

In the format provided by the authors and unedited.

A universal biomolecular integral feedback controller for robust perfect adaptation

Stephanie K. Aoki^{1,2}, Gabriele Lillacci^{1,2}, Ankit Gupta^{1,2}, Armin Baumschlager¹, David Schweingruber¹ & Mustafa Khammash^{1*}

¹Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland. ²These authors contributed equally: Stephanie K. Aoki, Gabriele Lillacci, Ankit Gupta.

*e-mail: mustafa.khammash@bsse.ethz.ch

Supplementary Materials for

A universal biomolecular integral feedback controller for robust perfect adaptation

Stephanie K. Aoki^{1,*}, Gabriele Lillacci^{1,*}, Ankit Gupta^{1,*}, Armin Baumschlager¹, David Schweingruber¹,
and Mustafa Khammash^{1,†}

¹ETH Zürich, Department of Biosystems Science and Engineering,
Mattenstrasse 26, 4058 Basel, Switzerland

*These authors contributed equally to this work.

†Correspondence: mustafa.khammash@bsse.ethz.ch

This PDF files includes:

Supplementary Text
Supplementary Methods
Tables S1, S2

Supplementary Text

S1 Modeling	3
S1.1 AIF circuit with dilution and negative disturbance	4
S1.1.1 Asymptotic behavior of the steady-state error	4
S1.1.2 Behavior of the steady-state error for realistic parameter values	5
S1.2 Accounting for actuator saturation	5
S1.3 Dependency of adaptation region on actuator gain k	6
S1.4 Parameter values used in the computational analysis	6
S1.4.1 Dilution rates δ	6
S1.4.2 Native degradation rate of the output γ	7
S1.4.3 Strength of negative disturbance w	7
S1.4.4 Annihilation rate η	7
S1.4.5 Controller constitutive production rate μ	8
S1.4.6 Regulated production rates θ and k	8
S1.4.7 Maximal saturated production rate V_{max}	8
S2 The universality of antithetic feedback controllers for robust perfect adaptation of noisy biomolecular networks	8
S2.1 Preliminaries	8
S2.1.1 The stochastic model for chemical kinetics	8
S2.1.2 The notion of stability	10
S2.1.3 Perfect adaptation and the Poisson equation	10
S2.1.4 Sensitivity and Robustness	11
S2.2 Main Results	12
S2.2.1 The biomolecular control design problem	12
S2.2.2 Parametric characterization of all RPA-achieving controller motifs	14
S2.2.3 The generalized antithetic motif as the universal RPA-achieving motif	19
S2.2.4 RPA in the deterministic setting	22
S3 Supplementary Methods	27

Supplementary Tables

S1 Suicide plasmids for allelic exchange	31
S2 Strains, plasmids and primers	35

Supplementary Text

S1 Modeling

The closed-loop antithetic integral feedback (AIF) circuit will achieve robust perfect adaptation, even when the dynamics are intrinsically noisy and modeled as a *continuous-time Markov chain* (CTMC) (see Section S2.1.1). However direct analysis and extensive simulations of this stochastic model, using Gillespie’s *stochastic simulation algorithm* [34] for example, is quite challenging and computationally demanding. To circumvent this issue, we appeal to the well-known limit theorem of Kurtz [35] which proves that when species constituting a reaction network are present in large copy-numbers, then the CTMC dynamics is well-approximated by the corresponding ODE dynamics. Even though this classical result only holds over finite time-intervals, it can be extended to the infinite time-interval $[0, \infty)$ in situations where the limiting ODE dynamics has a unique globally attracting fixed point (see Theorem 2.7 in [36]). In our engineered biological system we are comfortably in the region of high copy-numbers, a requirement to ensure that the fluorescent levels remain detectable. Therefore, to analyze the closed-loop antithetic integral feedback (AIF) circuit, we use a 3-dimensional ODE model of the circuit depicted in Fig. 1c. The circuit regulates a single gene that produces an output protein of interest X . In the antithetic topology, the regulation is carried out by two controller species, namely an *actuator* species (denoted Z_1) and a *sensor* species (denoted Z_2). A key feature of AIF is that Z_1 and Z_2 must *annihilate*, i.e. abolish each other’s biological activity upon reacting, for example by forming a high-affinity, biologically inert dimer. This is also known as *molecular sequestration*.

The perfect adaptation property of the AIF circuit requires that the two controller species are lost in equal quantities only due to the annihilation reaction. This condition is not met in live growing cells, since the cell growth will introduce an additional degradation term of the form $-\delta X$, where δ is the growth rate and X is the concentration of the chemical species being considered. We refer to this degradation mechanism as *dilution*. Dilution becomes a particular concern in fast-growing cells, such as *Escherichia coli* (*E. coli*), which double in size and divide every 20-30 minutes when in exponential growth phase.

In the ideal AIF circuit it is also assumed that there are no limitations in the production rates of circuit components. However, it is likely that the promoters used to drive the expression of the genes that code for the proteins involved in the circuit will not have infinite transcriptional capacity. This may reduce the space of achievable set-points, and affect the disturbance rejection properties of AIF.

We investigate these issues by analyzing dynamical models of non-ideal AIF circuits that account for dilution and saturation. We elucidate the parameters that are important to the circuit’s performance and we characterize the regions of the parameter space in which the performance is not far from that of an ideal AIF-controlled system. In the ideal case in which no dilution is present, the actuator species is produced constitutively and lost only through the annihilation reaction. The sensor species is produced proportionally to X and also lost only due to annihilation. The output is produced proportionally to the actuator and degraded according to some rate γ . These reactions can be summarized in the following equations:

$$\begin{cases} \dot{Z}_1 = \mu - \eta Z_1 Z_2 \\ \dot{Z}_2 = \theta X - \eta Z_1 Z_2 \\ \dot{X} = k Z_1 - \gamma X, \end{cases} \quad (\text{S1})$$

where all parameters are assumed strictly positive and all state variables are assumed non-negative. It is easy to see that the steady-state of X in (S1) is given by:

$$\bar{X} = \frac{\mu}{\theta}. \quad (\text{S2})$$

We note that (S2) defines the controller parameters regulating the set-point: the output steady state is robust to variations in any other parameters.

To study how the presence of dilution affects the disturbance rejection properties of the circuit, we add a degradation term with rate δ to all the 3 species in the model. We will further add a *negative disturbance* in the form of an additional degradation term for X : this simulates the effect of an orthogonal protease targeting X .

S1.1 AIF circuit with dilution and negative disturbance

For an AIF circuit with dilution subject to a negative disturbance we consider the following model:

$$\begin{cases} \dot{Z}_1 = \mu - \eta Z_1 Z_2 - \delta Z_1 \\ \dot{Z}_2 = \theta X - \eta Z_1 Z_2 - \delta Z_2 \\ \dot{X} = k Z_1 - (\gamma + w + \delta) X. \end{cases} \quad (\text{S3})$$

The additional parameters δ and w are assumed non-negative. The steady state of X can be computed analytically and is given by the following expression:

$$\bar{X}^{(nd)} = \frac{2\delta\mu k}{\delta + \gamma + w} \left(\sqrt{(\delta^2 + \mu\eta)^2 + \frac{4\delta\mu\theta\eta k}{\delta + \gamma + w}} + \delta^2 - \mu\eta \right)^{-1}, \quad (\text{S4})$$

where the superscript (nd) denotes the addition of the negative disturbance. Using the same formula and substituting $w = 0$, one can also obtain the steady state of X when no disturbance is present, which we denote $\bar{X}_0^{(nd)}$. The steady-state error is then given by:

$$\bar{E}^{(nd)} = \bar{X}_0^{(nd)} - \bar{X}^{(nd)}. \quad (\text{S5})$$

S1.1.1 Asymptotic behavior of the steady-state error

The steady-state error must reduce to zero if there is no disturbance ($w = 0$) or no dilution ($\delta = 0$). The limit

$$\lim_{w \rightarrow 0} \bar{E}^{(nd)} = 0$$

holds by definition of $\bar{E}^{(nd)}$ and as a sanity check we can verify the following limit:

$$\lim_{\delta \rightarrow 0} \bar{E}^{(nd)} = 0.$$

We now ask whether it is possible to make the error small by choosing limiting values of other controller parameters. A natural one to consider is the *sensor gain*, which in our setup is the sensor production rate θ . Does the error go to zero as this parameter becomes large? To answer this question, we can not simply take the limit as θ goes to infinity, because as θ goes to infinity the steady state of X goes to zero. So, naturally one would get $\bar{E}^{(nd)} = 0$, simply because $\bar{X}_0^{(nd)}$ and $\bar{X}^{(nd)}$ are both zero. Instead, as θ grows, we also need to increase μ accordingly. We substitute $\mu = r\theta$, so that the ratio of μ and θ stays constant. In other words, we let θ go to infinity along the line $\mu = r\theta$. We find:

$$\begin{aligned} E^{(1)} = \lim_{\theta \rightarrow \infty} \bar{E}^{(nd)} \Big|_{\mu=r\theta} &= \frac{r}{2(\delta + \gamma + w) \left(\sqrt{\eta r \left(\frac{4\delta k}{\delta + \gamma + w} + \eta r \right)} - \eta r \right)} \\ &\times \left(-\eta r (\delta + \gamma + w) - (\delta + \gamma + w) \left(\sqrt{\eta r \left(\frac{4\delta k}{\delta + \gamma} + \eta r \right)} - \sqrt{\eta r \left(\frac{4\delta k}{\delta + \gamma + w} + \eta r \right)} \right) \right. \\ &\quad \left. + \sqrt{\frac{(\eta r (\delta + \gamma) + 4\delta k) (\delta + \gamma + w) (\eta r (\delta + \gamma + w) + 4\delta k)}{\delta + \gamma}} - 4\delta k \right) \quad (\text{S6}) \end{aligned}$$

We find an expression that is *not* identically zero, suggesting that high gain alone is not sufficient to make the error vanish. Interestingly, if the annihilation rate η also goes to infinity, one obtains exactly zero error:

$$\lim_{\eta \rightarrow \infty} E^{(1)} = 0.$$

Consistent with this finding, if one takes the same sequence of limits of (S4), the expression of the steady state in the ideal case (S2) is recovered.

Another way to make the error small is to choose k sufficiently small and the product θk sufficiently large. This can be for example verified by substituting $\mu = r\theta$ and $\theta = k^{-2}$ in (S5), and then taking the limit for $k \rightarrow 0$. The limit of the steady-state error is 0, and the limit of the steady state becomes r as in the ideal case.

A third way is to choose η sufficiently large and the product θk sufficiently large. This can be verified similarly to the case above, for example by substituting $\mu = r\theta$ and $\theta = 1$ in (S5), and then taking the limit for $k \rightarrow \infty$. Again, the limit of the steady-state error is 0, and the limit of the steady state is r .

S1.1.2 Behavior of the steady-state error for realistic parameter values

The asymptotic analysis presented above suggests that there are at least three ways of reducing the steady-state error, and that by appropriately tuning the controller parameters one can circumvent the loss of perfect adaptation due to dilution. We now investigate numerically the dependency of the error on the parameters. To that end, we define the relative steady-state error as:

$$\bar{E}_r^{(nd)} = \frac{\bar{X}_0^{(nd)} - \bar{X}^{(nd)}}{\bar{X}_0^{(nd)}}. \quad (S7)$$

We evaluate (S7) numerically and look for regions of the parameter space where the error is within 5%. Realistic values or ranges for the parameters were chosen as discussed in Section S1.4. To simplify the visualization of the results, we fix δ , η and γ and w (for which realistic values could be chosen with more confidence), and we vary μ , θ , and k within 4 orders of magnitude. For the latter parameters, we consider a grid of $100 \times 100 \times 100$ logarithmically spaced values (total of 10^6 points), and we evaluate (S7) at each point. The results are presented in Extended Data Fig. 1 b, c for two different values of k . As it can be seen from the figure, for a fixed k the dependency of the output steady state on the controller parameters is still well approximated by a linear function of the ratio of μ and θ . Regarding the disturbance rejection performance the following observations can be made: (i) for fixed θ and k , there is a *minimum* value of μ required for small steady-state error; (ii) for fixed μ and k , there is an *interval* of θ in which the circuit displays small steady-state error.

We then repeated the calculations for two additional values of δ : $\ln(2)/120 \text{ min}^{-1}$, and $\ln(2)/1200 \text{ min}^{-1}$, corresponding to typical growth rates for yeast and mammalian cells respectively. As expected, the region in which the system shows near-perfect adaptation expands as δ becomes small (Extended Data Fig. 1b).

S1.2 Accounting for actuator saturation

When a negative disturbance is applied to the AIF circuit, the feedback regulation results in a lower production of the sensor species Z_2 . This causes a rise in the levels of the free actuator species Z_1 , which in turn increases the output production rate. In (S3) we assumed that such production rate is linearly related to the actuator concentration. The output promoter, however, does not have infinite capacity. To model this, we replace the linear function kZ_1 with a Michaelis-Menten function of the form:

$$f(Z_1) = \frac{V_{max}Z_1}{K + Z_1},$$

where V_{max} is the maximum allowable production rate. The choice of V_{max} is discussed in Section S1.4. To set K , we require f to have a slope comparable to the one of the line in its linear range. We do that by

assuming that the linear function is the line connecting the origin with the point where $f(Z_1) = V_{max}/2$. This yields:

$$K = \frac{V_{max}}{2k}.$$

When saturation is considered, analytical expressions for the steady states are no longer available. Resorting to numerical simulations, we can still characterize the regions with less than 5% relative steady-state error (Extended Data Fig. 1d). The main difference with Extended Data Fig. 1b, c is that for each fixed θ and k there is now an *interval* of μ values resulting in good disturbance rejection performance.

