

Supplementary Information

Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock

Jeffrey R. Chabot, Juan M. Pedraza, Prashant Luitel & Alexander van Oudenaarden

EXPERIMENTAL MATERIALS AND METHODS

Development of fluorescence reporter assay

We screened a panel of several commercially available fluorescent proteins with emission wavelengths covering the visible spectrum from cyan (CFP) to deep red (HcRed). First, we introduced the fluorescent reporter under the control of the strong synthetic *trc* promoter (P_{trc}) as a single copy in the chromosome. This promoter can be externally induced by the lactose analog IPTG (Fig. S1b, inset). We found that only a yellow-shifted variant of green fluorescent protein (YFP) was able to faithfully report the induction of P_{trc} with minimal background. In addition, chlorophyll and other bacteriophytochromes in *Synechococcus* do not absorb significantly at the YFP excitation wavelength, reducing the possibility of brief excitation pulses affecting the timing of the circadian clock. We did not detect any increased fluorescence above background upon induction for GFPmut3.1 and GFPuv. The red fluorescent reporters conflicted with the substantial red autofluorescence of chlorophylls. The cyan fluorescent protein (CFP) reported a modest increase in the presence of IPTG, but with a significant background owing to the large blue auto-fluorescence in cyanobacterial cells. We therefore selected YFP for further exploration.

For the fluorescent protein to be a sensitive reporter of expression fluctuations, the half-life of the reporter should preferably be much smaller than the 24 hour period of the circadian clock. We determined the *in vivo* half-life of YFP to be 12.8 hours, which is close to the cell doubling time under our test conditions, indicating that cell growth and division dominates the rate at which YFP is effectively removed from the cells (Fig. S1a). The YFP stability confers the advantage of higher signal levels; however, the sluggish response smoothes out the oscillating signal. To achieve a compromise between these factors, an *ssrA* degradation tag was appended to the *yfp* gene. This tagged version was called YFPLVA. The effective half-life of YFPLVA was reduced to 5.6 hours indicating that proteolysis in addition to cell growth determines the half-life of the tagged reporter. Because of its faster response YFPLVA was selected as the single cell reporter. Figure 1b shows the induction kinetics of YFPLVA demonstrating that the reporter is immediately

visible (within the time resolution of our experiment), implying a negligible lag time between induction and synthesis of functional reporter protein.

In the following step, strains were constructed expressing *yfpLVA* under a native *Synechococcus* promoter. After synchronisation to two 12 hour light/12 hour dark (LD) cycles, the expression of *yfpLVA* was tested under constant light (LL) conditions. Figure S1c shows YFPLVA expression driven by the *kaiBC* promoter (which bicistronically controls expression of two clock genes *kaiB* and *kaiC*) of a synchronised population. Additionally this assay can be used to monitor YFP expression in individual cells as they grow into micro-colonies using fluorescence time-lapse microscopy (Fig. 1, Supplementary Movie 1).

