S1. SUPPLEMENTARY FIGURES

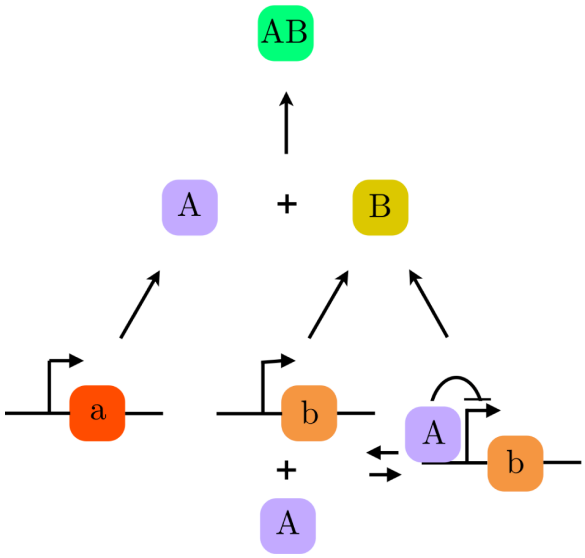

FIG. S1: **Illustration of the genetic reaction network in Eq. S1**

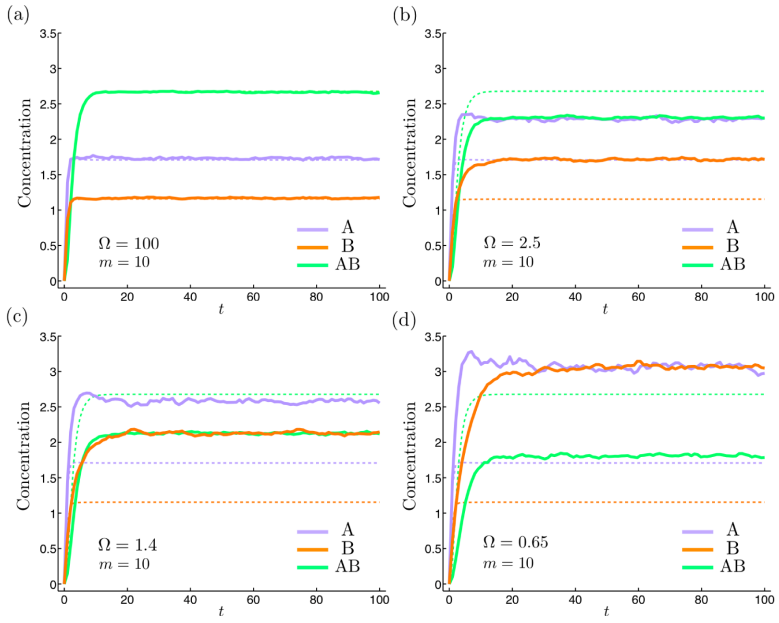

FIG. S2: **Ensemble-averaged concentrations of proteins A, B and their heterodimer AB as a function of time for a genetic reaction network confined in compartments of different volumes** The data (solid lines) are obtained from exact stochastic simulations using PDM SSA. The gene network is schematically shown in Fig. S1. The compartment volumes are: (a) $\Omega = 100$, (b) $\Omega = 2.5$, (c) $\Omega = 1.4$, (d) $\Omega = 0.65$. The rate constants are fixed to $k_1' = 2.0\Omega/m$, $k_2 = 0.34$, $k_3' = 0.74\Omega/m$, $k_4 = 0.1$, $k_5 = 0.72$, $k_6 = 0.53$, $k_7' = 2.0\Omega$, $k_8' = 2.0\Omega$, $k_9' = 2.0\Omega$ and the burst size to $m = 10$. The dashed lines show the RE predictions for the same parameter values. The simulations confirm the theoretical prediction of a discreteness-induced inversion below the largest critical volume, $\Omega_{\mathrm{A,AB}} = 3.35$, for species A and AB. Below this largest critical volume the ordering of steady-state concentrations predicted by the RE is incorrect. Note that time, concentration and volumes are in non-dimensional units.

S2. SUPPLEMENTARY METHODS

We present the existence of the concentration inversion effect in an *in silico* genetic network [54]:

$$\begin{aligned}
\mathrm{a} &\xrightarrow{k_1'} \mathrm{a} + m\mathrm{A}\\
\mathrm{A} &\xrightarrow{k_2} \emptyset\\
\mathrm{b} &\xrightarrow{k_3'} \mathrm{b} + m\mathrm{B}\\
\mathrm{B} &\xrightarrow{k_4} \emptyset\\
\mathrm{A} + \mathrm{B} &\xrightarrow{k_5} \mathrm{AB}\\
\mathrm{AB} &\xrightarrow{k_6} \emptyset\\
\mathrm{b} + \mathrm{A} &\xrightarrow{k_7'} \mathrm{bA}\\
\mathrm{bA} &\xrightarrow{k_8'} \mathrm{b} + \mathrm{A}\\
\mathrm{bA} &\xrightarrow{k_9'} \mathrm{bA} + \mathrm{B}.
\end{aligned} \tag{S1}$$

The k's are the macroscopic reaction rates and the reactions occurs in a reactor of volume Ω. In this network, a and b are genes. The corresponding proteins encoded by these genes are A and B, respectively. The synthesis of these proteins occur in bursts of size m. Proteins A and B form a heterodimer AB. This network models a frequently found motif in gene networks [54]. Protein A represses gene b at the transcriptional level. Protein B, however, acts through heterodimerization with A, the heterodimer AB being unable to repress gene b. Thus, if A is high, b is repressed, but if B is high, all of the As are heterodimerized, and B remains high [54] (see Supplementary Fig. S1 for an illustration).

