Figure S7A; the resulting mean number of mRNAs per cell is roughly 1, in line with our experimental results. The results of the simulations gave an mRNA correlation coefficient of around 0.176, with a protein correlation of around 0.162 (protein correlation shown in Figure S7B). This shows that, for parameters chosen to be reasonable to our case, the extrinsic variations in transcription rates do not propagate to correlations in protein

[5]Another way to see this is that, as shown in the previous section, reducing the protein degradation rate can reduce the intrinsic protein noise. If the intrinsic protein noise is decreased, then extrinsic effects will become more apparent.

number. Essentially, this is because we were able to set a reasonable lower bound on the rate of protein degradation. Since we know the mean number of mRNA at T_0 is roughly 1, the mean number of proteins is simply μ_p/δ_p. Given that we know that the number of proteins is very low (since fluorescent protein levels in non-competent cells are essentially undetectable) and that the Shine-Dalgarno sequence of the gene is close to optimal, we know that μ_p/δ_p must be small but μ_p must be large. This means that δ_p must also be large, thus precluding the possibility of translation acting as a filter. The active degradation of ComK proteins by the MecA/ClpC/ClpP complex is the most likely reason that the protein degradation rate is relatively large.

References

1. M. Albano, J. Hahn, D. Dubnau, *J Bacteriol* **169**, 3110 (1987).

2. D. Y. Vargas, A. Raj, S. A. Marras, F. R. Kramer, S. Tyagi, *Proc Natl Acad Sci U S A* **102**, 17008 (2005).

3. A. Raj, C. Peskin, D. Tranchina, D. Vargas, S. Tyagi, *PLoS Biol* **4**, e309 (2006).

4. L. Kong, K. Siranosian, A. Grossman, D. Dubnau, *Mol Microbiol* **9**, 365 (1993).

5. A. Guérout-Fleury, K. Shazand, N. Frandsen, P. Stragier, *Gene* **167**, 335 (1995).

6. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, *Science* **280**, 585 (1998).