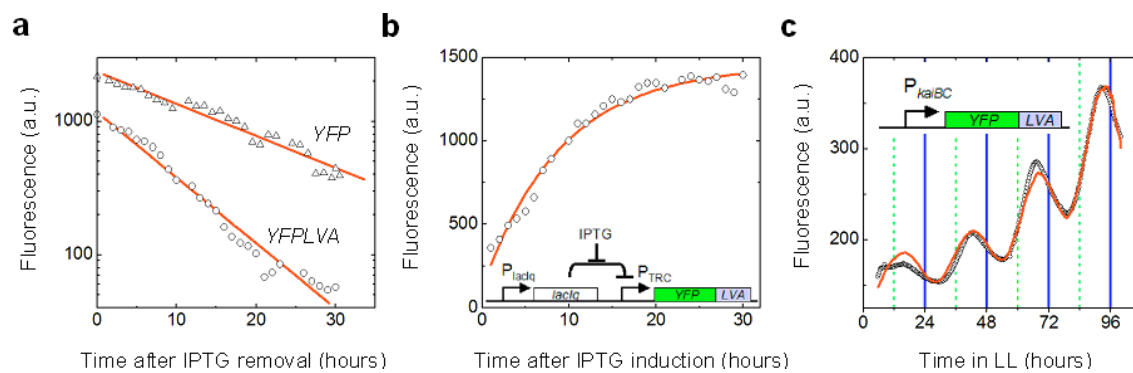


Figure S1. Experimental characterisation of the single-cell fluorescence reporter assay. **a**, Degradation dynamics of YFP and YFPLVA. Strain JRCS05 (P_{Trc} -*yfp*) and JRCS13 (P_{Trc} -*yfpLVA*) were fully induced with 1 mM IPTG for 36 hours to reach maximum expression levels. Subsequently, these strains were washed and resuspended in growth media without IPTG. Fluorescence was obtained by averaging the fluorescence of at least 10^4 single cells. Red lines denote linear fits on semi-log axes. **b**, Induction dynamics of YFPLVA. Strain JRCS13 (P_{Trc} -*yfpLVA*) was induced at $t = 0$ by adding 1 mM IPTG. Fluorescence was obtained by averaging the fluorescence of at least 10^4 single cells. Red line denotes a fit to an exponential approach to steady-state (inset) Schematic illustration of synthetic inducible system. **c**, Total fluorescence of a population of cells (strain JRCS32; P_{kaiBC} -*yfpLVA*) as a function of circadian time. Cells were synchronized by two 12:12 LD cycles prior to the start of the experiment. The fluorescence increases over time due to the increased number of cells in the population due to cell division. The

red line denotes a fit to the function $2^{t/\tau_{1/2}} \left[\cos\left(\frac{2\pi t}{T} + \phi\right) + C_1 \right] + C_2$, where $\tau_{1/2}$ and T are the

cyanobacterial doubling time and oscillation period respectively. The constants C_1 and C_2 are needed to account for the fluorescent background. This function models an exponential dividing culture of synchronised cells. The angle ϕ denotes the phase of the circadian rhythm.

Plasmid and strain constructions

yfp was PCR-amplified from pEYFP (Clontech), adding a BamHI site and the strong Shine-Delgarno sequence TAAGGAGGAAAAAAA before the start codon. To construct *yfpLVA*, the stop codon of *yfp* was replaced with the sequence

GCATGGACGAGCTGTACAAGGCAGCAAACGACGAAAACACTACGCTTTAGTAGC TTAAT encoding the *LVA* degradation tag. For both *yfp* and *yfpLVA*, the restriction sites *S*all and *S*acI were appended after the stop codon. These products were ligated into the *Synechococcus* neutral site I (NSI) plasmid pAM1303.

P_{trc} was PCR-amplified from pTrc99A (AP Biotech) using primers ATAAGAATGCGGCCGCTCTAGACTGAAATGAGCTGTTGACAATT (5'-primer, adding *Not*I and *X*baI) and CGGGATCCGCATGCGTGTGAAATTGTTATCCGCTC (3'-primer, adding *B*amHI and *S*phI). This region includes the active promoter elements of P_{trc} . This product was ligated into the *yfp* and *yfpLVA* containing NSI plasmids resulting in plasmids pJRC05 and pJRC13 respectively. These plasmids were transformed into *Synechococcus elongatus* PCC7942 resulting in *S. elongatus* strains JRCS05 and JRCS13, respectively.

To repress P_{trc} , the *lac* repressor gene driven by the strong constitutive P_{lacIq} promoter was inserted in the *S. elongatus* neutral site II (NSII) plasmid pAM1579.

The *kaiBC* promoter was cloned from the *S. elongatus* PCC7942 chromosome using primers ATAAGAATGCGGCCGCTCTAGACGCAAATTGCAATCTGCATTG (5'-primer; adding *Not*I and *X*baI) and GAAGATCTACGCAGATCAACGGGGTAG (3'-primer, adding *B*glII). This region includes 937 bases upstream of the *kaiB* start codon. This sequence was ligated into the NSI *yfpLVA* plasmid between the *Not*I and *B*amHI sites resulting in plasmid pJRC27. This plasmid was transformed into PCC7942 using standard *S. elongatus* protocols, resulting in *S. elongatus* strains JRCS32.