S1.3 Dependency of adaptation region on actuator gain k

As shown in Extended Data Fig. 1b, c, increasing the actuator gain k has the effect of shifting the near-perfect adaptation region of the AIF circuit toward smaller values of μ and θ . This has important implications in designing a circuit that has good adaptation performance around a desired set-point. To further illustrate this effect, suppose we are interested in regulating the output around 10000 nM, and that the actuator gain of our circuit is $k = 0.06 \text{ min}^{-1}$. Extended Data Fig. 2 shows that it is possible to achieve near-perfect adaptation for that output level for values of μ and θ approximately larger than 1600 nM min⁻¹ and 0.1 min⁻¹ respectively. If the output gain is increased 10 fold to $k = 0.6 \text{ min}^{-1}$, the set-point of 10000 nM moves towards the interior of the adaptation region, and near-perfect adaptation can be achieved for smaller values of μ and θ , roughly larger than 160 nM min⁻¹ and 0.015 min⁻¹ respectively (Extended Data Fig. 2). One practical way to increase the gain k is to increase the number of copies of the output gene: this is the motivation for using two copies of the output gene in our final circuit design.

S1.4 Parameter values used in the computational analysis

We now discuss how values and ranges of the parameters for the numerical simulations presented above were chosen. To interpret these, it is helpful to consider that 1 molecule in an *E. coli* cell of the volume of 1 fL will result in a concentration of approximately 1.7 nM:

$$\frac{1 \text{ molecule}}{6.022 \cdot 10^{23} \text{ molecules mol}^{-1}} \cdot \frac{1}{10^{-15} \text{ L}} = 1.66 \text{ nM}.$$

S1.4.1 Dilution rates δ

Dilution rates can be calculated on the basis of exponential decay processes related to cell growth. If a cell contains an initial amount X_0 of a chemical species X and no new X is produced, X will gradually dilute out due to cell growth. The corresponding differential equation:

$$\dot{X} = -\delta X$$

has solution:

$$X(t) = X_0 e^{-\delta t}.$$

We define the *doubling time* T_D as the average time it takes for a cell to double in size (which will eventually lead to cell division). After time T_D , the cell will contain half of the initial amount of X . Thus:

$$X(T_D) = X_0 e^{-\delta T_D} = \frac{X_0}{2},$$

which gives:

$$\delta = \frac{\log(2)}{T_D}.$$

The doubling times are well known for common laboratory organisms. We assumed 25 min for bacteria, 120 min for yeast and 1200 min (20 hours) for mammalian cells, in line with standard cellular biology knowledge. These respectively result in values of δ of 0.028 min^{-1} , 0.0058 min^{-1} and 0.00048 min^{-1} respectively.

S1.4.2 Native degradation rate of the output γ

We previously observed that fluorescent proteins without degradation tags are lost only due to dilution. In our circuit design output proteins however do have degradation tags. While these are not supposed to contribute any active degradation when the disturbance is not induced, we found that this is not actually the case: the proteins will still be less stable compared to untagged ones, presumably due to partial recognition of the tags by *E. coli* native proteases [37]. To model this effect, we set $\gamma = 0.5\delta$.

We remark that the fact that the output protein of interest degrades at a faster rate than dilution has to do with our specific implementation. However, it should not be viewed a requirement for the system to achieve near-perfect adaptation. In fact, within the parameter ranges we tested in our simulations, γ has little effect on the shape and size of the near-perfect adaptation region.

S1.4.3 Strength of negative disturbance w

A nominal value for this was chosen on the basis of Figure 1D of [37], where it is shown that different mflon protein degradation tags can take out anywhere from 10% to 90% of the initial protein amount. Based on the simple model:

$$\dot{X} = k - (\delta + w)X,$$

where X is protein amount, k is the production rate, δ is the dilution rate and w is the active degradation rate, one can write an equation for the relative drop due to active degradation at steady state:

$$\frac{\delta}{k} \left(\frac{k}{\delta} - \frac{k}{\delta + w} \right) = \frac{w}{\delta + w}.$$

So the drop is 50% when $w = \delta$, 90% when $w = 9\delta$, and so on. We assume an intermediate value of $w = 5\delta$.

S1.4.4 Annihilation rate η

The references available include *in vitro* measurements of sigma factor binding affinities [38], as well as previous modeling work [39]. Buchler and Louis summarize several previous studies in which binding affinities of proteins that display annihilation had been measured *in vitro* (Supplementary Table SII of [40]). For sigma and anti-sigma factor pairs the reported values range from $3.66 \cdot 10^{-3} \text{ nM}^{-1} \text{ min}^{-1}$ to $5.40 \cdot 10^{-2} \text{ nM}^{-1} \text{ min}^{-1}$.

The question of how well *in vitro* rates match the *in vivo* ones is not completely resolved, mostly due to the challenges associated with measuring the latter. For a large set of *E. coli* enzymes it has been shown that *in vitro* catalytic rates generally correlate well with the *in vivo* ones [41]. There are, however, some notable exceptions, in which differences of about 20-fold have been observed depending on the method used [42]. For the yeast galactose signaling network, Gencoglu *et al.* [43] report binding affinities *in vivo* of up to 40-fold stronger than their *in vitro* counterparts.

Given this uncertainty, and the indication that *in vivo* binding may be stronger than *in vitro* binding, we picked the value of $0.05 \text{ nM}^{-1} \text{ min}^{-1}$, closer to the upper limit of the *in vitro* range given above. Note that higher values will improve the circuit performance, so this nominal value should be viewed as a conservative lower bound.

S1.4.5 Controller constitutive production rate μ

To get a realistic upper bound, we referred to [44], where it is reported that maximal production rates for the two prototypical *E. coli* proteins LacI and LacZ are in the order of 100 molecules per minute. These rates were in turn derived from [45] and they refer to the genomic lac operon. Consistently with these, Buchler and Louis [40] give transcription factor synthesis rates of 0.5-50 nM min⁻¹ in bacteria.

Given that our circuits operate on a high copy plasmid with a ColE1 origin of replication yielding about 50-70 plasmid copies per cell [46], we multiplied the genomic rate by 60, to obtain 6000 molecules per minute. Using the equivalence between number of molecules and concentration given above, we approximated this to 10⁴ nM min⁻¹. This maximal value was then reduced by up to 4 orders of magnitude to obtain a search space for this parameter of 10⁰ to 10⁴ nM min⁻¹.

S1.4.6 Regulated production rates θ and k

When no saturation occurs, the regulated production rates are proportional to the concentration of regulator protein. To get a *lower* bound for these, we assume that the maximal rate of 10⁴ nM min⁻¹ occurs for the maximum possible concentration of regulator protein in the cell. If the regulator were constitutively produced from a high-copy plasmid, the maximal concentration at steady state would be:

$$R_{max} = \frac{10^4 \text{ nM min}^{-1}}{0.028 \text{ min}^{-1}} \approx 3.6 \cdot 10^5 \text{ nM}.$$

With this reasoning, one gets therefore a lower bound of:

$$\frac{10^4 \text{ nM min}^{-1}}{R_{max}} \approx 0.028 \text{ min}^{-1}.$$

We used this reference value to establish a search space of 4 orders of magnitude for θ and k of 10⁻² to 10² min⁻¹.

S1.4.7 Maximal saturated production rate V_{max}

The maximal saturated production rate for sensor and actuator was set to 10⁴ nM min⁻¹. This ensures that this values is never exceeded regardless of the amounts of the output and controller species respectively.

S2 The universality of antithetic feedback controllers for robust perfect adaptation of noisy biomolecular networks

S2.1 Preliminaries

S2.1.1 The stochastic model for chemical kinetics

The dynamics within cells is often very *noisy* and it has a random component that arises because of the intermittent and stochastic nature of the reactions. This nature is more pronounced in reactions involving biomolecular species with low population-counts or copy-numbers, like DNA transcripts, transcription factors, etc [47]. The randomness in the intracellular dynamics can lead to significant cell-to-cell heterogeneities [48, 49], which has been implicated in important functional roles [50, 51]. We are interested in investigating control mechanisms that can effectively operate despite the randomness in intracellular dynamics, and therefore we need to account for this randomness using stochastic models. We shall employ the most common stochastic model which describes the intracellular dynamics as a *continuous-time Markov chain* (CTMC) that keeps track of the copy-numbers of all the species as they participate in various reactions [34, 52].

We now describe the CTMC model in greater detail. Henceforth let $\mathbb{Z}$ and $\mathbb{N}_0$ denote the set of all integers and all nonnegative integers respectively. Consider a reaction network with N biomolecular species $\mathbf{X}_1, \dots, \mathbf{X}_N$ that interact among each other through reactions in the set $\mathcal{K}_N$. Each reaction $k \in \mathcal{K}_N$ is of the form



where ν_{ki} and ν'_{ki} denote the number of molecules of species $\mathbf{X}_i$ that are consumed and produced by reaction k . The state at time t is simply the vector of species copy-numbers $X(t) = (X_1(t), \dots, X_N(t)) \in \mathbb{N}_0^N$. The rate of firing of the k -th reaction is determined by its *propensity function* $\lambda_k : \mathbb{N}_0^N \rightarrow [0, \infty)$ and this reaction displaces the state by the *stoichiometric vector* $\zeta_k = (\nu'_{k1} - \nu_{k1}, \dots, \nu'_{kN} - \nu_{kN}) \in \mathbb{Z}^N$. When the network state is $x = (x_1, \dots, x_N)$, reaction k fires at rate $\lambda_k(x)$ and the state moves to $(x + \zeta_k)$. Letting $\nu_k = (\nu_{k1}, \dots, \nu_{kN}) \in \mathbb{N}_0^N$, we shall assume from now on that for each $k \in \mathcal{K}_N$

$$\lambda_k(x) > 0 \quad \Longleftrightarrow \quad x \geq \nu_k, \quad (\text{S9})$$

where the vector inequality $x \geq \nu_k$ is in the component-wise sense. This assumption essentially says that if for reaction k the required number of reactant molecules are present then this reaction has a positive rate of firing. Such an assumption is biologically very reasonable (see [53]) and it ensures that the dynamics stays in the non-negative integer orthant $\mathbb{N}_0^N$. Note that this assumption is trivially satisfied for reaction k , if its propensity function has the *mass-action* form [52]

$$\lambda_k(x_1, \dots, x_N) = \gamma_k \prod_{i=1}^N \frac{x_i(x_i - 1) \dots (x_i - \nu_{ki} + 1)}{\nu_{ki}!}, \quad (\text{S10})$$

where $\gamma_k > 0$ is the associated rate constant.

A common way to describe a stochastic process is via its *generator* which specifies the time-evolution of the probability distribution of the process (see [54]). For our CTMC model, the generator of the state dynamics $(X(t))_{t \geq 0}$ is given by

$$\mathbb{A}f(x) = \sum_{k \in \mathcal{K}_N} \lambda_k(x) (f(x + \zeta_k) - f(x)), \quad (\text{S11})$$

where f is a bounded real-valued function on the set $\mathcal{E} \subset \mathbb{N}_0^N$ of accessible states for the CTMC. If p_t is the probability distribution of $X(t)$ then its time-evolution is given by the well-known *Chemical Master Equation* (CME) [52], which can be expressed as

$$\frac{dp_t}{dt} = \mathbb{A}^* p_t, \quad (\text{S12})$$

where $\mathbb{A}^*$ is the *adjoint* of linear operator $\mathbb{A}$ given by

$$\mathbb{A}^* \xi(x) = \sum_{k \in \mathcal{K}_N} (\lambda_k(x - \zeta_k) \xi(x - \zeta_k) - \lambda_k(x) \xi(x)).$$

If we define the inner product $\langle f, \xi \rangle$ between a real-valued function f on $\mathcal{E}$ and a probability distribution ξ on $\mathcal{E}$ as

$$\langle f, \xi \rangle = \sum_{x \in \mathcal{E}} f(x) \xi(x)$$

then the operator $\mathbb{A}$ and its adjoint $\mathbb{A}^*$ satisfy

$$\langle \mathbb{A}f, \xi \rangle = \langle f, \mathbb{A}^* \xi \rangle. \quad (\text{S13})$$

S2.1.2 The notion of stability

The stability of the CTMC $(X(t))_{t \geq 0}$ is usually assessed by checking if the probability distribution p_t of the state $X(t)$ converges to a stationary distribution π as $t \rightarrow \infty$. We work with a strong notion of stability, also called *ergodicity* (see [55]), that demands that this convergence occurs in the *total variation* norm, irrespective of the initial state $X(0)$, i.e.

