

Stochastic simulation of prokaryotic two-component signalling indicates stochasticity-induced active-state locking and growth-rate dependent bistability - Supplementary Information

Katy Wei, Maxim Moinat, Timo R. Maarleveld, Frank J. Bruggeman

Contents

1	Signalling network reaction rate constants	1
2	Two-component signalling robustness - mathematical analysis	3
3	Converting Deterministic Bistability Model Parameters to Stochastic Model input	4
4	StochPy Cell Division extension	6
4.1	Settings for reproduction of results	7

1 Signalling network reaction rate constants

Parameter	Estimate	Reaction
k1f	$0.01 \text{ nM}^{-1} \text{ s}^{-1}$	$S+L \rightarrow SL$
k1b	0.059 s^{-1}	$SL \rightarrow S+L$
k2	0.1 s^{-1}	$SL (+ATP) \rightarrow SLP (+ADP)$
k3f	0.001 s^{-1}	$SLP \rightarrow SP$
k3b	0.001 s^{-1}	$SP \rightarrow SLP$
k4f	$0.0238 \text{ nM}^{-1} \text{ s}^{-1}$	$SLP+R \rightarrow SLP\cdot R$
k4b	0.001 s^{-1}	$SLP\cdot R \rightarrow SLP+R$
k5f	$0.0238 \text{ nM}^{-1} \text{ s}^{-1}$	$SP+R \rightarrow SP\cdot R$
k5b	0.001 s^{-1}	$SP\cdot R \rightarrow SP+R$
k6f	19.8 s^{-1}	$SLP\cdot R \rightarrow SL+RP$
k6b	$0.001 \text{ nM}^{-1} \text{ s}^{-1}$	$SL+RP \rightarrow SLP\cdot R$
k7f	19.8 s^{-1}	$SP\cdot R \rightarrow S+RP$
k7b	$0.001 \text{ nM}^{-1} \text{ s}^{-1}$	$S+RP \rightarrow SP\cdot R$
k8f	$0.0238 \text{ nM}^{-1} \text{ s}^{-1}$	$S+RP \rightarrow RP\cdot S$
k8b	0.2 s^{-1}	$RP\cdot S \rightarrow S+RP$
k9	0.5 s^{-1}	$RP\cdot S \rightarrow S+R (+Pi)$
k10f	0.001 s^{-1}	$SLP\cdot R \rightarrow SP\cdot R$
k10b	0.001 s^{-1}	$SP\cdot R \rightarrow SLP\cdot R$
k11	0.0001 s^{-1}	$RP \rightarrow R (+Pi)$
k12	0.0001 s^{-1}	$SP \rightarrow S (+Pi)$
k13	0.0001 s^{-1}	$SLP \rightarrow SL (+Pi)$

Table S1. Rate constants for signalling reactions illustrated in Fig. 1 of the main text. Numbers in the parameter names correspond to the reaction numbers, and the letters f, b denote forward and backward reaction rates respectively for reversible reactions. The third column gives the reaction corresponding to each rate constant—species in brackets were not considered explicitly in the model, and the units for the rate constants reflect this. Note that $1\text{nM} \approx 1$ molecule per cell.

The rate constants used in our signalling model (Fig. 1 main text) are shown in Table S1. All values were estimated accounting for physical limitations. For example, second-order reactions cannot occur faster than the diffusion-limited rate. We originally tried to fit *in vitro* kinetics data from well-studied two-component systems in order to deduce the parameter values more precisely. However, the model did not give reasonable results with these data. More specifically, reaction rates were generally unrealistically slow, leading to signalling response times on the order of cell generation times. We hypothesised that this was due to the enzymes being much less efficient *in vitro* than *in vivo*. Many rate constants also have not yet been measured in the laboratory. Therefore, we did not rely heavily on published data regarding specific reaction rates for specific two-component systems. Instead, we focused on matching our model with more global properties. Examples of these global properties are a reasonably sensitive input-output response curve, reasonable response times and observed robustness of the output with respect to variations in the total protein concentrations.