Fluorescence microscopy

A Nikon Eclipse 800 upright microscope was used to collect bulk fluorescence data on *S. elongatus* reporter strains. 300 μ L of mid exponential cells were loaded into a 96 well plate (Corning). Every 20 minutes, the plate was scanned using a motorized stage (Prior Scientific) controlled by Metamorph software (Universal Imaging), and 5 second fluorescence exposures were collected by a CoolSnap HQ camera (Photometrics) using a 10x objective. The fluorescence source was a xenon arc lamp (Sutter Instruments); this source provides a temporally stable illumination. The cells were kept in a LL environment (except during the fluorescence scans) by a circular cool fluorescent bulb (Sylvania) surrounding the objective; this light was controlled by Metamorph. Custom-made scripts were developed in MATLAB to automatically segment cells (using watershed algorithms) and track the fluorescence of single cells over time.

Automatic entrainment

To collect data on differently phased samples, we constructed a device which carries 24 tubes evenly placed around a circle through a light and a dark (< 50 lux) environment. The spinner rotates once per day, giving each tube a 12:12 LD exposure, with each tube phased one hour differently than the tubes beside it. After at least 2 complete 12:12 LD cycles, the tubes were removed (while in the light environment) to a matching LL incubator, and samples from each tube were measured. Complete sets of these measurements were performed less than 24 hours apart, so individual tubes could be seen oscillating.

Flow cytometry

Flow cytometry was performed on a FACScan cytometer (BD Biosciences), using a 488 nm excitation laser and a 530/30 emission filter (which can effectively measure GFP or YFP). Cells were selected from a small gate near the center of the population in forward scatter/side scatter space; this gate typically encompassed ~10% of all scattering events. The gate was fixed for all measurements of a particular experiment. The resulting data files were converted into ASCII format using MFI (E. Martz, University of Massachusetts, Amherst). To confirm that the cell size does not significantly change during the circadian cycle, we measured the cell length during a circadian period using phase contrast microscopy (Fig. S2). Although we observe a slight increase in cell length around CT = 16 hrs, this relative increase is very small compared to the observed changes in stochastic gene expression.

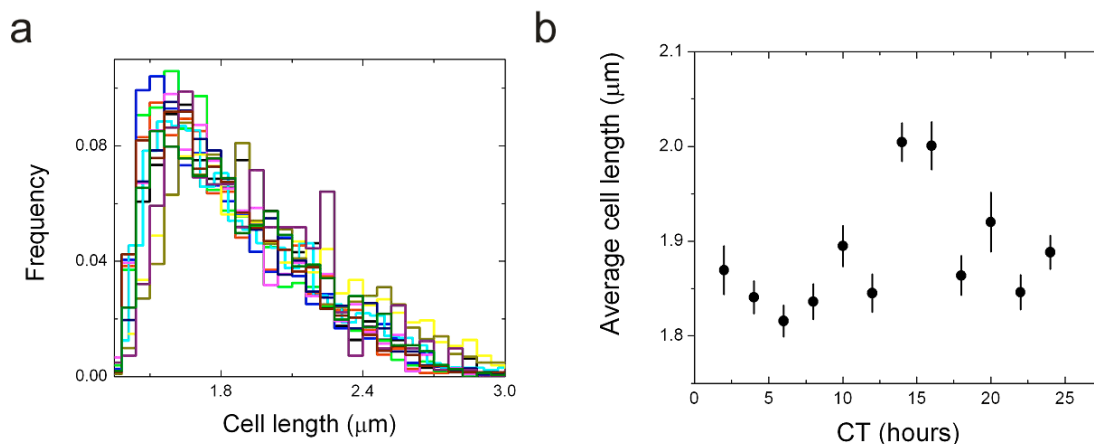


Figure S2. **a**, Length distributions at the different circadian time as measured with phase contrast microscopy. Each color reflects a length distribution determined at a different circadian time. **b**, Average cell length as a function of circadian time. Approximately 1000 cells were evaluated for each distribution.