$$\lim_{t \rightarrow \infty} \sup_{A \subseteq \mathcal{E}} |p_t(A) - \pi(A)| = 0 \quad \text{for any} \quad X(0) \in \mathcal{E}$$

where for any probability distribution ξ on $\mathcal{E}$

$$\xi(A) = \sum_{x \in A} \xi(x).$$

Note that the stationary distribution π is a fixed point for the CME (S12) which satisfies

$$\mathbb{A}^* \pi = \mathbf{0}. \quad (\text{S14})$$

Ergodicity implies that this fixed point exists uniquely and it is globally attracting over the space of probability distributions over $\mathcal{E}$. As the state-space $\mathcal{E}$ is typically infinite, direct verification of ergodicity is nearly impossible. Fortunately an indirect approach exists for this purpose, which involves first checking that the state-space $\mathcal{E}$ is *irreducible* and then constructing a suitable Foster-Lyapunov function over the state-space $\mathcal{E}$ [56]. Computational approaches for identifying the irreducible state-spaces for CTMC models of reaction networks and constructing a Foster-Lyapunov function are provided in [53] and [57] respectively. For a state-space $\mathcal{E}$ to be irreducible, for any two states $x, y \in \mathcal{E}$, there is a positive probability for the CTMC starting at x to reach y in a finite time, i.e.

$$\mathbb{P}(X(t) = y | X(0) = x) > 0 \quad \text{for some} \quad t > 0. \quad (\text{S15})$$

Due to condition (S9), reaction k has a positive probability of firing at state z if and only if $z \geq \nu_k$. Using this fact one can see that for (S15) there must exist a sequence of reactions $k_1, \dots, k_n \in \mathcal{K}_N$, such that if we define $z_0 = x$ and

$$z_j = z_0 + \sum_{l=1}^j \zeta_{k_l} \quad \text{for} \quad j = 1, \dots, n$$

then we have $z_n = y$ and

$$z_{j-1} \geq \nu_{k_j} \quad \text{for each} \quad j = 1, \dots, n.$$

S2.1.3 Perfect adaptation and the Poisson equation

Suppose that the CTMC $(X(t))_{t \geq 0}$ is ergodic with stationary distribution π and for a real-valued function f on state-space $\mathcal{E}$ we have

$$\langle f, \pi \rangle = \sum_{x \in \mathcal{E}} f(x) \pi(x) = \mu$$

for some $\mu \in \mathbb{R}$. Usually the ergodicity of $(X(t))_{t \geq 0}$ would imply that the convergence

$$\lim_{t \rightarrow \infty} \mathbb{E}(f(X(t))) = \mu \quad (\text{S16})$$

holds irrespective of the initial condition $X(0)$. Therefore this input-output system, with input given by the initial state $X(0)$ and output given by the steady-state expectation $\lim_{t \rightarrow \infty} \mathbb{E}(f(X(t)))$, exhibits *Perfect Adaptation* (PA) with respect to constant input disturbances. If the input is perturbed by a

constant, then this perturbation is *rejected* and the system output returns to the pre-perturbation level μ after a transient period.

Under very general conditions, the PA condition (S16) is equivalent to the existence of a function $F : \mathcal{E} \rightarrow \mathbb{R}$ that solves the following Poisson equation (see Chapter 8 in [58])

$$\mathbb{A}F(x) = \mu - f(x) \quad \text{for all } x \in \mathcal{E}. \quad (\text{S17})$$

This solution function F is unique up to addition by a constant. This relationship between PA and the Poisson equation will be useful for us later.

S2.1.4 Sensitivity and Robustness

We now assume that the reaction propensity functions λ_k -s depend on a scalar parameter γ (like a reaction rate constant, cell volume, temperature etc.) in addition to the current state x . Hence the rate of firing of reaction k at state x is $\lambda_k(x, \gamma)$, and we denote the corresponding γ -dependent CTMC as $(X_\gamma(t))_{t \geq 0}$ and the associated generator as $\mathbb{A}_\gamma$ and the stationary distribution as π_γ . We assume that γ can take any value in some feasible *open* set $\Gamma \subset \mathbb{R}$. Typically for mass-action kinetics (S10), γ is the reaction rate constant and $\Gamma = (0, \infty)$ is the set of all positive real numbers.

We now identify the necessary condition for PA (S16) to hold *robustly* with respect to parameter γ , along with the initial state $X(0)$. This brings us to the notion of *Robust Perfect Adaptation* (RPA) which demands that in addition to (S16) the following holds

$$\lim_{t \rightarrow \infty} \frac{\partial}{\partial \gamma} \mathbb{E}(f(X_\gamma(t))) = 0 \quad \text{for any } \gamma \in \Gamma. \quad (\text{S18})$$

According to this condition, constant perturbations in the parameter γ are rejected, and the infinitesimal sensitivity of the output with respect to parameter γ is zero at steady-state. This relationship between RPA and zero steady-state sensitivity has been explored in [59, 60] for deterministic models of reaction networks (see Section S2.2.4 for more details).

Due to ergodicity of the CTMC $(X_\gamma(t))_{t \geq 0}$ it can be expected that (see [61] for more details)

$$\lim_{t \rightarrow \infty} \frac{\partial}{\partial \gamma} \mathbb{E}(f(X_\gamma(t))) = \frac{\partial}{\partial \gamma} \langle f, \pi_\gamma \rangle.$$

We now discuss how the steady-state sensitivity

$$\frac{\partial}{\partial \gamma} \langle f, \pi_\gamma \rangle$$

can be computed, using the solution F_γ of the Poisson equation (S17) for generator $\mathbb{A}_\gamma$. Note that equation (S14) implies that $\mathbb{A}_\gamma^* \pi_\gamma = \mathbf{0}$, and differentiating this relation with respect to γ we obtain

$$-\mathbb{A}_\gamma^* \left(\frac{\partial}{\partial \gamma} \pi_\gamma \right) = \left(\frac{\partial}{\partial \gamma} \mathbb{A}_\gamma^* \right) \pi_\gamma.$$

Taking inner production with function F_γ , and using (S13) along with linearity of the derivative operator we get

$$\begin{aligned} -\left\langle F_\gamma, \mathbb{A}_\gamma^* \left(\frac{\partial}{\partial \gamma} \pi_\gamma \right) \right\rangle &= \left\langle F_\gamma, \left(\frac{\partial}{\partial \gamma} \mathbb{A}_\gamma^* \right) \pi_\gamma \right\rangle \\ -\left\langle \mathbb{A}_\gamma F_\gamma, \left(\frac{\partial}{\partial \gamma} \pi_\gamma \right) \right\rangle &= \left\langle \left(\frac{\partial}{\partial \gamma} \mathbb{A}_\gamma \right) F_\gamma, \pi_\gamma \right\rangle \\ -\left\langle \mu - f, \left(\frac{\partial}{\partial \gamma} \pi_\gamma \right) \right\rangle &= \left\langle \left(\frac{\partial}{\partial \gamma} \mathbb{A}_\gamma \right) F_\gamma, \pi_\gamma \right\rangle \\ \left\langle f, \left(\frac{\partial}{\partial \gamma} \pi_\gamma \right) \right\rangle - \left\langle \mu, \left(\frac{\partial}{\partial \gamma} \pi_\gamma \right) \right\rangle &= \left\langle \left(\frac{\partial}{\partial \gamma} \mathbb{A}_\gamma \right) F_\gamma, \pi_\gamma \right\rangle \\ \frac{\partial}{\partial \gamma} \langle f, \pi_\gamma \rangle &= \left\langle \left(\frac{\partial}{\partial \gamma} \mathbb{A}_\gamma \right) F_\gamma, \pi_\gamma \right\rangle, \end{aligned}$$

where the last equality holds because

$$\left\langle \mu, \left(\frac{\partial}{\partial \gamma} \pi_\gamma \right) \right\rangle = 0$$

as the probability measure π_γ sums up to 1 for any choice of γ . This calculation shows that the steady-state parameter sensitivity can be computed as

$$\begin{aligned} \frac{\partial}{\partial \gamma} \langle f, \pi_\gamma \rangle &= \left\langle \left(\frac{\partial}{\partial \gamma} \mathbb{A}_\gamma \right) F_\gamma, \pi_\gamma \right\rangle \\ &= \sum_{k \in \mathcal{K}_N} \sum_{x \in \mathcal{E}} \frac{\partial \lambda_k}{\partial \gamma}(x, \gamma) (F_\gamma(x + \zeta_k) - F_\gamma(x)) \pi_\gamma(x). \end{aligned} \quad (\text{S19})$$

A formal derivation of this expression for steady-state parameter sensitivity is provided in [62]. Even though we considered γ as a scalar parameter, the generalization of the results in this section to the vector-valued case is quite straightforward.

S2.2 Main Results

S2.2.1 The biomolecular control design problem

We consider the problem of designing an *in vivo* biomolecular controller that can effectively regulate an unknown or poorly characterized intracellular network, even in the presence of randomness in the dynamics and other environmental fluctuations (see Box 1 (Main Text), Fig. a). We shall use the CTMC-based modeling framework (see Section S2.1.1) to describe the dynamics of both the controller network and the controlled network. The controlled network consists of N biomolecular species $\mathbf{X}_1, \dots, \mathbf{X}_N$ that can undergo reactions in the set $\mathcal{K}_N$ of the form (S8). Let $\gamma_N = (\gamma_{N,1}, \gamma_{N,2}, \dots)$ be the set of network parameters that influence the propensity functions λ_k -s for the reactions in $\mathcal{K}_N$.

Henceforth let e_m denote the vector of appropriate dimensions whose m -th component is one and the rest are all zeros. We suppose that the biomolecular controller network comprises C biomolecular species $\mathbf{Z}_1, \dots, \mathbf{Z}_C$ that interact among each other through reactions in the set $\mathcal{K}_C$. Each reaction $k \in \mathcal{K}_C$ is of the form



where α_{ki} and α'_{ki} denote the number of molecules of species $\mathbf{Z}_i$ that are consumed and produced by reaction k . For each controller state $z = (z_1, \dots, z_C)$, the rate of firing of the k -th reaction is $\lambda_k(z, \gamma_C)$ and it displaces the controller state to $(z + \zeta_k)$, where $\zeta_k = (\alpha'_{k1} - \alpha_{k1}, \dots, \alpha'_{kN} - \alpha_{kN})$. Here $\gamma_C = (\gamma_{C,1}, \gamma_{C,2}, \dots)$ is the vector of controller parameters. Apart from reactions in $\mathcal{K}_N \cup \mathcal{K}_C$ we also have interaction reactions $\mathcal{K}_I$ between the controller and the network and an additional reaction that encodes the user-specified set-point μ for the controller. This set-point encoding reaction (denoted by k_μ) is a reaction of the form (S20) with mass-action kinetics (S10),

$$\text{rate constant } \gamma_{k_\mu} = \mu \quad \text{and stoichiometric vector } \zeta_{k_\mu} = \pm e_1, \quad (\text{S21})$$

i.e. this reaction only produces or consumes a single molecule of species $\mathbf{Z}_1$. Examples of such a set-point encoding reaction include the zeroth-order reaction



or the first-order reaction



The set $\mathcal{K}_I$ consists of *exactly one sensing/measurement* reaction (denoted by k_θ), which measures the output species $\mathbf{X}_L$. This sensing reaction k_θ is again of the form (S20) with mass-action kinetics (S10),

$$\text{rate constant } \gamma_{k_\theta} = \theta x_L \quad \text{and stoichiometric vector } \zeta_{k_\theta} = \pm e_i, \quad (\text{S24})$$

for some $i \in \{1, \dots, C\}$. Such a reaction only produces or consumes a single molecule of some controller species $\mathbf{Z}_i$, at a rate proportional to the copy-number x_L of output species $\mathbf{X}_L$. In particular, this sensing reaction could be *catalytic production* of $\mathbf{Z}_i$



or *catalytic degradation* of $\mathbf{Z}_i$



The set of interaction reactions $\mathcal{K}_I$ also consists of actuation reactions which can be arbitrary (in both their number and their type), as long as they only affect the state $x = (x_1, \dots, x_N)$ of the controlled network and do not affect the state $z = (z_1, \dots, z_C)$ of the controller network. For definiteness we assume from now on that there is only one actuation reaction of the form



where $\mathbf{Z}_j$ is one of the controller species. We assume mass-action kinetics for reaction (S27) but the kinetics for this reaction can be completely general.

Note that the mass-action form (S10), allows us to write the propensity functions for the set-point encoding reaction k_μ and the sensing/measurement reaction k_θ as $\mu \lambda_{k_\mu}(z)$ and $\theta x_L \lambda_{k_\theta}(z)$ respectively. The closed-loop reaction dynamics $(X_\gamma(t), Z_\gamma(t))_{t \geq 0}$ can be described as a CTMC whose generator is given by

$$\begin{aligned} \mathbb{A}_\gamma f(x, z) = & \sum_{k \in \mathcal{K}_N} \lambda_k(x, \gamma_N) (f(x + \zeta_k, z) - f(x, z)) \\ & + \sum_{k \in \mathcal{K}_C} \lambda_k(z, \gamma_C) (f(x, z + \zeta_k) - f(x, z)) \\ & + \kappa z_j (f(x + e_1, z) - f(x, z)) + \mu \lambda_{k_\mu}(z) (f(x, z \pm e_1) - f(x, z)) \\ & + \theta x_L \lambda_{k_\theta}(z) (f(x, z \pm e_i) - f(x, z)) \end{aligned} \quad (\text{S28})$$

where $\gamma = (\gamma_N, \gamma_C)$ is the vector of all parameters and f is any bounded real-valued function on the state-space $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$ of the dynamics with $\mathcal{E}_1 \subset \mathbb{N}_0^N$ and $\mathcal{E}_2 \subset \mathbb{N}_0^C$.