For the autoregulatory studies, further reactions were considered relating to protein production, degradation, dimerisation and DNA-binding. These reaction rate constants were estimated, since measured values can vary greatly depending on factors such as the promoter strength and incorporation of activated degradation, and we were not aiming to match a particular system in our study. The parameters used that were found to give bistability are shown in Table S2. Note that kd2 can be interpreted not only as a protein degradation rate, but also as a cell growth rate. This alternative interpretation was used in our final study on the effects of cell growth and division on the system.

Parameter	Estimated Value	Reaction
kp1	0.003 nM s ⁻¹	$\emptyset \rightarrow \text{mRNA}$ (constitutive)
kp2	0.1 nM s ⁻¹	$\emptyset \rightarrow \text{mRNA}$ (RP-activated)
kp3	0.002 s ⁻¹	mRNA \rightarrow mRNA+R, mRNA \rightarrow mRNA+S
kd1	0.002 s ⁻¹	mRNA $\rightarrow \emptyset$
kd2	0.0005 (0.00025) s ⁻¹	R $\rightarrow \emptyset$, RP $\rightarrow \emptyset$, S $\rightarrow \emptyset$, SP $\rightarrow \emptyset$, SLP $\rightarrow \emptyset$
kDimerf	0.001 nM ⁻¹ s ⁻¹	RP+RP \rightarrow RP ₂
kDimerb	1 s ⁻¹	RP ₂ \rightarrow RP+RP
kActf	0.001 nM ⁻¹ s ⁻¹	RP ₂ +DNA \rightarrow DNA*
kActb	0.01 s ⁻¹	DNA* \rightarrow RP ₂ +DNA

Table S2. Rate constants for the additional reactions considered when modelling autoregulatory signalling (producing bistability). These constants were used for simulations both with and without cell division; the only difference was in the choice of kd2. For this case, the parameter for no cell division is shown, with the value used for cell division simulations also shown in brackets. The value is lower for cell division in order to allow protein populations to at least double within one cell cycle. In the parameter names, ‘p’ denotes reactions relating to mRNA/protein production, ‘d’ denotes degradation, ‘Dimer’ denotes dimerisation, and ‘Act’ denotes activation of the RP-responsive promoter. DNA and DNA* species thus represent the inactive and active states of the RP-responsive promoter respectively. Note that kp3 and kd2 correspond to multiple reactions, in these cases all the relevant reactions are listed.

Finally, we used (stochastic) delay times in transcription and translation using average transcription/translation rates from *Escherichia coli* (*E. coli*) and sequence lengths from the PhoP/PhoQ two-component system (Table S3).

Reaction	Distribution	Mean	Variance
Rp1a	gamma(0.018,672)	12.096	8128.512
Rp2b	gamma(0.018,1461)	26.298	38421.378
Rp3a	gamma(0.059,223)	13.157	2934.011
Rp3b	gamma(0.059,486)	28.674	13935.564
Rp2	gamma(0.018,672)	12.096	8128.512

Table S3. An overview of all reactions that were modelled with a delay. In all cases, a gamma-distribution was used; values shape and scale parameters are given in brackets.

2 Two-component signalling robustness - mathematical analysis

Considering the signalling network shown in Fig. 1 of the main text, we can show mathematically that the output RP levels are robust to variations in total numbers of response regulator and histidine kinase proteins (R_T and S_T). This analysis requires minimal assumptions to be made, which are described next.