MODELING MATERIALS AND METHODS

Calculation of delay in fluorescence reporter protein

Consider a driving signal with frequency ω and arbitrary phase $\phi = 0$. A reporter being produced with a production rate k modulated by such a signal, with an constant production offset of A and a first order decay with constant γ , is described by

$$\dot{y} = k[\sin(\omega t) + A] - \gamma y \quad [1]$$

The solution to [1] takes the form

$$y = y_0 \sin(\omega t + \phi) + c \quad [2]$$

Inserting this solution in [1] yields

$$y_0 \omega \cos(\omega t + \phi) = k \sin(\omega t) - \gamma y_0 \sin(\omega t + \phi) \quad [3]$$

Trigonometric substitution gives

$$y_0 \omega [\cos(\omega t) \cos(\phi) - \sin(\omega t) \sin(\phi)] = k \sin(\omega t) - \gamma y_0 [\sin(\omega t) \cos \phi + \cos(\omega t) \sin(\phi)] \quad [4]$$

Dividing by $\cos(\omega t)$ and grouping terms independent of t reveals

$$\tan^{-1}(\phi) = \frac{\omega}{\gamma} \quad [5]$$

This term represents the phase difference observed between the reporter and the driving signal. Experimentally we found a half-life of the YFPLVA reporter of 5.6 hours (Fig.

1a). This results in a degradation rate $\gamma = \frac{\ln 2}{t_{1/2}} \approx 0.12 \text{ hr}^{-1}$. For a frequency $\omega = \frac{2\pi}{24} \approx 0.26$

hr^{-1} , equation [5] predicts a phase lag ϕ of about 65° . This implies that the reporter protein lags about 4.3 hours, consistent with the experimentally observed lag (Fig. 2).

Langevin modeling for an out-of-equilibrium system

We model the system as a set of two variables, x and y , where x represents the amount of mRNA and y is the amount of the protein. The variables follow the deterministic,

macroscopic differential equations plus Langevin terms corresponding to different sources of noise. We assume that the time dependence is introduced only in the transcription rate, as do external sources of noise. For simplicity, we also ignore the intrinsic noise coming from the individual protein creation and destruction events, as this is relatively small for the estimated numbers of proteins. This results in the following set of equations:

$$\begin{aligned}\frac{dx}{dt} &= k_R(t) - \gamma_R x + q_{local}(t)\xi_{local}(t) + q_{global}(t)\xi_{global}(t) \\ \frac{dy}{dt} &= k_P x - \gamma_P y\end{aligned}\quad [6]$$

where γ_R and γ_P are the rate constants for RNA degradation and protein degradation, respectively. The function $k_R(t)$ describes the periodic transcription rate and k_P is the translation rate. The noise terms $\xi_i(t)$ reflect Gaussian noise with zero mean, $\langle \xi_i(t) \rangle = 0$, and have exponentially decaying correlation, $\langle \xi_i(t)\xi_i(\tau) \rangle = \frac{\gamma_N}{2} e^{-\gamma_N|t-\tau|}$ with a decay rate γ_N , which is assumed equal to the rate constant of protein destruction γ_P . The index i denotes global or local. The sources of the local and global noise are assumed to be uncorrelated with each other and by definition the global noise is fully correlated from gene to gene while the local noise is uncorrelated between two genes. We use a mean-field approximation to assume that the noise magnitudes $q_i(t)$ depend on the average $\langle x(t) \rangle$ but not on the random trajectories $x(t)$. This limits the validity of the approach in the case of very few molecules.

In Fourier space, the dynamic equation for the fluctuation $\delta x = x - \langle x \rangle$ becomes

$$\begin{aligned}\tilde{\delta x}(\omega) &= \frac{1}{i\omega + \gamma_R} (\overbrace{q_{local}\xi_{local}(\omega)} + \overbrace{q_{global}\xi_{global}(\omega)}) = \\ &= \tilde{f}(\omega) (\overbrace{q_{local}\xi_{local}(\omega)} + \overbrace{q_{global}\xi_{global}(\omega)})\end{aligned}\quad [7]$$

where $f(t) = e^{-\gamma_R t} Z(t)$ with $Z(t)$ the step function ($Z(t < 0) = 0; Z(t \geq 0) = 1$). The notation

$\tilde{f}(\omega)$ denotes the Fourier transform of function $f(t)$.