7. J. Spizizen, *Proc Natl Acad Sci U S A* **44**, 1072 (1958).

8. R. Magnuson, J. Solomon, A. Grossman, *Cell* **77**, 207 (1994).

9. L. Hamoen, A. Van Werkhoven, J. Bijlsma, D. Dubnau, G. Venema, *Genes Dev* **12**, 1539 (1998).

10. J. Hasty, J. Pradines, M. Dolnik, J. J. Collins, *Proc Natl Acad Sci U S A* **97**, 2075 (2000).

11. K. Turgay, J. Hahn, J. Burghoorn, D. Dubnau, *EMBO J* **17**, 6730 (1998).

12. G. Süel, J. Garcia-Ojalvo, L. Liberman, M. Elowitz, *Nature* **440**, 545 (2006).

13. D. T. Gillespie, *J Phys Chem* **81**, 2340 (1977).

14. W. K. Smits, T. T. Hoa, L. Hamoen, O. P. Kuipers, D. Dubnau, *Molecular Microbiology* **64**, 368 (2007).

15. H. Maamar, D. Dubnau, *Mol Microbiol* **56**, 615 (2005).

16. C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, New York, 2004), third edn.

17. M. Thattai, A. van Oudenaarden, *Proc Natl Acad Sci U S A* **98**, 8614 (2001).

18. E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, A. van Oudenaarden, *Nat Genet* **31**, 69 (2002).

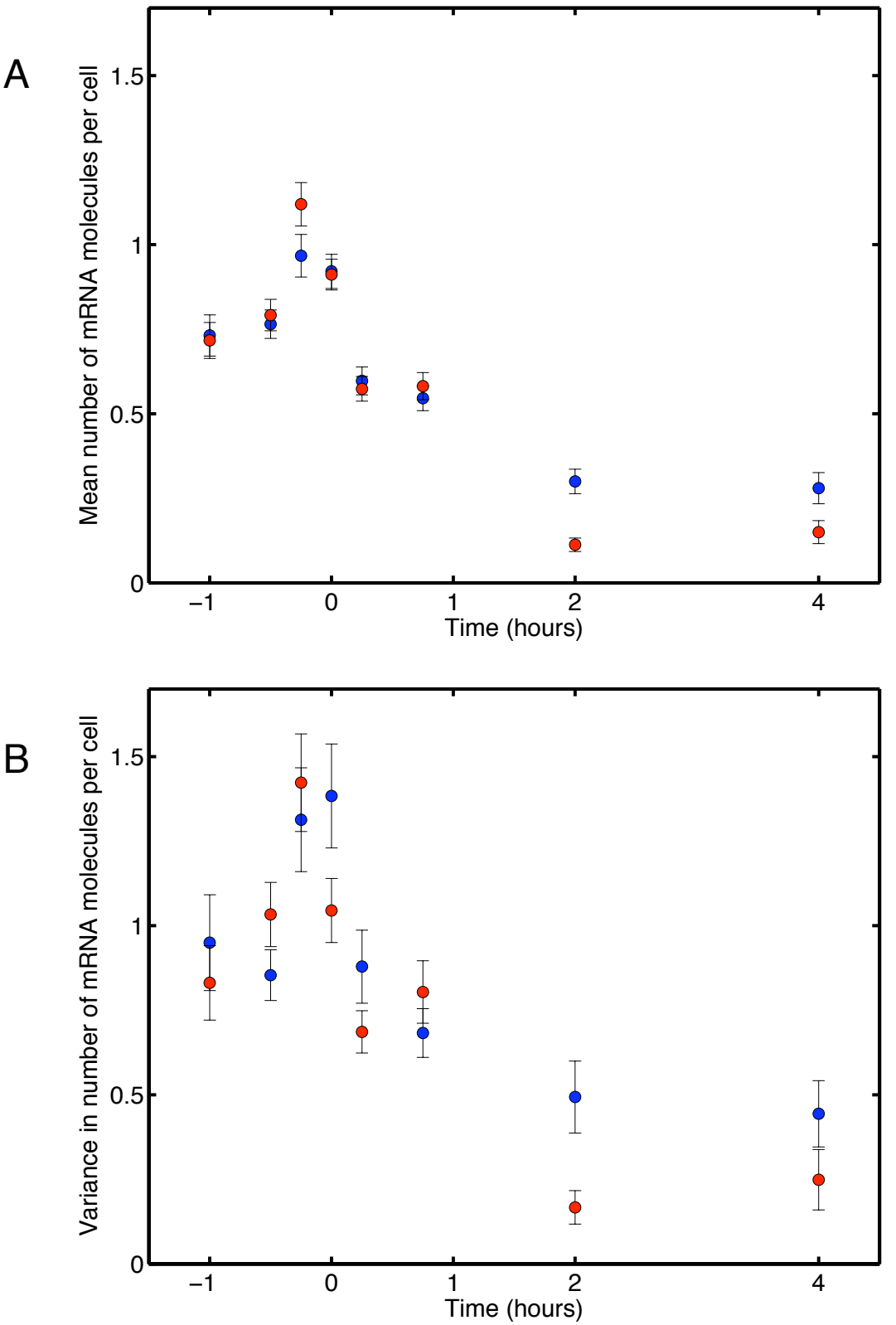

Figure S1: Comparison of the statistical properties of the number of *comK* (blue) and *comK-M2* (red) mRNA per non-competent cell in the wild-type strain. A) Mean number of mRNA per cell; **B**) Variance in the number of mRNA per cell. Error bars were obtained by bootstrap resampling.