In steady state, there should be no net accumulation or loss of phosphate from the system (so input and output rates should balance). Using mass-action kinetics, and rate constant labels shown in Table S1, we thus have:

$$\frac{d[P]}{dt} = k_2[SL] - k_9[RP \cdot S] = 0 \quad (\text{S1})$$

where we have represented phosphate by P and used square brackets to denote species concentrations (since this analysis is deterministic and hence assumes continuous variables). The effect of autodephosphorylation reactions 11, 12, and 13 is negligible compared to reactions 2 and 9 (Table S1). For this reason, we neglected these reactions. Now, if we assume that reaction 1 ($S+L \leftrightarrow SL$) occurs relatively rapidly, such that the three species are (approximately) in thermodynamic equilibrium, we have:

$$k_1f[S][L] = k_1b[SL] \quad (\text{S2})$$

Rewriting gives

$$[SL] = \frac{k_1f}{k_1b}[S][L] \quad (\text{S3})$$

Furthermore, in steady state $[RPS]$ is constant, so considering reactions 8 and 9 we have:

$$k_8f[RP][S] = (k_8b + k_9)[RP \cdot S] \quad (\text{S4})$$

Rewriting gives

$$[RP \cdot S] = \frac{k_8f}{k_8b + k_9}[RP][S] \quad (\text{S5})$$

Now substituting Equations (S3) and (S5) in Equation (S1) we find:

$$\frac{dP}{dt} = k_2 \frac{k_1f}{k_1b}[S][L] - \frac{k_9 \cdot k_8f}{k_8b + k_9}[RP][S] = 0 \quad (\text{S6})$$

Assuming $[S] \neq 0$ (valid for the deterministic case), we can divide the above equation by $[S]$ to obtain:

$$[RP] = k2 \frac{k1f (k8b + k9)}{k1b \cdot k9 \cdot k8f} [L] \quad (S7)$$

which is the remarkable result that in steady state the concentration of RP increases linearly in proportion to the signal strength, independent of any other species in the network. We have, however, implicitly assumed that $[RP]$ can increase without bound in this analysis. In reality, $[RP]$ is bounded by R_T , which explains the saturating profile shown in Fig. 2 of the main text.

3 Converting Deterministic Bistability Model Parameters to Stochastic Model input

In the deterministic model, we model the proportion of time for which DNA is activated by the Hill function,

$$P(\text{active}) = \frac{[RP]^n}{Kd^n + [RP]^n} \quad (S8)$$

where Kd is the microscopic dissociation constant of RP to its regulatory site, n is the multimerisation of RP required to bind to DNA, $[RP]$ is the concentration of RP in nM, and 1 nM is approximately equivalent to 1 molecule per cell in *E. coli*.

In order to make our stochastic model as realistic as possible, we wished to decompose the autoregulation of RP into the two fundamental processes of RP multimerisation and DNA binding, using only mass-action rate constants to parametrise the kinetics (Fig. S1). Since dimers are the simplest multimers that can lead to bistable RP expression, and the *E. coli* response regulator protein OmpR is known to dimerise before having regulatory functionality¹, we chose $n = 2$. Below we relate Kd to the four rate constants defined in Table S2 and used in the stochastic model.

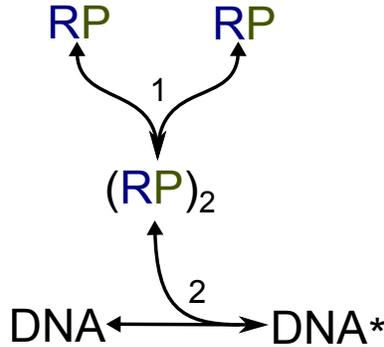


Fig. S1. RP dimerisation and binding to autoregulatory DNA site. Reactions are labelled ‘1’ and ‘2’ instead of ‘Dimer’ and ‘Act’ respectively (from Table S2) for ease of reading, but should not be confused with reactions 1 and 2 from the signalling network diagram in Fig. 1 of the main text. Products of forward reactions are indicated by pointed arrows, while products of back-reactions are indicated by triangular arrows. Reaction rates $k1f, k1b, k2f, k2b$ referred to in the text correspond to the forward and backward rates of reaction 1 and the forward and backward rates of reaction 2 respectively.