Note that since there is one gene of a and one gene of b, the concentration of gene a in the rate equation (RE) model is constant and equal to $[a] = 1/\Omega$, whereas the concentration of b and bA are $[\mathrm{b}] = x/\Omega$ and $[\mathrm{bA}] = y/\Omega$, respectively, where $x + y = 1$. This follows from the fact that gene b exists either in unbound or bound form. The term in the RE model for the rate of production of A due to transcription is $k_1'[a]m = k_1'm/\Omega$; other terms such as those leading to transcription of B and those modeling the reversible interaction of b and A also show this explicit volume dependence. This implies that the solution of the RE model for this system (and indeed for any genetic system) will depend on the volume of the compartment. This dependence stems from the fact that the copy number of genes is fixed, independent of the volume. Taking into account discreteness generates a volume dependence on top of this pre-existing volume dependence. To clearly distinguish the first from the second, we scale the rates such that we eliminate the volume dependence of the REs. This is achieved by setting $k_1' = k_1\Omega/m$, $k_3' = k_3\Omega/m$, $k_7' = k_7\Omega$, $k_8' = k_8\Omega$ and $k_9' = k_9\Omega$. Note that here we have additionally scaled some rates by the burst size m in order to eliminate also dependence on the latter; this is convenient, but not essential.

We further set the parameters of this network so that the system is monostable: $k_1 = 2.0$, $k_2 = 0.34$, $k_3 = 0.74$, $k_4 = 0.1$, $k_5 = 0.72$, $k_6 = 0.53$, $k_7 = 2.0$, $k_8 = 2.0$ and $k_9 = 2.0$. We investigate the existence of the concentration inversion effect in the concentration levels of proteins A, B and their heterodimer AB. Initially, the number of molecules of gene a and b are set to 1, and the concentrations of all other chemical species are 0.

S3. SUPPLEMENTARY DISCUSSION

Using burst size $m = 10$, the time evolution of the average concentrations of proteins A, B and their heterodimer AB predicted by the chemical master equation (CME) are shown in Supplementary Fig. S2 for four different reactor volumes $\Omega = 100, 2.5, 1.4, 0.65$. At large Ω, the concentration predicted by the RE and the average concentration predicted by the CME are in good agreement (Supplementary Fig. S2(a)). The theory using the effective mesoscopic reaction rate equation (EMRE) predicts the largest critical volume as $\Omega_{\mathrm{A,AB}} = 3.35$. The simulations indeed show that at $\Omega = 2.5$ the average steady-state concentrations of A and AB are equal (Supplementary Fig. S2(b)). Below this volume the ordering of the average steady-state concentration levels of A and AB is reversed with respect to their concentration levels at large Ω (Supplementary Figs. S2(c) and (d)). The theory further predicts two addition critical volumes $\Omega_{\mathrm{B,AB}} = 1.93$ and $\Omega_{\mathrm{A,B}} = 0.41$. The values of these critical volumes from stochastic simulations are 1.4 and 0.65, respectively (Supplementary Figs. S2(c) and (d)). Note that all possible concentration inversions are seen for this genetic network.

For $m = 1$, the theoretical predictions of the critical volumes are $\Omega_{\mathrm{A,B}} = 0.39$, $\Omega_{\mathrm{B,AB}} = 0.32$ and $\Omega_{\mathrm{A,AB}} = 0.27$. The corresponding values from stochastic simulations are 0.4, 0.28 and 0.18, respectively. For $m = 30$, the theoretical predictions of the critical volumes are $\Omega_{\mathrm{A,AB}} = 10.38$, $\Omega_{\mathrm{B,AB}} = 5.51$ and $\Omega_{\mathrm{A,B}} = 0.43$. The corresponding values from simulations are 8, 4 and 1.2, respectively. From these results we observe that the critical volumes increase with

increasing m, as indeed was also the case for the trimerization reaction studied in the main text. In addition, the largest critical volume predicted by the EMRE is the most accurate in comparison to the ones observed from exact stochastic simulations. This volume is also the most important, since it demarcates regions with inversion from regions with no inversion.

S4. SUPPLEMENTARY REFERENCES

[54] Francois, P. & Hakim, V. Design of genetic networks with specified functions by evolution *in silico*. *Proc. Natl. Acad. Sci. USA* **101**, 580–585 (2004).