Using the property that the product of two functions in the time-domain equals the convolution (denoted by \otimes) of the corresponding functions in the frequency-domain and vice-versa, we obtain

$$\begin{aligned}
 \overline{\sigma_x^2}(\omega) &= \overline{\langle \delta x \delta x \rangle}(\omega) = \langle \overline{\delta x}(\omega) \otimes \overline{\delta x}(\omega) \rangle = \\
 &= \langle \overline{f}(\omega) (\overbrace{q_{local} \xi_{local}}(\omega) + \overbrace{q_{global} \xi_{global}}(\omega)) \otimes \overline{f}(\omega) (\overbrace{q_{local} \xi_{local}}(\omega) + \overbrace{q_{global} \xi_{global}}(\omega)) \rangle \\
 &= \langle \overline{f}(\omega) \overbrace{q_{local} \xi_{local}}(\omega) \otimes \overline{f}(\omega) \overbrace{q_{local} \xi_{local}}(\omega) \rangle + \langle \overline{f}(\omega) \overbrace{q_{global} \xi_{global}}(\omega) \otimes \overline{f}(\omega) \overbrace{q_{global} \xi_{global}}(\omega) \rangle \\
 &= \langle \overbrace{f \otimes q_{local} \xi_{local}}(\omega) \otimes \overbrace{f \otimes q_{local} \xi_{local}}(\omega) \rangle + \langle \overbrace{f \otimes q_{global} \xi_{global}}(\omega) \otimes \overbrace{f \otimes q_{global} \xi_{global}}(\omega) \rangle \\
 &= \langle \overbrace{(f \otimes q_{local} \xi_{local})^2}(\omega) \rangle + \langle \overbrace{(f \otimes q_{global} \xi_{global})^2}(\omega) \rangle
 \end{aligned} \tag{8}$$

Upon inverse-transforming, this results in

$$\begin{aligned}
 \sigma_x^2(t) &= \langle (f \otimes q_{local} \xi_{local})^2 \rangle + \langle (f \otimes q_{global} \xi_{global})^2 \rangle \\
 &= \iint q_{local}(\tau) q_{local}(\tau') \langle \xi_{local}(\tau) \xi_{local}(\tau') \rangle f(\tau) f(\tau') d\tau d\tau' \\
 &\quad + \iint q_{global}(\tau) q_{global}(\tau') \langle \xi_{global}(\tau) \xi_{global}(\tau') \rangle f(\tau) f(\tau') d\tau d\tau' \\
 &= \int_{-\infty}^t \int_{-\infty}^t (q_{local}(\tau) q_{local}(\tau') + q_{global}(\tau) q_{global}(\tau')) \frac{\gamma_N}{2} e^{-\gamma_N |\tau - \tau'|} e^{-\gamma_R(t-\tau) - \gamma_R(t-\tau')} d\tau d\tau'
 \end{aligned} \tag{9}$$

This result makes intuitive sense, as it means that the fluctuations at different times are added up with their decaying correlation and their effect decreases exponentially with the time constant of the decay.

An analogous procedure is used to calculate the dynamics of the variance in the protein concentration $\sigma_y^2(t)$:

$$\begin{aligned}
 \overline{\delta y}(\omega) &= \frac{k_P \overline{\delta x}(\omega)}{i\omega + \gamma_P} = \frac{k_P}{(i\omega + \gamma_P)(i\omega + \gamma_R)} (\overbrace{q_{local} \xi_{local}}(\omega) + \overbrace{q_{global} \xi_{global}}(\omega)) \\
 &= \overline{f_{RP}}(\omega) (\overbrace{q_{local} \xi_{local}}(\omega) + \overbrace{q_{global} \xi_{global}}(\omega))
 \end{aligned} \tag{10}$$

with $f_{RP}(t) = \left(\frac{e^{-\gamma_R t}}{\gamma_P - \gamma_R} + \frac{e^{-\gamma_P t}}{\gamma_R - \gamma_P} \right) Z(t)$. Using the same procedure as outlined by equations [6]-[9]:

$$\sigma_y^2(t) = k_P^2 \left(\left\langle \left(f_{RP} \otimes q_{local} \xi_{local} \right)^2 \right\rangle + \left\langle \left(f_{RP} \otimes q_{global} \xi_{global} \right)^2 \right\rangle \right) \quad [11]$$