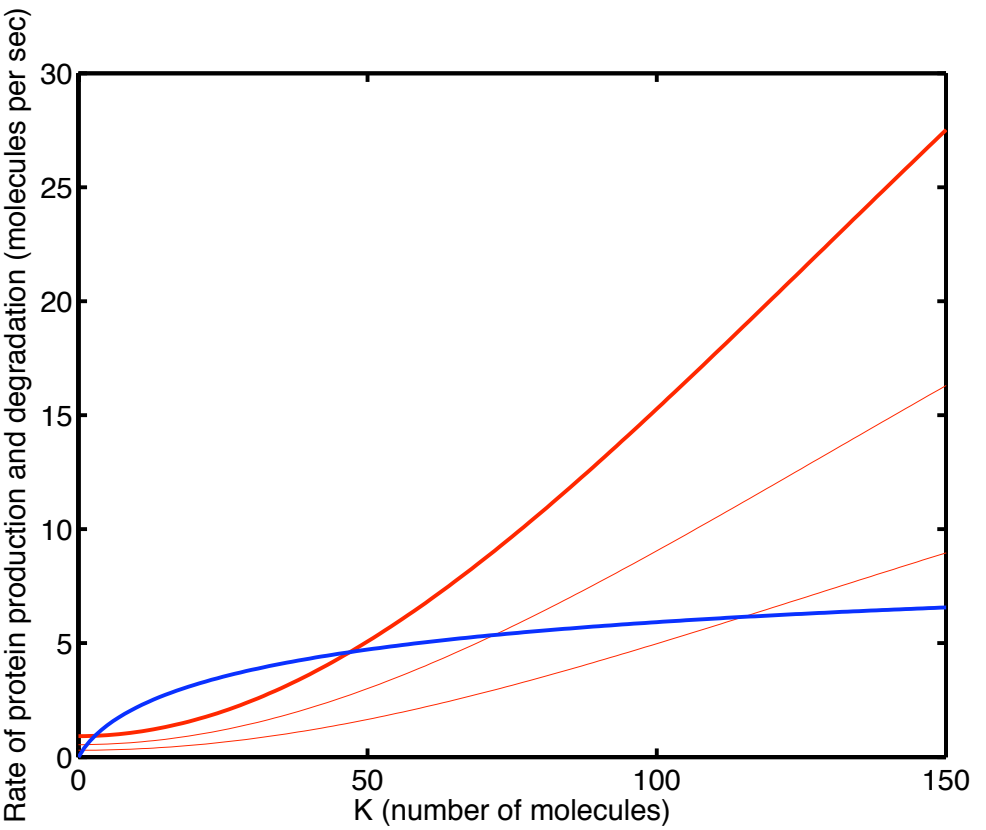

Figure S2: ComK production (red) and degradation (blue) as a function of the number of ComK dimers. The intersections of the graphs are fixed points of the deterministic dynamical system, with the lower fixed point being stable and the upper fixed point, corresponding to the threshold, being unstable. The thick red line corresponds to the production rate at T_0 with mRNA mean of .92, whereas the thin red lines correspond $T_{0.75}$ and T_2, where the decreased transcription rate resulted in mRNA means of 0.55 and 0.30, respectively.