Objective: Note that the output being sensed by the controller is θx_L . Our goal here is to design a biomolecular controller that can *robustly* drive this output to the steady-state value $\mu > 0$. In other words the following *perfect adaptation* (PA) property (see Section S2.1.3) must hold for any initial state $(x_0, z_0) \in \mathcal{E}$

$$\lim_{t \rightarrow \infty} \mathbb{E}(\theta X_{\gamma, L}(t)) = \mathbb{E}_{\pi_\gamma}(\theta X_{\gamma, L}) = \mu, \quad (\text{S29})$$

where $\mathbb{E}$ denotes the expectation operator and $\mathbb{E}_{\pi_\gamma}$ denotes expectation under the stationary distribution π_γ . Moreover we desire the controller to achieve *robust perfect adaptation* (RPA) (see Section S2.2.4) in the sense that the PA condition (S29) should be satisfied for any choice of parameters $\gamma = (\gamma_N, \gamma_C)$ in some feasible set $\Gamma = \Gamma_N \times \Gamma_C$. This ensures constant disturbances in these parameters will be *rejected*. Finally we would like the proposed biomolecular controller to achieve RPA for any unknown or poorly characterized intracellular network of interest, as long as certain minimal assumptions are satisfied which shall be elaborated later.

The requirements we are placing on the proposed controller are inspired by synthetic biology applications where the structure, the kinetic details (for e.g. mass-action, Hill-type or Michaelis-Menten kinetics) and parameters for the controlled network are often unknown. Therefore it is natural to desire robustness with respect to these attributes. Additionally we also require robustness with respect to all controller parameters except μ and θ which are associated with the set-point encoding reaction (S21) and the sensing/measurement reaction (S24). Due to this strong robustness propensity the topology of the controller that achieves RPA is severely restricted. Our aim is to show that any RPA-achieving controller motif needs to be of the antithetic type in a certain generalized sense. Note that even though we are considering RPA at the population-level by defining it in terms of the expectation or the population-average of several identical single-cells, under the assumption of ergodicity for the single-cell dynamics (see [57]), RPA will also hold for the long-term average output of a single cell given by $\lim_{T \rightarrow \infty} T^{-1}(\int_0^T X_{\gamma,L}(t)dt)$.

S2.2.2 Parametric characterization of all RPA-achieving controller motifs

The aim of this section is to derive the necessary and sufficient conditions that an RPA-achieving controller must satisfy to regulate any intracellular network. Before we proceed with our analysis we list all the assumptions we will be making in our analysis.

Assumption S2.1. (A) *There is a set-point encoding reaction k_μ satisfying (S21), sensing/measurement reaction k_θ satisfying (S24) and only one actuation reaction of the form (S27). All these reactions have mass-action kinetics and the constants μ, κ and θ are strictly positive. All other network parameters γ_N and controller parameters γ_C belong to feasible sets Γ_N and Γ_C respectively. Furthermore, $\Gamma = \Gamma_N \times \Gamma_C$ is an open subset of some Euclidean space $\mathbb{R}^n$. The propensity function $\lambda_k(\cdot, \gamma_N)$ for each network reaction satisfies (S9) for every $\gamma_N \in \Gamma_N$. Similarly the propensity function $\lambda_k(\cdot, \gamma_C)$ for each controller reaction satisfies (S9) for every $\gamma_C \in \Gamma_C$.*

(B) *The state-space $\mathcal{E}_N \subset \mathbb{N}_0^N$ is irreducible (see Section S2.1.2) for the controlled network with only the reactions in the set $\mathcal{K}_N$. Similarly the state-space $\mathbb{N}_0^C$ is irreducible for the controller network with reactions in the set $\mathcal{K}_C$, the set-point encoding reaction and the sensing/measurement reaction with a fixed output copy-number x_L . Furthermore the closed-loop CTMC dynamics described by generator $\mathbb{A}_\gamma$ is ergodic with irreducible state-space $\mathcal{E} = \mathcal{E}_N \times \mathbb{N}_0^C$ and a unique stationary distribution π_γ (see Section S2.1.2). Finally, the first two moments of the controlled network state $X_\gamma(t)$ converge to their steady-state values (determined by the stationary distribution π_γ) as $t \rightarrow \infty$.*

(C) *Each controller reaction $k \in \mathcal{K}_C$ of the form (S20) is elementary in the sense that the total number of reactant molecules is strictly less than 3 (i.e. $\sum_i \alpha_{ki} < 3$). Moreover each controller reaction's propensity function $\lambda_k(z, \gamma_C)$ depends non-trivially on at least one independent parameter in the controller parameter set γ_C .*

Recall the relationship between perfect adaptation (PA) (S29) and the Poisson equation from Section S2.1.3. Define a real-valued function f over the state-space $\mathcal{E}$ by

$$f(x, z) = \theta x_L.$$

Then PA is equivalent to the existence of another real-valued function $F_\gamma : \mathcal{E} \rightarrow \mathbb{R}$ which solves the Poisson equation

$$\mathbb{A}_\gamma F_\gamma(x, z) = \mu - f(x, z) \quad \text{for all} \quad (x, z) \in \mathcal{E}. \quad (\text{S30})$$

The following lemma proves that this function F_γ needs to be invariant under translations by the stoichiometric vectors corresponding to reactions in the set $\mathcal{K}_N \cup \mathcal{K}_C$.

Lemma S2.2. *Pick a RPA-achieving controller and suppose that Assumptions S2.1 hold. Fix a feasible parameter set $\gamma \in \Gamma$ and let F_γ be a solution to the Poisson equation (S30). Let $\nu_k = (\nu_{k1}, \dots, \nu_{kN})$ and $\alpha_k = (\alpha_{k1}, \dots, \alpha_{kC})$ for each reaction k be of the form (S8) and (S20) respectively. Then for every $(x, z) \in \mathcal{E} := \mathcal{E}_N \times \mathbb{N}_0^C$ this function needs to satisfy*

$$F_\gamma(x + \zeta_k, z) = F_\gamma(x, z) \quad \text{for each } k \in \mathcal{K}_N \quad \text{such that } x \geq \nu_k \quad (\text{S31})$$

$$\text{and} \quad F_\gamma(x, z + \zeta_k) = F_\gamma(x, z) \quad \text{for each } k \in \mathcal{K}_C \quad \text{such that } z \geq \alpha_k. \quad (\text{S32})$$

Furthermore, this function F_γ does not depend on the network state x , i.e. the map $x \mapsto F_\gamma(x, z)$ is a constant for any fixed z .

Proof. The proposed controller needs to achieve RPA for any parametrization of the network of interest. Hence to derive the necessary conditions for F_γ we can assume that for each reaction $k \in \mathcal{K}_N$, the propensity function depends on a single independent parameter γ_k . Using (S19) and writing each propensity function as

$$\lambda_k(x, \gamma_N) = \lambda_k(x, \gamma_k)$$

we can express the zero-sensitivity condition (S18) as

$$\frac{\partial}{\partial \gamma} \langle f, \pi_\gamma \rangle = \sum_{(x, z) \in \mathcal{E}} \frac{\partial \lambda_k}{\partial \gamma_k}(x, \gamma_k) (F_\gamma(x + \zeta_k, z) - F_\gamma(x, z)) \pi_\gamma(x, z) = 0.$$

As the parametrization of the controlled network is completely arbitrary, we can replace γ_k with another parameter γ'_k close to it and write the propensity function for reaction k as

$$\lambda_k(x, \gamma_N) = \lambda_k(x, \gamma'_k) \exp((\gamma'_k - \gamma_k)\rho(x))$$

where ρ is any real-valued function on $\mathcal{E}_N$. Noting that

$$\lambda_k(x, \gamma_N) \Big|_{\{\gamma'_k = \gamma_k\}} = \lambda_k(x, \gamma_k)$$

and

$$\frac{\partial \lambda_k}{\partial \gamma'_k}(x, \gamma_N) \Big|_{\{\gamma'_k = \gamma_k\}} = \frac{\partial \lambda_k}{\partial \gamma_k}(x, \gamma_k) + \lambda_k(x, \gamma_k)\rho(x),$$

we can write the zero-sensitivity condition (S18) at $\gamma'_k = \gamma_k$ as

$$\begin{aligned} \frac{\partial}{\partial \gamma} \langle f, \pi_\gamma \rangle &= \sum_{(x, z) \in \mathcal{E}} \left[\frac{\partial \lambda_k}{\partial \gamma_k}(x, \gamma_k) + \lambda_k(x, \gamma_k)\rho(x) \right] (F_\gamma(x + \zeta_k, z) - F_\gamma(x, z)) \pi_\gamma(x, z) \\ &= 0. \end{aligned} \quad (\text{S33})$$

Here we have used the fact that the set of feasible parameters Γ_N is open and hence parameter γ'_k is admissible if its value is sufficiently close to γ_k . Secondly the stationary distribution π_γ and the Poisson equation solution F_γ will not change due to this parameter switch, as the kinetics (i.e. the generator $\mathbb{A}_\gamma$) remains the same when we set $\gamma'_k = \gamma_k$. Observe that as the state-space $\mathcal{E}$ is irreducible, $\pi_\gamma(x, z) > 0$ for each $(x, z) \in \mathcal{E}$ and if $x \geq \nu_k$ then $\lambda_k(x, \gamma_k) > 0$ due to part (A) of Assumption S2.1. Since (S33) holds for any choice of function ρ we must have that condition (S31) is satisfied. The proof of (S32) is similar except that zero-sensitivity condition needs to be applied with respect to controller parameters in the set γ_C . Here part (C) of Assumption S2.1 is needed to ensure that for each controller reaction $k \in \mathcal{K}_C$ we have an independent controller parameter γ_k that affects its propensity function.

Due to part (B) of Assumption S2.1, the state-space $\mathcal{E}_N$ is irreducible for the controlled network with only the reactions in the set $\mathcal{K}_N$. For any $x_1, x_2 \in \mathcal{E}_N$ there exists a sequence of reactions $k_1, \dots, k_n \in$

$\mathcal{K}_N$, such that there is a positive probability of going from state x_1 to state x_2 via these reactions. Due to part (A) of Assumption S2.1, for this to hold we must have

$$y_j := x_1 + \sum_{l=1}^j \zeta_{k_l} \geq \nu_{k_{j+1}} \quad \text{for each } j = 0, 1, \dots, n-1$$

and

$$y_n := x_1 + \sum_{l=1}^n \zeta_{k_l} = x_2.$$

Setting $y_0 = x_1$ and using (S31) we note that

$$\begin{aligned} F_\gamma(x_2, z) - F_\gamma(x_1, z) &= \sum_{j=1}^n (F_\gamma(y_j, z) - F_\gamma(y_{j-1}, z)) \\ &= \sum_{j=1}^n (F_\gamma(y_{j-1} + \zeta_{k_j}, z) - F_\gamma(y_{j-1}, z)) \\ &= 0. \end{aligned}$$

This shows that the map $x \mapsto F_\gamma(x, z)$ is constant for any fixed z . This completes the proof of this lemma. $\square$

We now use this lemma to prove a key result, Proposition S2.3, which shows that the antithetic controller described in [63] is the *minimal* RPA-achieving controller in a certain sense. This antithetic controller has only two species (i.e. $C = 2$) and only one reaction in $\mathcal{K}_C$, which is the *sequestration* or *annihilation* reaction of the form



The set-point encoding reaction has the form (S22) and the sensing/measurement reaction has the form (S25) with $\mathbf{Z}_i = \mathbf{Z}_2$. However the actuation mechanism could be more general than (S27) (see Supplementary Material of [63]).

Proposition S2.3. *Select any RPA-achieving controller and suppose that Assumptions S2.1 hold. Then this controller must necessarily have at least two species (i.e. $C \geq 2$), the set-point encoding reaction must have the form (S22) and the sensing reaction must be of the form (S25) with $\mathbf{Z}_i \neq \mathbf{Z}_1$. Moreover if the number of species is 2, then $\mathbf{Z}_i = \mathbf{Z}_2$ and the controller must have the sequestration reaction (S34).*

Remark S2.4. *This proposition shows that the antithetic controller of [63] has the least number of reactions among all the two-species controllers that achieve RPA. Since all such controllers must have at least two species, we can regard this antithetic controller as the minimal RPA-achieving controller.*

Proof. As Lemma S2.2 shows, the map $x \mapsto F_\gamma(x, z)$ is constant for any fixed z . Therefore we write the function $F_\gamma(x, z)$ as $F_\gamma(z)$, dropping the dependence on x . Applying the generator (S28) for closed-loop CTMC dynamics on function F_γ we obtain

$$\begin{aligned} \mathbb{A}_\gamma F_\gamma(z) &= \mu \lambda_{k_\mu}(z) (F_\gamma(z \pm e_1) - F_\gamma(z)) + \theta x_L \lambda_{k_\theta}(z) (F_\gamma(z \pm e_i) - F_\gamma(z)) \\ &\quad + \sum_{k \in \mathcal{K}_C} \lambda_k(z, \gamma_C) (F_\gamma(z + \zeta_k) - F_\gamma(z)). \end{aligned} \quad (\text{S35})$$

Observe that the last term on the right does not contain the variable x_L and so for (S30) to hold for all possible parameter values γ_C we must have that

$$\lambda_{k_\mu}(z) (F_\gamma(z \pm e_1) - F_\gamma(z)) = 1 \quad (\text{S36})$$

and

$$\lambda_{k_\theta}(z)(F_\gamma(z \pm e_i) - F_\gamma(z)) = -1. \quad (\text{S37})$$