For the case where two copies of the gene are inserted in neutral sites N_I and N_{II} :

$$\begin{aligned} \frac{dx_1}{dt} &= k_R(t) - \gamma_R x_1 + q_{1local}(t) \xi_{1local}(t) + q_{global}(t) \xi_{global}(t) \\ \frac{dx_2}{dt} &= k_R(t) - \gamma_R x_2 + q_{2local}(t) \xi_{2local}(t) + q_{global}(t) \xi_{global}(t) \\ \frac{dy_{1+2}}{dt} &= k_P(x_1 + x_2) - \gamma_P y \end{aligned} \quad [12]$$

This implies

$$\overbrace{\delta y_{1+2}}(\omega) = k_P \overbrace{f_{RP}}(\omega) \left(\overbrace{q_{1local} \xi_{1local}}(\omega) + \overbrace{q_{2local} \xi_{2local}}(\omega) + 2 \overbrace{q_{global} \xi_{global}}(\omega) \right) \quad [13]$$

Which, by the same procedure as above, results in

$$\sigma_{y_{1+2}}^2(t) = k_P^2 \left(\left\langle \left(f_{RP} \otimes q_{1local} \xi_{1local} \right)^2 \right\rangle + \left\langle \left(f_{RP} \otimes q_{2local} \xi_{2local} \right)^2 \right\rangle + 4 \left\langle \left(f_{RP} \otimes q_{global} \xi_{global} \right)^2 \right\rangle \right) \quad [14]$$

Using this expression it is possible to separate each term from the experimental measurements of $\sigma_{y_{1+2}}^2(t)$, $\sigma_{y_1}^2(t)$ and $\sigma_{y_2}^2(t)$ (Fig. 4). It is straightforward then to numerically solve for the different $q_i(t)$.

Langevin equivalent for a time-independent multiplicative noise source

In the case where noise arises from a time-independent multiplicative factor in the creation rate of a chemical species, the equivalent Langevin term can be determined as follows: if the creation rate $k_x(t)$ can be decomposed into a time-dependent factor $h(t)$ and a fluctuating factor y which is time independent, the equation

$$\dot{x} = k(t) - \gamma x = h(t)y - \gamma x \Rightarrow \delta\dot{x} = h(t)\delta y - \gamma\delta x \quad [15]$$

can be rewritten as

$$\delta\dot{x} = -\gamma x + q_{eq}(t)\xi_{eq} \Rightarrow \dot{x} = \langle k(t) \rangle - \gamma x + q_{eq}(t)\xi_{eq} \quad [16]$$

The equivalent stochastic term ξ_{eq} must contain the variation δy , but for proper normalization it should be divided by the static variance σ_y . For exponentially decaying autocorrelation of y , this results in a correlation

$$\left\langle \frac{\delta y(t)}{\sigma_y} \frac{\delta y(t+\tau)}{\sigma_y} \right\rangle = \frac{A_y(\tau)}{A_y(0)} = e^{-\gamma_y|\tau|} \quad [17]$$

To have units that match the definition in the intrinsic noise case, we can impose the additional normalization condition $\int \langle \xi_{eq}(t)\xi_{eq}(t+\tau) \rangle d\tau = 1$, which requires

$\xi_{eq}(t) = \sqrt{\frac{\gamma_y}{2}} \frac{\delta y(t)}{\sigma_y}$. This in turn implies that

$$q_{eq}(t) = h(t)\sigma_y \sqrt{\frac{2}{\gamma_y}} = k(t)\eta_y \sqrt{\frac{2}{\gamma_y}} \quad [18]$$