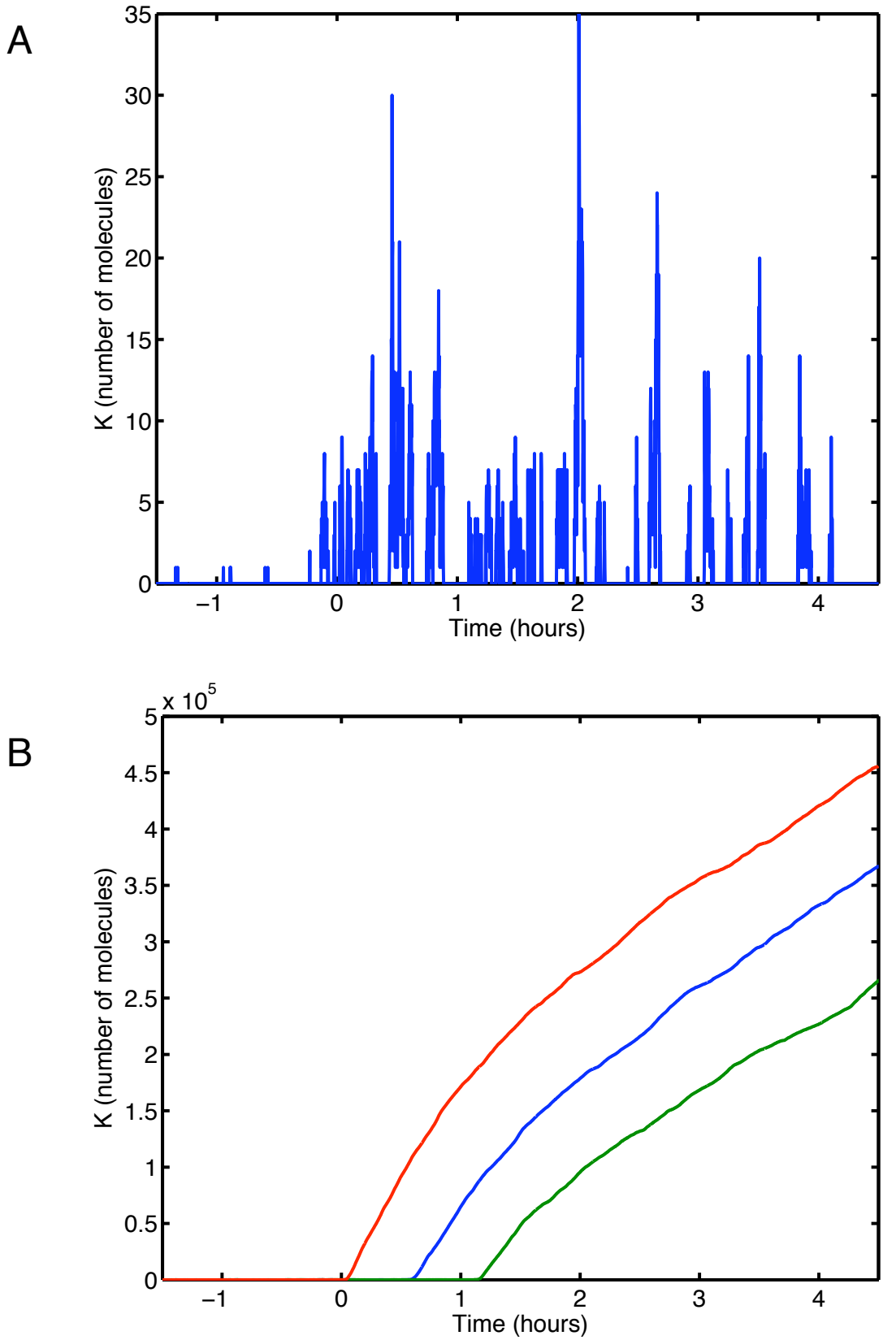

Figure S3: Realizations of stochastic simulations admitting competence. **A**) Time course of the number of ComK molecules in a simulation run not resulting in competence. **B**) Time courses of the number of ComK molecules in three simulation runs resulting in competence at different times during stationary phase.