In steady state, the time-averaged concentration of active and inactive RP-DNA binding sites in the cell (denoted $[DNA*]$ and $[DNA]$ respectively) is constant, hence:

$$k2f[RP_2][DNA] = k2b[DNA*] \quad (S9)$$

Similarly, the time-averaged concentration of dimerised RP (denoted RP_2) is also constant in steady state, hence:

$$k1f[RP]^2 + k2b[DNA*] = k1b[RP_2] + k2f[RP_2][DNA] \quad (S10)$$

By rearranging Equations (S9) and (S10), we obtain, respectively:

$$\frac{[DNA*]}{[DNA]} = \frac{k2f}{k2b}[RP_2] \quad (S11)$$

$$[RP_2] = \frac{k1f[RP]^2 + k2b[DNA*]}{k1b + k2f[DNA]} \quad (S12)$$

Next, combining both Equations (S11) and (S12) then yields:

$$\frac{[DNA*]}{[DNA]} = \frac{k2f}{k2b} \cdot \frac{k1f[RP]^2 + k2b[DNA*]}{k1b + k2f[DNA]} \quad (S13)$$

Now, using the Hill equation,

$$\frac{[DNA*]}{[DNA] + [DNA*]} = \frac{[RP]^2}{K_d^2 + [RP]^2}, \quad (S14)$$

we deduce that when $[RP] = K_d$, after time-averaging we have $[DNA*] = [DNA] = 0.5$. Note that a Hill coefficient of 2 has been assumed, which can be interpreted as RP binding only when it is in a dimerised form. Substituting this result in Equation (S13) gives the final relation:

$$K_d^2 = \frac{k2b}{k2f} \cdot \frac{k1b}{k1f} \quad (S15)$$

We can estimate K_d using the following arguments. Suppose first that there are no competing binding sites for RP in the cell. Then it is reasonable to suppose that having two RP molecules per cell (i.e. one RP dimer) may be sufficient to activate $[DNA*]$ half the time. This suggests that—using units of nM for concentrations—we can estimate that $K_d \approx 2$. However, since RP molecules are responsible for regulating the entire cellular response to L in the two-component system, there are usually multiple regulatory sites on the genome to which RP can bind. For example, estimates for the number of genes activated by the phosphorylated PhoP response regulator in *E. coli* range from 40 up to 236^{2,3}. If we assume that RP has the same affinity to all potential binding sites, the effective number of RP molecules available for binding to a particular binding site is divided by two times the total number of sites (since each site is only bound by an RP dimer). Denoting m as the total number of binding sites in the cell, we can modify Equation (S14) to account for this effect as:

$$\frac{[DNA*]}{[DNA] + [DNA*]} = \frac{([RP]/2m)^2}{K_d^2 + ([RP]/2m)^2} = \frac{[RP]^2}{(2mK_d)^2 + [RP]^2} \quad (S16)$$

where $[DNA*]$ and $[DNA]$ still refer only to the activation state of the autoregulatory binding site, and $2mK_d$ can be considered as the new ‘effective’ dissociation constant. Thus, if we consider $m = 50$, we implement an effective K_d of 100. Substituting this into Equation (S15) then gives:

$$10000 = \frac{k2b}{k2f} \cdot \frac{k1b}{k1f} \quad (S17)$$

Any combination of the four parameters $k1f, k1b, k2f, k2b$ that satisfies Equation (S17) is sufficient to account for a K_d choice of 100. However, additional constraints were placed on $k2f$ and $k2b$ so that transitions between activated and basal transcription rates occurred on a minute timescale; this allows multiple transitions to occur during a single cell cycle yet still allows a degree of transcriptional ‘bursting’, necessary to produce bistability. The final parameters chosen are shown in Table S2.