Recall that the set-point encoding reaction k_μ has the following form

$$\sum_{j=1}^C \alpha_{kj} \mathbf{Z}_j \longrightarrow (\alpha_{k1} \pm 1) \mathbf{Z}_1 + \sum_{j=1}^C \alpha_{kj} \mathbf{Z}_j$$

and it obeys mass-action kinetics (S10). Note that if any of the α_{kj} -s is strictly positive then (S36) will be violated at $z = (0, \dots, 0)$ because $\lambda_{k_\mu}(0, \dots, 0) = 0$. Hence each $\alpha_{kj} = 0$ and this reaction must necessarily be of the form (S22) in order to satisfy condition (S9) and prevent the copy-number of $\mathbf{Z}_1$ from becoming negative. Similar reasoning shows that the sensing reaction must be of the form (S25). Clearly $i \neq 1$ and so the sensing species $\mathbf{Z}_i$ must be different from the set-point encoding species $\mathbf{Z}_1$ and therefore the controller must have *at least* two species. Without loss of generality we can assume that $i = 2$ and then to satisfy (S36) and (S37) for all $z \in \mathbb{N}_0^C$, F_γ must have the form

$$F_\gamma(z) = z_1 - z_2 + G_\gamma(\bar{z}), \quad (\text{S38})$$

where $\bar{z} = (z_3, \dots, z_C)$ and G_γ is some function which does not depend on z_1 and z_2 . If there are only two controller species (i.e. $C = 2$) then the function G_γ must be a constant and due to (S32) we must have that

$$\zeta_{k1} = \zeta_{k2} \quad \text{for all} \quad k \in \mathcal{K}_C. \quad (\text{S39})$$

For closed-loop stability (see part (B) of Assumption S2.1) there must exist at least one reaction $k \in \mathcal{K}_C$ such that $\zeta_{k1} < 0$ (i.e. reaction k causes net loss in the number of molecules of species $\mathbf{Z}_1$). As this reaction needs to be elementary (part (C) of Assumption S2.1) the only possible choice of such a reaction is the *sequestration* reaction (S34). This completes the proof of this proposition. $\square$

For a controller with C species and K reactions in the set $\mathcal{K}_C$, define the $C \times K$ stoichiometric matrix as

$$S = \text{Col}(\zeta_1, \dots, \zeta_K), \quad (\text{S40})$$

where for each $k \in \mathcal{K}_C$, $\zeta_k = (\zeta_{k1}, \dots, \zeta_{kC})$ is the corresponding stoichiometry vector. We now prove our main result which characterizes all RPA-achieving controllers through a simple linear-algebraic condition, specified in terms of matrix S .

Theorem S2.5. *Consider a controller with C species $\mathbf{Z}_1, \dots, \mathbf{Z}_C$ and K reactions in the set $\mathcal{K}_C$, in addition to the actuation reactions, the set-point encoding reaction (S22) and the sensing/measurement reaction (S25). Suppose that Assumptions S2.1 hold, there are at least two control species (i.e. $C \geq 2$) and the sensing species $\mathbf{Z}_i$ is $\mathbf{Z}_2$. Then this controller achieves RPA if and only if there exists a $C \times 1$ vector $q = (q_1, \dots, q_C)$ such that*

$$q_1 = 1, \quad q_2 = -1 \quad \text{and} \quad q^T S = \mathbf{0} \quad (\text{S41})$$

where S is the $C \times K$ stoichiometry matrix (S40) for the controller.

Remark S2.6. *It will be evident from the proof of this theorem that if such a vector q exists then all its components will necessarily be integers.*

Proof. We first prove the “if” part of the theorem which is simpler. Suppose we have a $C \times 1$ vector $q = (q_1, \dots, q_C)$ satisfying (S41). Define a function $F : \mathbb{N}_0^C \rightarrow \mathbb{R}$ by

$$F(z_1, \dots, z_C) = \sum_{i=1}^C q_i z_i = z_1 - z_2 + \sum_{i=3}^C q_i z_i. \quad (\text{S42})$$

With a slight abuse of notation we may also regard F as a function on $\mathcal{E}_N \times \mathbb{N}_0^C$, given by $F(x, z) = F(z)$. It is easy to see that F satisfies (S31) and using condition (S41) we can also see that F satisfies (S32). Replacing F_γ by F in (S35), we can verify that F solves the Poisson equation (S30) for any choice of parameters $\gamma = (\gamma_N, \gamma_C)$, and hence the PA condition (S29) is achieved. Moreover from the proof of Lemma S2.2 it is immediate that the zero-sensitivity condition (S18) will hold with respect to all these parameters and therefore the controller achieves RPA.

We now prove the “only if” part of the theorem. Suppose that we have a controller that achieves RPA and we fix a set of parameters $\gamma = (\gamma_N, \gamma_C)$. Then there exists a solution F_γ to the Poisson equation (S30), which is unique up to addition by a constant, and due to Lemma S2.2, we can consider $F_\gamma(x, z)$ as only a function of z (i.e. $F_\gamma(x, z) = F_\gamma(z)$). This function satisfies condition (S32) which can be written as

$$F_\gamma(z + \zeta_k) = F_\gamma(z) \quad \text{for each } z \in \mathbb{N}_0^C \quad \text{and } k \in \mathcal{K}_C, \quad (\text{S43})$$

such that $z \geq \alpha_k$, where $\alpha_k = (\alpha_{k1}, \dots, \alpha_{kC})$ for reaction of the form (S20).

Without loss of generality we can assume that $F_\gamma(\mathbf{0}) = 0$ where $\mathbf{0}$ is the C -dimensional vector of zeros. Let e_m be the C -dimensional vector whose m -th component is one and the rest are all zeros. Note that apart from reactions in $\mathcal{K}_C$ the state of the control species can change due to the set-point encoding reaction (S22) (reaction $K + 1$) with stoichiometry vector e_1 and the sensing reaction (S25) (reaction $K + 2$) with stoichiometry vector e_2 . Let $\mathcal{K}'_C = \mathcal{K}_C \cup \{K + 1, K + 2\}$ be the augmented set of reactions that includes these two reactions and the reactions in $\mathcal{K}_C$. As the state-space $\mathbb{N}_0^C$ is irreducible for control dynamics, for each species $\mathbf{Z}_i$ there exists a sequence of reactions $k_1^{(i)}, \dots, k_n^{(i)}$ in $\mathcal{K}'_C$ such that there is a positive probability of going from state $\mathbf{0}$ to state e_i via these reactions. Due to part (A) of Assumption S2.1, for this to hold we must have

$$z_j := \sum_{l=1}^j \zeta_{k_l^{(i)}} \geq \alpha_{k_{j+1}^{(i)}} \quad \text{for each } j = 0, 1, \dots, n-1$$

and

$$z_n := \sum_{l=1}^n \zeta_{k_l^{(i)}} = e_i.$$

Moreover for any $z \in \mathbb{N}_0^C$ this same sequence of reactions will also allow the dynamics to go from state z to state $(z + e_i)$ with a positive probability (see [53]). Setting $z_0 = \mathbf{0}$ we note that

$$\begin{aligned} F_\gamma(z + e_i) - F_\gamma(z) &= \sum_{j=1}^n (F_\gamma(z + z_j) - F_\gamma(z + z_{j-1})) \\ &= \sum_{j=1}^n \left(F_\gamma(z + z_{j-1} + \zeta_{k_j^{(i)}}) - F_\gamma(z + z_{j-1}) \right). \end{aligned}$$

Observe that because of condition (S43) and the form (S38) of the function F_γ we obtain

$$F_\gamma(z + z_{j-1} + \zeta_{k_j^{(i)}}) - F_\gamma(z + z_{j-1}) = \begin{cases} 0 & \text{if } k_j^{(i)} \in \mathcal{K}_C \\ 1 & \text{if } k_j^{(i)} = K + 1 \\ -1 & \text{if } k_j^{(i)} = K + 2. \end{cases}$$

Therefore if n_i^+ and n_i^- denote the number of set-point encoding reactions ($K + 1$) and sensing reactions ($K + 2$) in the sequence $k_1^{(i)}, \dots, k_n^{(i)}$, then we have

$$F_\gamma(z + e_i) - F_\gamma(z) = q_i := n_i^+ - n_i^- \quad (\text{S44})$$

for any $z \in \mathbb{N}_0^C$. Following this procedure we obtain an integer q_i for each controller species and this is sufficient to show that F_γ must be of the form (S42). Note that since the solution F_γ of the Poisson equation (S30) is unique (up to addition by a constant), the quantity q_i will be unique for each species even though the sequence of reaction $k_1^{(i)}, \dots, k_n^{(i)}$ used to go from state $\mathbf{0}$ to state e_i may not be unique. Clearly $q_1 = 1$ as only one set-point encoding reaction is needed to produce a molecule of species $\mathbf{Z}_1$ and similarly $q_2 = -1$ as only one sensing reaction is needed to produce a molecule of species $\mathbf{Z}_2$. Letting $q = (q_1, \dots, q_C)$ and writing function F_γ as

$$F_\gamma(z) = \sum_{i=1}^C q_i z_i = q^T z$$

we see that since (S43) holds, q must be orthogonal to stoichiometric vector ζ_k for each reaction $k \in \mathcal{K}_C$. Therefore we must have

$$q^T S = \mathbf{0}$$

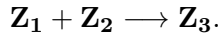
and this finishes the proof of this theorem. $\square$

S2.2.3 The generalized antithetic motif as the universal RPA-achieving motif

In the preceding section we proved that the antithetic controller proposed in [63] is the *minimal* RPA-achieving controller (see Remark S2.4). In this section we strengthen this result by showing that every RPA-achieving controller can be viewed as a *generalized* antithetic controller which embeds the minimal controller. Let $q = (q_1, \dots, q_C)$ be the vector in Theorem S2.5 that characterizes a RPA-achieving controller. We can view q as the vector of *charges* which assigns the charge q_i to each controller species $\mathbf{Z}_i$ based on the number of set-point encoding reactions (S22) and sensing reactions (S25), required to produce exactly one molecule of this species from nothing (see (S44)). Based on vector q we partition the set of all species into a set of *positively charged* species $\mathcal{C}_+$, a set of *negatively charged* species $\mathcal{C}_-$ and a set of *neutral* species $\mathcal{C}_0$, given by

$$\mathcal{C}_+ = \{\mathbf{Z}_i : q_i > 0\}, \quad \mathcal{C}_- = \{\mathbf{Z}_i : q_i < 0\} \quad \text{and} \quad \mathcal{C}_0 = \{\mathbf{Z}_i : q_i = 0\}.$$

The set $\mathcal{C}_+$ consists of all those controller species whose production of exactly one molecule requires *strictly* more set-point encoding reactions (S22) than sensing reactions (S25). Similarly $\mathcal{C}_-$ includes all those controller species for which the opposite is true and $\mathcal{C}_0$ consists of all those controller species whose production of exactly one molecule requires *equal* number of set-point encoding and sensing reactions. Observe that the sets $\mathcal{C}_\pm$ cannot be empty, as species $\mathbf{Z}_1$ is in $\mathcal{C}_+$ and species $\mathbf{Z}_2$ is in $\mathcal{C}_-$. However the set $\mathcal{C}_0$ can be empty and in what follows, we assume that this set always contains the *null* species $\emptyset$. This null species does not require any reactions to get produced and so it trivially satisfies the requirement to belong to set $\mathcal{C}_0$. Another example of species that would lie in $\mathcal{C}_0$ would be a *dimeric* complex (say $\mathbf{Z}_3$) between $\mathbf{Z}_1$ and $\mathbf{Z}_2$ which would be produced via the reaction



We shall later explore the stoichiometric conditions for species to belong to $\mathcal{C}_+$, $\mathcal{C}_-$ and $\mathcal{C}_0$ in greater detail.

We now prove that any RPA achieving controller must be a *generalized* antithetic controller (see Box 1 (Main Text), Fig. b) in the sense that it must necessarily have a reaction of the form



where $\mathbf{Z}_i \in \mathcal{C}_+$, $\mathbf{Z}_j \in \mathcal{C}_-$ and $\mathcal{C}_0^*$ denotes any combination of species that belong to the set $\mathcal{C}_0$. Such a reaction can be seen as the generalized sequestration reaction (S34).

Proposition S2.7. Consider a RPA-achieving controller with C species $\mathbf{Z}_1, \dots, \mathbf{Z}_C$ and suppose that Assumptions S2.1 hold. Then it is necessary that there exists at least one reaction of the form (S45). Moreover reactions of this form are the only possible reactions where the products are species in $\mathcal{C}_0$ and one of the reactant species is in $\mathcal{C}_+ \cup \mathcal{C}_-$.

Remark S2.8. Since the null species $\emptyset$ is in $\mathcal{C}_0$, this proposition shows that for a RPA-achieving controller, any species in $\mathcal{C}_+ \cup \mathcal{C}_-$ cannot degrade spontaneously through a reaction of the form $\mathbf{Z} \rightarrow \emptyset$. Therefore if we have cell dilution where all the species have a first-order degradation rate $\delta > 0$, then there cannot exist a controller that perfectly achieves RPA.