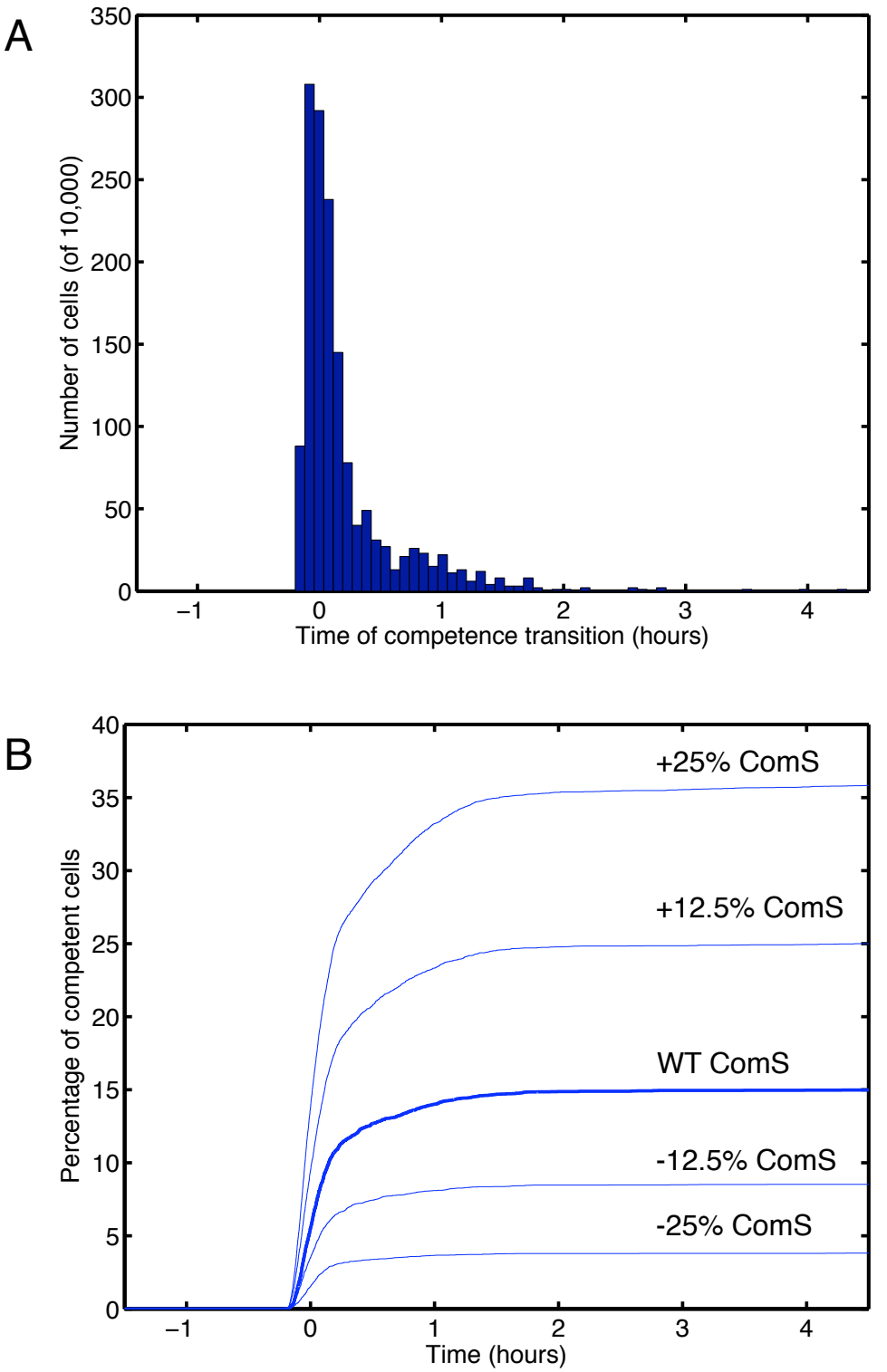

Figure S4: Probability of transition as a function of time. **A**) The distribution of transition times for wild-type parameters, computed from 10,000 total simulation runs. **B**) Percentage of competent cells as a function of time. The thick blue line corresponds to the wild-type rate of ComS transcription, while the thin blue lines represent the percentage of competent cells with the indicated increase or decrease in rate of ComS transcription.