4 StochPy Cell Division extension

Important sources of non-genetic cellular heterogeneity are cell growth and division. Cell division is a stochastic process which can have a significant impact on molecular copy numbers in single-cells. An evident source of cell-to-cell heterogeneity is the stochastic partitioning of molecules at cell division which distributes most molecules unequally between both daughter cells⁴. In addition, cell-to-cell heterogeneity can also be caused by heterogeneity of cell volume, which is caused by three different processes: (1) heterogeneity in the mother cell volume at division, (2) imprecise volume division from mother to daughter cells, and (3) volume growth rate heterogeneity. We incorporated, with the exception of a heterogeneity of growth rates, these sources of non-genetic cellular heterogeneity in Stochastic modelling in Python (StochPy)⁵.

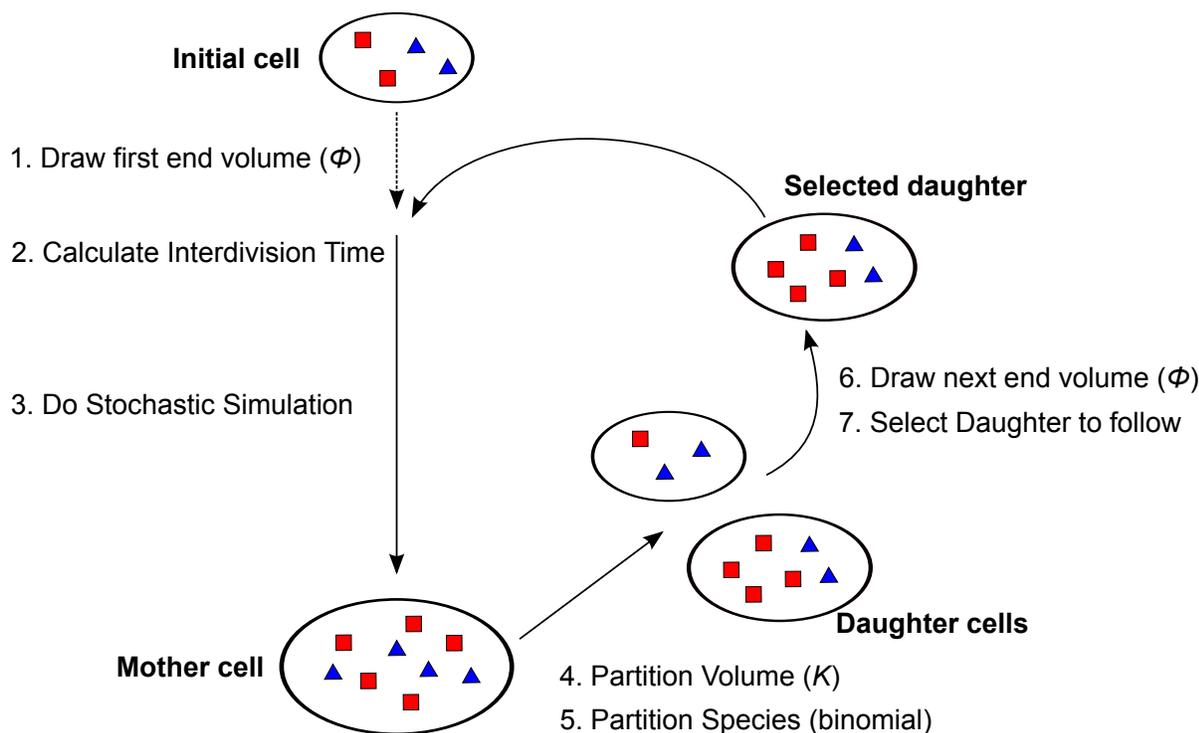


Fig. S2. Cartoon which illustrates how StochPy incorporates cell division events in stochastic simulations. Blue triangles and red squares represent molecules which are stochastically partitioned at cell division between both daughters. In this example, the largest daughter cell—which also has a larger chance to be selected—is selected for the next generation. The stochastic simulation is interrupted at each cell division and stops when (1) the number of generations is reached, (2) the desired end time is reached, (3) the desired number of time steps is reached, or (4) all reactions are exhausted.