Proof. Let I_0 be the set of addresses of species in $\mathcal{C}_0$ respectively, i.e.

$$I_0 = \{i = 1, \dots, C : \mathbf{Z}_i \in \mathcal{C}_0\}.$$

Define $\mathcal{S}_0 \subset \mathbb{N}_0^C$ to be the set given by

$$\mathcal{S}_0 = \{z = (z_1, \dots, z_C) \in \mathbb{N}_0^C : z_i = 0 \text{ for each } i \notin I_0\}.$$

Hence if the controller state is in $\mathcal{S}_0$ then all the species that are not in $\mathcal{C}_0$ have zero copy-numbers. Pick any species $\mathbf{Z}_m \in \mathcal{C}_+ \cup \mathcal{C}_-$. Let $\mathcal{K}'_C$ be the augmented set of reactions defined in the proof of Theorem S2.5. As the state-space $\mathbb{N}_0^C$ is irreducible for control dynamics, there exists a sequence of reactions $k_1, \dots, k_n$ in $\mathcal{K}'_C$ such that there is a positive probability of going from state e_m to state $\mathbf{0}$ via these reactions. Due to part (A) of Assumption S2.1, for this to hold we must have

$$z_j := e_m + \sum_{l=1}^j \zeta_{k_l} \geq \alpha_{k_{j+1}} \quad \text{for each } j = 0, 1, \dots, n-1$$

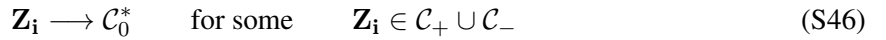
and

$$z_n := e_m + \sum_{l=1}^n \zeta_{k_l} = \mathbf{0}.$$

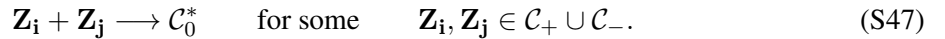
Note that $z_n \in \mathcal{S}_0$ and $z_0 = e_m \notin \mathcal{S}_0$. Let k_* be the first reaction in this sequence which originates from a state outside $\mathcal{S}_0$ and leads to a state within $\mathcal{S}_0$, i.e. $k_* = k_l$ where

$$l = \min\{j = 1, \dots, n : z_{j-1} \notin \mathcal{S}_0 \text{ and } z_j \in \mathcal{S}_0\}.$$

Note that reaction k_* needs to eliminate some copy-numbers of species in $\mathcal{C}_+ \cup \mathcal{C}_-$ and all its products must be species in $\mathcal{C}_0$. Hence this reaction k_* cannot be the set-point encoding (S22) or the sensing reaction (S25). So $k_* \in \mathcal{K}_C$ and since only elementary reactions are allowed (Assumption S2.1, part (C)), reaction k_* must necessarily be of the form



or



From condition (S41) it is easy to see that if reaction k_* has the form (S46), then we must have that $q_i = 0$ which is not possible because $\mathbf{Z}_i \in \mathcal{C}_+ \cup \mathcal{C}_-$. Therefore reaction k_* must have the form (S47) and in this case condition (S41) implies that $q_i + q_j = 0$ or

$$q_i = -q_j.$$

Clearly this can only happen if one of the species $\mathbf{Z}_i$ and $\mathbf{Z}_j$ belongs to $\mathcal{C}_+$ and the other belongs to $\mathcal{C}_-$. This proves the existence of the generalized sequestration reaction (S45) and it also shows that the two species that react in such a reaction should have weights with equal magnitudes but with opposite signs.

This analysis shows that reactions of the form (S46) cannot happen and for reactions of the form (S47) to happen, one of the reactant species must be in $\mathcal{C}_+$ and the other is in $\mathcal{C}_-$. This completes the proof of this proposition. $\square$

Owing to Proposition S2.7 and Theorem S2.5, we can now define the *generalized antithetic controller* as any C -species RPA achieving controller for which there exists a charge-vector $q = (q_1, \dots, q_C)$ satisfying (S41). We now investigate the structure of the generalized antithetic controller in greater detail. Henceforth we shall regard $\mathbf{Z}_0$ as the null species $\emptyset$ with charge $q_0 = 0$ and we adopt the rule that

$$\mathbf{Z}_0 + \mathbf{Z}_i \equiv \mathbf{Z}_i$$

for any species $\mathbf{Z}_i$. In addition to the controller reactions (S20) being elementary (see part (C) of Assumption S2.1) we shall also assume from now on that the number of reactant and product molecules in each reaction $k \in \mathcal{K}_C$ cannot exceed 3, i.e.

$$\sum_{i=1}^C (\alpha_{ki} + \alpha'_{ki}) \leq 3$$

and if this inequality is strict then $\sum_{i=1}^C \alpha_{ki} > 0$. This implies that only three possible types of reactions are allowed:

1. **Conversion reactions:** These reactions are of the form

$$\mathbf{Z}_i \longrightarrow \mathbf{Z}_j \quad \text{for} \quad i, j \in \{0, 1, \dots, C\}. \quad (\text{S48})$$

2. **Splitting reactions:** These reactions are of the form

$$\mathbf{Z}_m \longrightarrow \mathbf{Z}_i + \mathbf{Z}_j \quad \text{for} \quad i, j, m \in \{0, 1, \dots, C\} \quad \text{with} \quad i, j \neq 0. \quad (\text{S49})$$

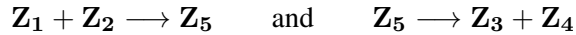
3. **Complex-formation reactions:** These reactions are of the form

$$\mathbf{Z}_i + \mathbf{Z}_j \longrightarrow \mathbf{Z}_m \quad \text{for} \quad i, j, m \in \{0, 1, \dots, C\} \quad \text{with} \quad i, j \neq 0. \quad (\text{S50})$$

Note that as $\mathbf{Z}_i$ can be $\emptyset$, the set of conversion reactions include constitutive production $\emptyset \longrightarrow \mathbf{Z}_j$ and spontaneous degradation $\mathbf{Z}_j \longrightarrow \emptyset$. Similarly, as $\mathbf{Z}_m$ can be $\emptyset$, the set of complex-formation reactions includes the sequestration reaction (S34). Furthermore, letting $m = i$ in (S49), we see that the set of splitting reactions include catalytic reactions of the form $\mathbf{Z}_i \longrightarrow \mathbf{Z}_i + \mathbf{Z}_j$ which are biologically very important (for e.g. to model mRNA translating into protein or transcription of mRNA). Our assumption of allowing only the three types of reactions ((S48), (S49) and (S50)) is not very restrictive as more general reaction forms can usually be decomposed into such reactions, by introducing additional species. For example a reaction of the form



can be viewed as a combination of two reactions



after allowing for an additional species $\mathbf{Z}_5$.

Given a RPA-achieving C -species generalized antithetic controller with $q = (q_1, \dots, q_C)$ satisfying (S41) we can naturally partition the set of control species $\mathcal{C} = \{\mathbf{Z}_0, \dots, \mathbf{Z}_C\}$ according to their charges. For each $n \in \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ define the *charge-class* $\mathcal{C}_n$ as

$$\mathcal{C}_n = \{\mathbf{Z}_j \in \mathcal{C} : q_j = n\}.$$

Note that the definition of $\mathcal{C}_0$ is consistent with the definition stated earlier, and

$$\mathcal{C}_+ = \bigcup_{n=1}^{\infty} \mathcal{C}_n \quad \text{and} \quad \mathcal{C}_- = \bigcup_{n=1}^{\infty} \mathcal{C}_{-n}.$$

From (S41) it is easy to see that the conversion reaction (S48) is *charge-preserving*, i.e. the presence of reaction $\mathbf{Z}_i \longrightarrow \mathbf{Z}_j$ implies that $q_i = q_j$ and hence species $\mathbf{Z}_i$ and $\mathbf{Z}_j$ must belong to the same charge-class $\mathcal{C}_n$. On the other hand the complex formation reaction (S50) is *charge-additive* which means that the reaction $\mathbf{Z}_i + \mathbf{Z}_j \longrightarrow \mathbf{Z}_m$ implies that if $q_i = n_1$ and $q_j = n_2$, then $q_m = (n_1 + n_2)$. A similar additive relationship holds in the case of a splitting reaction (S49) of the form $\mathbf{Z}_m \longrightarrow \mathbf{Z}_i + \mathbf{Z}_j$. Observe that since n_1 and n_2 can be of opposite signs, this additive relationship can also be *annihilative*, in the sense that if $n_1 > 0$ and $n_2 < 0$, then positive and negative charges cancel each other, in appropriate amounts, to either produce (reaction (S50)) or consume (reaction (S49)) species $\mathbf{Z}_m$ with charge $(n_1 + n_2) = (n_1 - |n_2|)$. As Proposition S2.7 indicates each generalized antithetic controller must have at least one such annihilation reaction of the form (S50) with $n_1 = -n_2$.

This discussion suggests that we can express reactions (S48), (S49) and (S50), in terms of charge-classes as below:

Type	Reaction form	Property
Conversion	$\mathcal{C}_n \longrightarrow \mathcal{C}_n$	charge-preserving
Splitting	$\mathcal{C}_{n_1+n_2} \longrightarrow \mathcal{C}_{n_1} + \mathcal{C}_{n_2}$	charge additive/annihilative
Complex-formation	$\mathcal{C}_{n_1} + \mathcal{C}_{n_2} \longrightarrow \mathcal{C}_{n_1+n_2}$	charge additive/annihilative

Here n, n_1 and n_2 are integers and charge-class $\mathcal{C}_n$ is a proxy for any species belonging to this class.

Such a structured form gives us the following recipe to construct the generalized antithetic controller with any arbitrary number of species $C \geq 2$:

- Step 1: Let $\mathbf{Z}_0 = \emptyset$ and select a positive integer $N \geq 1$.
- Step 2: Partition the set of species $\mathcal{C} = \{\mathbf{Z}_0, \dots, \mathbf{Z}_C\}$ into charge-classes $\mathcal{C}_n$ for $n = 0, \pm 1, \pm 2, \dots, \pm N$. Ensure that $\mathbf{Z}_0 \in \mathcal{C}_0$, $\mathbf{Z}_1 \in \mathcal{C}_1$ and $\mathbf{Z}_2 \in \mathcal{C}_{-1}$, but some of the other charge-classes can be empty.
- Step 3: For each nonempty charge-class $\mathcal{C}_n$, add any number of conversion reactions among the species within this class $\mathcal{C}_n$.
- Step 4: Add any number of splitting and complex-formation reactions among species in $\mathcal{C}_{n_1}, \mathcal{C}_{n_2}$ and $\mathcal{C}_{n_1+n_2}$.
- Step 5: Check if the state-space $\mathbb{N}_0^C$ is irreducible for the control dynamics (using techniques in [53] for example). If this holds, then we have successfully constructed a generalized antithetic controller.

Note that due to Proposition S2.7, step 5 can only be successful if we have at least one generalized sequestration reaction (S45) which is a complex-formation reaction that can be written in terms of charge-classes as

$$\mathcal{C}_n + \mathcal{C}_{-n} \longrightarrow \mathcal{C}_0 \quad \text{for} \quad n \neq 0. \quad (\text{S51})$$

S2.2.4 RPA in the deterministic setting

So far we have worked in the stochastic setting, where the intracellular dynamics is described as a continuous-time Markov chain (CTMC). However most of the existing work on biomolecular controllers is carried out in the deterministic setting, where the dynamics is described by a system of ODEs [64]. Such deterministic models are accurate when all the constituent species are in high copy-numbers, and in fact these models can be viewed as approximations of the corresponding stochastic CTMC-based models in the limit of infinite system-size [35].

Theoretical studies on RPA in the deterministic setting have mostly focused on finding RPA enabling network topologies by either exhaustively exploring the dynamics of all possible networks of a fixed size [64], or by constructing an algebraic signature for RPA and using this as a means to identify topologies that demonstrate RPA [59, 60]. These approaches, however, apply to networks with fixed known structures and deterministic dynamics. Hence they do not address scenarios when the controlled

network of interest has either unknown players or interactions, nor do they address the important effects of intrinsic biochemical noise that can be quite consequential at the molecular level [47]. To deal with these issues, in this work the “controller network” and the “controlled network” are explicitly separated, and the latter can be completely arbitrary while the former can have uncertain parameters. Both networks are subjected to intrinsic stochastic noise and we also allow multiple potential input nodes for the disturbances to enter. This enables consideration of more realistic scenarios in cell biology, and is particularly suited for synthetic biology applications, where the intracellular network to be controlled is often poorly characterized, and there are several sources of disturbances due to the effects of context dependence [65].

Given the two modeling approaches, stochastic and deterministic, it is important to understand the dependence between the classes of RPA-achieving controllers in the two modeling paradigms. The goal of the analysis in this section is to provide such an understanding. Specifically we prove that the class of RPA-achieving controllers in the stochastic setting (denoted by $\mathcal{S}$) is *strictly contained* in the class of RPA-achieving controllers in the deterministic setting (denoted by $\mathcal{D}$), i.e.

$$\mathcal{S} \subsetneq \mathcal{D}. \quad (\text{S52})$$

Recall that we argued in Section S2.2.2 that all controllers in class $\mathcal{S}$ must contain at least two species. Hence the strict containment in (S52) is testified by the one-species controller in [66] which has set-point encoding reaction of the form (S23), the measurement reaction of the form (S26) and no reactions in the set $\mathcal{K}_C$. This controller does not embed the antithetic motif but it achieves RPA in the deterministic setting and so it belongs to $\mathcal{D}$. This controller does not belong to $\mathcal{S}$ because in the presence of noise the dynamics will get trapped in an absorbing state in the long run, leading to a loss of feedback and a state of dysregulation. However the expected time to reach this catastrophic eventuality depends on several factors, like the abundance of molecular species used in the implementation, and may not necessarily preclude near perfect adaptation over shorter time-scales.