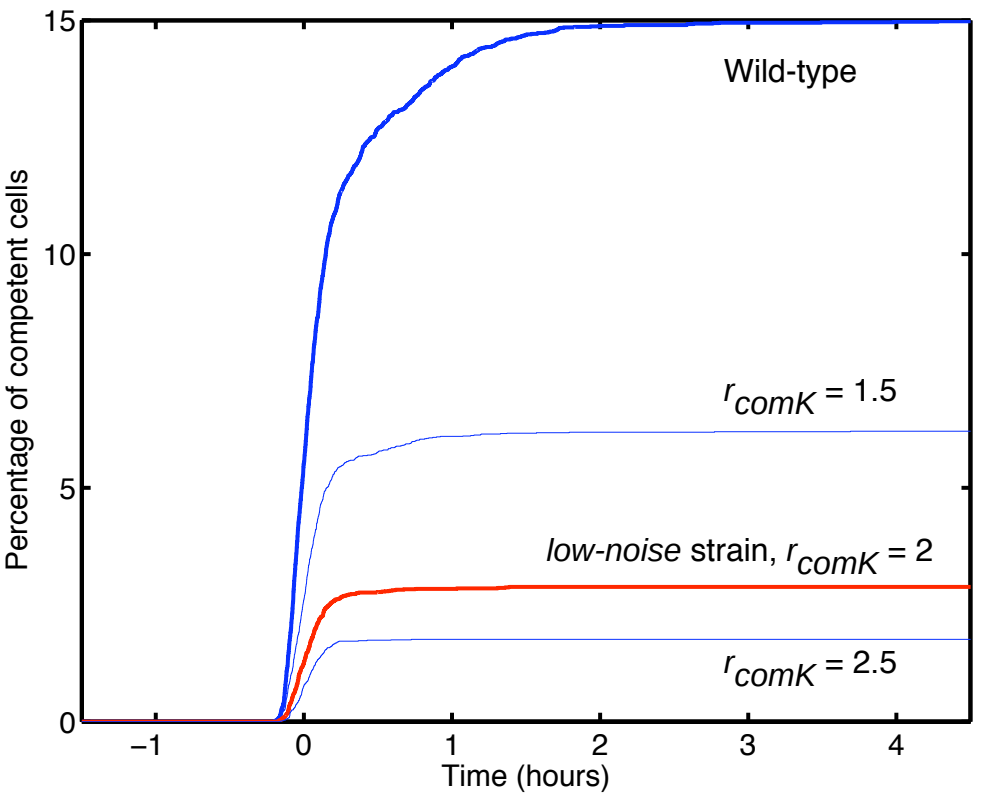

Figure S5: Percentage of competent cells as a function of time for different values of r_{comK}. For the wild-type strain r_{comK} = 1, with the percentage indicated by the thick blue line. The percentage of competent cells for the value of r_{comK} corresponding to that expected in the *low-noise* strain is given by the thick red line.

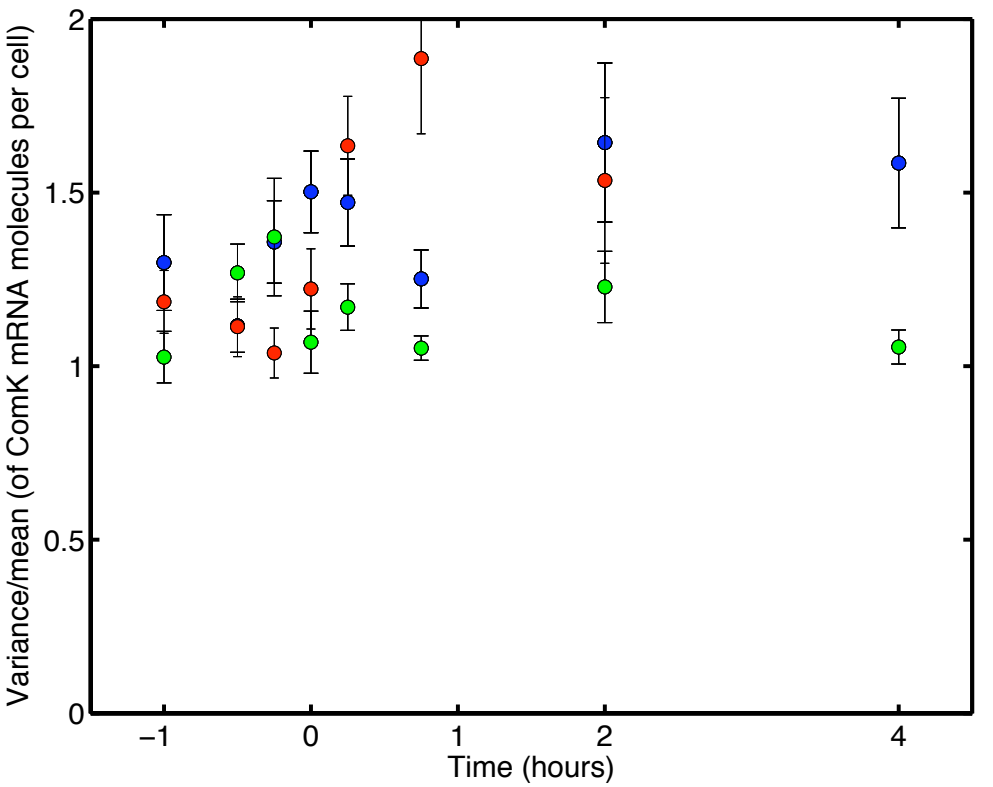

Figure S6: Fano factor for mRNA distributions for the wild-type strain (blue), the *comK* strain (green) and the *rok* strain (red). The Fano factor is defined as the variance divided by the mean. Error bars were obtained by bootstrap resampling.