An overview of the incorporation of cell division into StochPy is shown in Figure S2. Note that a more detailed explanation will be provided when the next version of StochPy is released. Within StochPy, the growth function, initial cell volume, specific division properties, volume dependencies, and if present in your model, exact and non-dividing species have to be set before a simulation can be initiated. Of course, for each of these the default values can also be used. StochPy supports both exponential and linear volume growth. The time for a given daughter cell to reach division is called the interdivision time (IDT), which is determined from the specified growth function, the initial volume and the volume at division. By default, the initial cell volume is set to 1 fl, which corresponds to the cell volume of *E. coli*. Note that both the cell volume at division and the cell volume partitioning ratios are drawn from user-specified probability density functions

ϕ and K , respectively. Modelling processes in a non-constant volume environment affects the propensities of volume-dependent reactions, which are second and higher-order reactions. In StochPy, volume-dependent propensities are, by default, updated upon a change in cell volume as is shown in Equation (S18).

$$a_V(t) = \frac{a(t)}{V(t)^{\text{order}-1}} \text{ with order } \geq 2 \quad (\text{S18})$$

Here, $a(t)$ is the reaction propensity in an environment with a constant cell volume and $V(t)$ represents the time-dependent cell volume. Alternatively, StochPy allows users to modify the volume-dependency of each individual reaction. More specifically, StochPy allows users to specify, for each reaction, the order of the volume dependency. Note that cell volume is not modelled as an additional (continuous) reaction, but is updated after firing of each individual reaction and is therefore a discrete variable.

Finally, in StochPy users can specify species which should not be subject to stochastic partitioning at cell division. More specifically, StochPy distinguishes between non-dividing species (e.g. DNA if its replication is not modelled) and exact-dividing species (e.g. chromosomes). All other species are subject to stochastic partitioning at cell division which, in StochPy, also depends on the daughter cell volumes. Namely, a daughter with a larger volume has the tendency to inherit more molecules than the other daughter. More specifically, the probability that a molecule is inherited by a daughter cell is directly proportional to the volume ratio of daughter and mother cell:

$$P(\text{molecule inheritance} | V_{d_1}, V_m) = \frac{V_{d_1}}{V_m} \quad (\text{S19})$$

Here, V_{d_1} is the daughter cell volume and V_m is the mother cell volume. The number of molecules, with copy number n , inherited by daughter d_1 can thus be drawn as a random sample from a binomial distribution with n number of trials and success probability $\frac{V_{d_1}}{V_m}$ in each trial. Note that StochPy tracks a single lineage through time, which means that after division, only one daughter is selected with a certain probability for the next generation.

4.1 Settings for reproduction of results

In Table S4, we provide the settings used for the explicit simulation of cell division.

Setting	Value
Initial cell volume	1 fl
Exponential Growth rate	0.00025 s^{-1}
ϕ (mother cell volume)	Beta(2,2) + 2
K (partition distribution)	Beta(5,5)
Non-dividing species	DNA and DNA*

Table S4. An overview of the settings used in the cell division simulations. Note that an exponential growth rate of 0.00025 s^{-1} corresponds to a doubling time of about 46 minutes.

References

- [1] C. M. Barbieri, T. Wu and A. M. Stock, *J Mol Biol*, 2013, **425**, 1612–1626.
- [2] S. Minagawa, H. Ogasawara, A. Kato, K. Yamamoto, Y. Eguchi, T. Oshima, H. Mori, A. Ishihama and R. Utsumi, *Journal of Bacteriology*, 2003, **185**, 3696–3702.
- [3] Y. Ishihama, T. Schmidt, J. Rappsilber, M. Mann, F. U. Hartl, M. J. Kerner and D. Frishman, *BMC Genomics*, 2008, **9**, 102.

[4] D. Huh and J. Paulsson, *Nat Genet*, 2011, **43**, 95–100.

[5] T. R. Maarleveld, B. G. Olivier and F. J. Bruggeman, *PLoS One*, 2013, **8**, e79345.