The structure and parametric constraints for a controller motif to belong to class $\mathcal{D}$ has been investigated thoroughly in [59] and [60]. We shall discuss later how our results relate to these existing results.

Analysis of deterministic models

Consider a reaction network as in Section S2.1.1, with N species $\mathbf{X}_1, \dots, \mathbf{X}_N$ and reactions in the set $\mathcal{K}_N$ of the form (S8). In the deterministic setting, the state at time t is the vector of species concentrations $x(t) = (x_1(t), \dots, x_N(t)) \in \mathbb{R}_+^N$ where $\mathbb{R}_+ = [0, \infty)$ is the set of nonnegative real numbers. For each reaction $k \in \mathcal{K}_N$ we denote its stoichiometric vector by ζ_k and we let $\lambda_k(x, \gamma)$ be the associated propensity function which depends on a scalar parameter γ . The dynamics $(x_\gamma(t))_{t \geq 0}$ is described by the following system of ODEs

$$\frac{dx}{dt} = D(x, \gamma) := \sum_{k \in \mathcal{K}_N} \lambda_k(x, \gamma) \zeta_k. \quad (\text{S53})$$

We shall assume that propensity functions $\lambda_k(x, \gamma)$ are continuously differentiable and there exists an open set $\mathcal{E} \times \Gamma \subset \mathbb{R}_+^N \times \mathbb{R}$ such that for any $\gamma \in \Gamma$, if the dynamics starts in $\mathcal{E}$ (i.e. $x_\gamma(0) \in \mathcal{E}$) then it stays in $\mathcal{E}$ for all times (i.e. $x_\gamma(t) \in \mathcal{E}$ for all $t \geq 0$). Furthermore we shall assume asymptotic stability for each fixed parameter value $\gamma \in \Gamma$, in the sense that there exists a globally attracting fixed-point $\bar{x}_\gamma \in \mathcal{E}$ such that

$$x_\gamma(t) \rightarrow \bar{x}_\gamma \quad \text{as} \quad t \rightarrow \infty$$

for any initial state $x_\gamma(0) \in \mathcal{E}$. The map $\gamma \mapsto \bar{x}_\gamma$ is assumed to be continuously differentiable as well.

Note that since $\bar{x}_\gamma$ is a fixed-point for the vector field $D(x, \gamma)$ we have

$$D(\bar{x}_\gamma, \gamma) = \mathbf{0} \quad (\text{S54})$$

and as this fixed-point is attracting the Jacobian matrix for this vector field with respect to x , denoted by $\mathcal{J}_x D(x, \gamma)$ must be a Hurwitz stable matrix at $x = \bar{x}_\gamma$, i.e. all the eigenvalues of this matrix have strictly negative real parts. In particular, this matrix is invertible and differentiating (S54) with respect to parameter γ , we get

$$\mathcal{J}_x D(\bar{x}_\gamma, \gamma) \partial_\gamma \bar{x}_\gamma + \partial_\gamma D(\bar{x}_\gamma, \gamma) = \mathbf{0}$$

which yields

$$\partial_\gamma \bar{x}_\gamma = -[\mathcal{J}_x D(\bar{x}_\gamma, \gamma)]^{-1} \partial_\gamma D(\bar{x}_\gamma, \gamma), \quad (\text{S55})$$

where ∂_γ denotes the partial-differential operator $\partial/\partial\gamma$. This proves that for any real-valued continuously differentiable function $f : \mathcal{E} \rightarrow \mathbb{R}$, the steady-state parametric sensitivity is given by

$$\partial_\gamma f(\bar{x}_\gamma) = (\nabla_x f(\bar{x}_\gamma))^T \partial_\gamma \bar{x}_\gamma = -(\nabla_x f(\bar{x}_\gamma))^T [\mathcal{J}_x D(\bar{x}_\gamma, \gamma)]^{-1} \partial_\gamma D(\bar{x}_\gamma, \gamma), \quad (\text{S56})$$

where ∇_x denotes the gradient operator.

Now suppose for some real-valued continuously differentiable function $f : \mathcal{E} \rightarrow \mathbb{R}$ and some $\mu > 0$ we have *Perfect Adaptation* (PA) (see (S16)) given by

$$\lim_{t \rightarrow \infty} f(x_\gamma(t)) = f(\bar{x}_\gamma) = \mu \quad (\text{S57})$$

for any initial state $x_\gamma(0) \in \mathcal{E}$. For this PA condition to hold *robustly* with respect to the choice of parameter γ the following zero-sensitivity condition (see Section S2.2.4) must be satisfied

$$\lim_{t \rightarrow \infty} \partial_\gamma f(x_\gamma(t)) = \partial_\gamma f(\bar{x}_\gamma) = 0 \quad \text{for any } \gamma \in \Gamma. \quad (\text{S58})$$

Recall that for any invertible matrix $M = [M_{ij}]$ with determinant $\det(M) \neq 0$, the inverse matrix M^{-1} can be written in terms of the *cofactor* matrix $C = [C_{ij}]$ as

$$M^{-1} = \frac{1}{\det(M)} C^T,$$

where C_{ij} is the product of $(-1)^{i+j}$ and the determinant of the minor matrix of M formed by removing its i -th row and j -th column. Using the sensitivity formula (S56), the RPA condition (S58) can be expressed in terms of the cofactor matrix C of the Jacobian matrix $\mathcal{J}_x D(\bar{x}_\gamma, \gamma)$ as

$$(\nabla_x f(\bar{x}_\gamma))^T C^T \partial_\gamma D(\bar{x}_\gamma, \gamma) = 0, \quad (\text{S59})$$

for any $\gamma \in \Gamma$. This “cofactor condition” for RPA was reported in [59] and used for the purpose of designing topologies can generate RPA under certain parameter constraints.

The RPA condition (S59) can be further simplified in the case of a *single-input single-output* system. Suppose parameter γ in an “input” to the system which only affects the dynamics of the input species $\mathbf{X}_1$. Therefore we can write

$$\partial_\gamma D(\bar{x}_\gamma, \gamma) = \delta_i e_i$$

where δ_i is some non-zero real number and e_i is the vector in $\mathbb{R}^N$ whose i -th component is 1 and the rest are all zeros. Furthermore, suppose that constant perturbations in the input are “rejected” by steady-state concentration of output species $\mathbf{X}_L$, which implies that (S59) holds for $f(x) = \theta x_L$, where θ is some positive constant. Since $\nabla_x f(\bar{x}_\gamma) = \theta e_L$ we can express condition (S59) as

$$e_L^T C^T e_i = C_{iL} = 0. \quad (\text{S60})$$

This condition simply says that determinant of the input-output minor matrix of the Jacobian matrix $\mathcal{J}_x D(\bar{x}_\gamma, \gamma)$ is zero. This input-output minor matrix is the $(N - 1) \times (N - 1)$ matrix created by removing the i -th row and the L -th column from the $N \times N$ Jacobian matrix. The determinant condition (S60) was systematically disentangled by Araujo and Liotta in [60], to yield a *topological* basis for all RPA-achieving networks in the deterministic setting. This basis consists of two types of subnetworks, which are called *Opposer Modules* and *Balancer Modules* respectively, and they correspond to two characteristically different ways of robustly encoding integral action that is required for RPA. The Opposer Modules have the feedback structure (see Box 1 (Main Text), Fig. a) and in particular, the antithetic controller motif in [63] is an example of a two-node Opposer Module. On the other hand, the Balancer Module generalizes the notion of an *incoherent feed-forward* (IFF) motif that has been found in many biological examples [67, 68, 69]. As explained in [70], designing an IFF-based controller requires some information about the structure of the network being controlled, and hence this controlled network cannot be completely arbitrary as in the scenario we consider in this paper (see Section S2.2.1). Therefore only feedback controllers are considered in this paper which are composed of Opposer Modules in the deterministic setting. We shall soon prove that relation (S52) holds, and hence all RPA-achieving controllers in the stochastic setting will also be composed of these Opposer Modules. Our results show that in fact these Opposer Modules must necessarily have the structure of the generalized antithetic controller (see Section S2.2.3). Note that the analysis in [60] was at the level of interaction fluxes among network species, and the exact form of the reactions that implement these interactions is not important in the deterministic case. However the exact form of the reactions becomes important in the stochastic setting and this leads to stricter conditions on RPA-achieving controllers as demonstrated by our results.

We end this section with a simple proposition that provides an alternative way to check RPA in the deterministic setting.

Proposition S2.9. *Suppose there exists a twice-continuously differentiable function $F_\gamma : \mathcal{E} \rightarrow \mathbb{R}$ such that*

$$(\nabla_x F_\gamma(x))^T D(x, \gamma) = \mu - f(x) \quad \text{for all } x \in \mathcal{E}. \quad (\text{S61})$$

Then PA (S57) is satisfied and the zero-sensitivity condition (S58) holds if and only if

$$(\nabla_x F(\bar{x}_\gamma))^T \partial_\gamma D(\bar{x}_\gamma, \gamma) = 0, \quad (\text{S62})$$

for any $\gamma \in \Gamma$. In particular condition (S62) is equivalent to the “cofactor condition” (S59).

Remark S2.10. *Note that equation (S61) is essentially the Poisson equation (S30) in the deterministic setting. The solution of this equation F_γ specifies the “integral action” in the sense that if $u(t) = F_\gamma(x_\gamma(t))$ then*

$$\frac{du}{dt} = (\nabla_x F_\gamma(x_\gamma(t)))^T D(x_\gamma(t), \gamma) = \mu - f(x_\gamma(t)), \quad (\text{S63})$$

which shows that $u(t)$ is the time-integral of the “error” $(\mu - f(x_\gamma(t)))$ in the PA property (S57). Under certain conditions, the existence of F_γ (and hence the integral action) can be guaranteed by the Internal Model Principle in control theory [71, 72, 73, 74].

Proof. Equation (S63) along with asymptotic stability of the dynamics implies PA (S57). Differentiating (S61) with respect to x we obtain

$$D(x, \gamma)^T \mathcal{H}_x F_\gamma(x) + (\nabla_x F_\gamma(x))^T \mathcal{J}_x D(x, \gamma) = -(\nabla_x f(x))^T,$$

where $\mathcal{H}_x F_\gamma(x)$ denotes the Hessian matrix of function F_γ . Setting $x = \bar{x}_\gamma$ and using (S54) we obtain

$$(\nabla_x F_\gamma(\bar{x}_\gamma))^T \mathcal{J}_x D(\bar{x}_\gamma, \gamma) = -(\nabla_x f(\bar{x}_\gamma))^T$$

which proves that

$$(\nabla_x f(\bar{x}_\gamma))^T [\mathcal{J}_x D(\bar{x}_\gamma, \gamma)]^{-1} = -(\nabla_x F_\gamma(\bar{x}_\gamma))^T.$$

Plugging this in the sensitivity formula (S56) the zero-sensitivity condition (S58) yields (S62) and this completes the proof of this proposition. $\square$

Even though we consider a scalar parameter γ , the analysis in this section can be easily generalized to vector-valued parameters.

Proof of the containment relation (S52)

We adopt the same notation as in Section S2.2.1. Consider a C -species (viz. $\mathbf{Z}_1, \dots, \mathbf{Z}_C$) RPA achieving controller in class $\mathcal{S}$ which controls an arbitrary intracellular network with N species (viz. $\mathbf{X}_1, \dots, \mathbf{X}_N$). To prove (S52) we need to show that this controller achieves RPA when we model the dynamics in the deterministic setting. In this setting the state of the closed-loop network at any time t is given by a vector of species concentrations $(x_\gamma(t), z_\gamma(t))$ where $x_\gamma(t) = (x_{\gamma,1}(t), \dots, x_{\gamma,N}(t))$ is the network state, $z_\gamma(t) = (z_{\gamma,1}(t), \dots, z_{\gamma,C}(t))$ is the controller state and $\gamma = (\gamma_N, \gamma_C)$ is the set of all parameters. The closed-loop dynamics $(x_\gamma(t), z_\gamma(t))_{t \geq 0}$ is described by the following set of ODEs:

$$\frac{dx}{dt} = \sum_{k \in \mathcal{K}_N} \lambda_k(x, \gamma_N) \zeta_k + \kappa z_j e_1 \quad (\text{S64})$$

$$\frac{dz}{dt} = \sum_{k \in \mathcal{K}_C} \lambda_k(z, \gamma_C) \zeta_k + \mu e_1 + \theta x_L e_i. \quad (\text{S65})$$

As in the preceding section, we shall assume that all propensity functions are continuously differentiable in their arguments, and there exists an open set $\mathcal{E} \subset \mathbb{R}_+^N \times \mathbb{R}_+^C$ such that for any $\gamma \in \Gamma = \Gamma_N \times \Gamma_C$, if the dynamics starts in $\mathcal{E}$ (i.e. $(x_\gamma(0), z_\gamma(0)) \in \mathcal{E}$) then it stays in $\mathcal{E}$ for all times (i.e. $(x_\gamma(t), z_\gamma(t)) \in \mathcal{E}$ for all $t \geq 0$). Furthermore we shall assume asymptotic stability for each fixed parameter value $\gamma \in \Gamma$, i.e. there exists a globally attracting fixed-point $(\bar{x}_\gamma, \bar{z}_\gamma) \in \mathcal{E}$ such that

$$(x_\gamma(t), z_\gamma(t)) \rightarrow (\bar{x}_\gamma, \bar{z}_\gamma) \quad \text{as} \quad t \rightarrow \infty$$

for any initial state $(x_\gamma(0), z_\gamma(0)) \in \mathcal{E}$. The map $\gamma \mapsto (\bar{x}_\gamma, \bar{z}_\gamma)$ is assumed to be continuously differentiable as well.