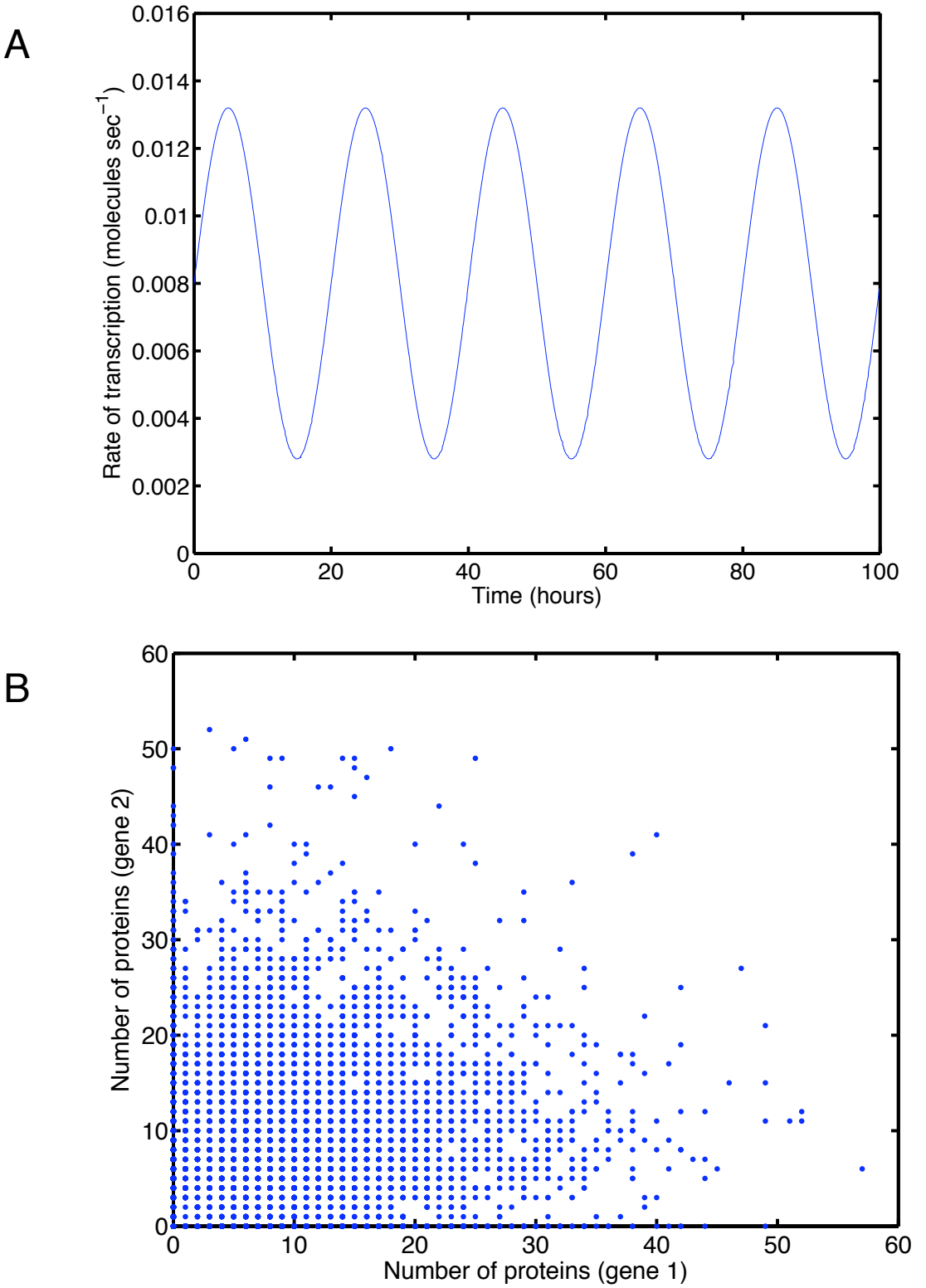

Figure S7: Propagation of extrinsic noise in transcription to protein levels. **A**) The time-dependent transcription rate of genes 1 and 2, conferring some measure of extrinsic noise to the system. **B**) Scatter plot of the number of protein molecules of genes 1 and 2. The correlation coefficient is 0.162.

Movie S1. Time lapse experiment showing growth and competence development of the wild type strain carrying a comK-yfp fusion. Cells were deposited on an agarose pad containing competence medium and pictures were taken every 10 to 15 minutes after the first hour of growth.