Observe that one can write the ODEs (S64)-(S65) as

$$\frac{d}{dt} \begin{bmatrix} x \\ z \end{bmatrix} = D(x, z, \gamma) := \begin{bmatrix} \sum_{k \in \mathcal{K}_N} \lambda_k(x, \gamma_N) \zeta_k + \kappa z_j e_1 \\ \sum_{k \in \mathcal{K}_C} \lambda_k(z, \gamma_C) \zeta_k + \mu e_1 + \theta x_L e_i \end{bmatrix}.$$

Let $q = (q_1, \dots, q_C)$ be the vector in Theorem S2.5 that characterizes our RPA-achieving controller in the stochastic setting. Define the function $F_\gamma : \mathcal{E} \rightarrow \mathbb{R}$ by

$$F_\gamma(x, z) = q^T z = \sum_{i=1}^C q_i z_i.$$

Then for any $(x, z) \in \mathcal{E}$

$$\nabla_{(x,z)} F_\gamma(x, z) = (\mathbf{0}, q)$$

which shows that

$$(\nabla_{(x,z)} F_\gamma(x, z))^T D(x, z, \gamma) = \mu - \theta x_L + \sum_{k \in \mathcal{K}_C} \lambda_k(z, \gamma_C) q^T \zeta_k = \mu - \theta x_L,$$

where the last equality holds because $q_1 = -q_i = 1$ and q is orthogonal to all the stoichiometric vectors for reactions in $\mathcal{K}_C$ (see (S41)). Therefore (S61) holds with $f(x) = \theta x_L$ and for any parameter γ_i in the parameter vector $\gamma = (\gamma_N, \gamma_C)$ we have

$$(\nabla_{(x,z)} F_\gamma(x, z))^T \partial_{\gamma_i} D(x, z, \gamma) = \sum_{k \in \mathcal{K}_C} \partial_{\gamma_i} \lambda_k(z, \gamma_C) q^T \zeta_k = 0.$$

This proves condition (S62) with respect to all the parameters and hence Proposition S2.9 implies that RPA is achieved by the controller in the deterministic setting as well. Hence our controller is in class $\mathcal{D}$ which proves relation (S52).

As mentioned before, the fact that this containment is strict is demonstrated by the existence of a one-species controller (see [66]) that achieves RPA in the deterministic setting but not in the stochastic setting. In this controller there are no reactions in $\mathcal{K}_C$, and a single control species $\mathbf{Z}_1 = \mathbf{Z}$ is responsible for both set-point encoding (via reaction (S23)) and output sensing/measurement (via reaction (S26)). Hence the closed-dynamics is given by

$$\frac{d}{dt} \begin{bmatrix} x \\ z \end{bmatrix} = D(x, z, \gamma) := \begin{bmatrix} \sum_{k \in \mathcal{K}_N} \lambda_k(x, \gamma_N) \zeta_k + \kappa z e_1 \\ z(\mu - \theta x_L) \end{bmatrix}.$$

If we define the function $F_\gamma : \mathcal{E} \rightarrow R$ by $F_\gamma(x, z) = \log(z)$ then

$$\nabla_{(x,z)} F_\gamma(x, z) = (\mathbf{0}, z^{-1})$$

which shows that

$$(\nabla_{(x,z)} F_\gamma(x, z))^T D(x, z, \gamma) = z^{-1} z(\mu - \theta x_L) = \mu - \theta x_L.$$

Hence (S61) holds with $f(x) = \theta x_L$, and since there are no reactions in $\mathcal{K}_C$, condition (S62) holds trivially. Therefore Proposition S2.9 implies that this single-species controller achieves RPA in the deterministic setting, thereby proving that the containment relation (S52) is strict.

S3 Supplementary Methods

Considerations for fixing circuit set-point

In the current design, both the closed- and open-loop circuits operate at low concentrations of HSL (nM). As a result, the output fluorescence is sensitive to small absolute changes in HSL concentrations. Thus, when fixing the set-point, it is important to measure concentrations of HSL with care, as small pipetting errors can have outsized effects. It is also important not to store the HSL in aqueous medium for long periods of time in order to minimize degradation of the inducer. If longer-term storage is required, storage of the medium at 4°C is advised.

Construction of strains and plasmids

SKA703

The host strain MG1655 $\Delta araCBAD \Delta lacIZYA \Delta araE \Delta araFGH attB::lacYA177C \Delta rhaSRT \Delta rhaBADM Tn7::tetR$, referred to as SKA703, was constructed as follows.

Deletion of arabinose, lactose, and rhamnose metabolic pathways

The endogenous genes for arabinose, lactose, and rhamnose metabolism, regulation, and transport were deleted from MG1655 (ATCC 47076). Suicide plasmids for allelic exchange were constructed using suicide vector pRE112 [75]. Briefly, MG1655 genomic DNA was used as a template to amplify 500 bp

regions flanking the genes to be deleted (Supplementary Table S1). Resulting PCR products were combined using overlapping-extension PCR (OE-PCR), ligated into pRE112 between KpnI and SacI, and cloned into either CC118(λ pir) [76] or DH5 α pir [77, 78] (Supplementary Table S1 and Extended Data Fig. 9a). The suicide plasmids were then used for sequential deletion in MG1655 by allelic exchange using standard methods [75].

λ -integration of *lacYA177C* into host chromosome

Integration vector pFL503 [79] containing relaxed specificity lac permease variant *lacYA177C* under constitutive *rrnBP1*_{CGC-ATA} promoter control [79] was modified for our purposes. The FRT-flanked kanamycin-resistance cassette with a flanking NotI site added was amplified from pKD13 [80] using primers 84 and 85. The resulting PCR product was cut with NheI and ligated into pFL503 cut with XbaI, replacing the *pir* gene with the antibiotic resistance cassette to construct pSKA27 (Extended Data Fig. 9b). The new integration plasmid pSKA27 was cut with NotI (Extended Data Fig. 9b) and religated together to form a closed circle containing the FRT-flanked kanamycin-cassette, *lacYA177C*, and the *attP* site. The closed circle was then integrated into the host chromosome at the *attB* site using λ integrase plasmid pJW27 as previously described [79]. The kanamycin-resistance gene was then excised from the chromosome using FLP-recombinase plasmid pCP20 as previously described [81].

***Tn7*-integration of *tetR* into host chromosome**

The FRT-flanked kanamycin-resistance cassette with a flanking NotI site added was amplified from pKD13 [80] using primers 84 and 85. The resulting PCR product was ligated into pUC18R6KT-mini-Tn7T [82] between SalI and NheI and cloned into CC118(λ pir) to construct pSKA207 (pUC18R6KT-mini-Tn7T-FRT-*kan*-FRT, unpublished, sequence available upon request). A *tetR* repressor gene under the control of constitutive N25 promoter control was amplified from pZEtetR21-gfp [83] using primers 382 and 383. The resulting PCR product was ligated into pSKA207 between XhoI and SpeI and cloned into CC118(λ pir) to construct pSKA273 (Extended Data Fig. 9c). The new integration plasmid and *Tn7* transposase plasmid pTNS3 [84] were transformed into MFDpir [85] and cells were maintained by the addition of diaminopimelic acid (100 μ g/mL) to the media. MFDpir pSKA273 and MFDpir pTNS3 were used to integrate the FRT-flanked kanamycin-resistance cassette and *tetR* into the host chromosome using tri-parental mating [84][85]. The kanamycin-resistance gene was then excised from the chromosome using FLP-recombinase plasmid pCP20 as previously described [81].

Negative perturbation plasmid (pSKA417, Genbank accession #MK775703)

Initially, the *Mesoplasma florum* Lon protease gene *mflon* was amplified from P_{LacO-1}-mflon (Addgene #45152) [86] using primers 510 and 511 and cut with BglII and XbaI. *luxR* and *lux* promoter (P_{LUX}) [87] flanked by a T7 terminator [88, 89] was amplified from pSKA353 (pZAspec-luxR-P_{LUX}-gfpmut3, unpublished, sequence available upon request) using primers 518 and 268 and cut with AatII and BglII. An *rrnB* T1 terminator, p15A origin of replication, and spectinomycin-resistance cassette was amplified from pSKA353 using primer pairs 519/504 and 349/505 to mutate an internal BglII site within the spectinomycin-resistance promoter region. The resulting PCRs were combined by OE-PCR using primers 519 and 349 and cut with XbaI and AatII. All pieces were ligated together and cloned into NEB Turbo (New England Biolabs) to construct pSKA357 (pZAspec-luxR-P_{LUX}-*mflon*, unpublished, sequence available upon request).

The plasmid pSKA357 was subsequently converted to an anhydrotetracycline-inducible promoter by amplifying a synthetic TetR-repressible promoter (P_{TET}), based on *LtetO-1* [46] and *tac* [90] promoters from pSKA332 (pZA2-P_{TET}-*cherry::lacI::LVA*, unpublished, sequence available upon request) and a strong ribosomal binding site (sRBS, 5'-AGGAGGT-3') with primers 441 and 342. The resulting product was cut with AatII and BglII, ligated into pSKA357, and cloned into SKA703 to construct pSKA417 (pZAspec-P_{TET}-*mflon*, Extended Data Fig. 10a).

Antithetic closed-loop circuit precursor plasmid (pSKA538, Genbank accession #MK775704)

Antithetic circuit precursor plasmid pSKA538 was modularly constructed using isothermal assembly [91] and standard restriction-ligation methods. Three modules have been designed to be amplifiable with primer pairs oGL312/oGL110, oGL311/ oGL310, and oGL308/oGL309. The modules can be isothermally assembled between AatII and XbaI in a pZE1 plasmid backbone amplified from pZE1R-gfp [83] with primers oGL115 and oGL116. Modules are in the following order: AatII-Module 1-Module 2-Module 3-XbaI (Extended Data Fig. 10b).

Module 1: $P_{sigW-sRBS-V5::araC::pdt\#3c-FLAG::sfgfp::pdt\#1}$

Arabinose-dependent transcriptional activator *araC* from pBAD33 [92] with a N-terminal simian virus 5 V protein epitope tag (V5, GKIPNPLLGLDST) [93] and a C-terminal *mf*-Lon-specific Pdt#3c degradation tag (AANKNEENTNEVAQPMLNAGQANRRRV) [37] ($V5::araC::pdt\#3c$) was placed under control of a *Bacillus subtilis* strain 168 SigW-activated promoter containing an *E. coli* consensus UP element (P_{sigW}) [94] and a strong ribosomal binding site (5'-AGGAGGT-3'). *B. subtilis* str. 168 genomic DNA was a gift from S. Panke, ETH Zurich. Superfolder GFP (*sfgfp*) from pJT119b [95] with an N-terminal FLAG epitope tag (DYKDDDDK) [96] and C-terminal Pdt#1 degradation tag (AANKNEENTNEVPTFMLNAGQANRLQL) [37] ($FLAG::sfgfp::pdt\#1$) was transcriptionally fused immediately downstream of $V5::araC::pdt\#3c$ with a strong ribosomal binding site (5'-AGGAGGT-3'). An *E. coli rrnB T1T2* transcriptional terminator from pBAD33 [92] was placed immediately downstream of $FLAG::sfgfp::pdt\#1$. Module 1 is amplifiable with primers oGL312 and oGL110.

Module 2: $luxR-P_{LUX-RBS5000-sigW}$.

Sigma factor *sigW* [94] from *B. subtilis* str. 168 genomic DNA was placed under control of an HSL-dependent *Vibrio fischeri* LuxR-activated promoter (P_{LUX}) from pVRc11_987 (Addgene #49733) [87] and a synthetic ribosomal binding site with a translation initiation rate of 5000 units (RBS5000, RBS Library Calculator v2.0, 5'-CTTCTAGATCGCTAAAACGGGCTAGC-3') [97]. An *E. coli rrnB T1* transcriptional terminator from pBAD33 [92] was placed immediately downstream of *sigW*. The *luxR* regulator under constitutive control from pVRc11_987 (Addgene #49733) [87] was followed by a bacteriophage *T7g10* transcriptional terminator [88, 89] and placed divergent to $P_{LUX-RBS5000-sigW}$. Module 2 is amplifiable with primers oGL311 and oGL310.

Module 3: $P_{BAD-RBS5000-rsiW}$.

The cytoplasmic domain of anti-sigma factor *rsiW* [94] from *B. subtilis* str. 168 genomic DNA was placed under control of an arabinose-dependent AraC-activated promoter (P_{BAD}) from pBAD33 [92] deleted for the cAMP receptor protein binding site and a synthetic ribosomal binding with a translation initiation rate of 5000 units (RBS5000, RBS Library Calculator v2.0, 5'-CTTCTAGATCGCTAAAACGGGCTAGC-3') [97]. An *E. coli rrnB T1T2* transcriptional terminator from pBAD33 [92] was placed immediately downstream of *rsiW*. Module 3 is amplifiable with primers oGL308 and oGL309.

Open-loop control precursor plasmid (pSKA539, Genbank accession #MK775705)

The open-loop control plasmid pSKA539 is a variant of antithetic circuit plasmid pSKA538 deleted for the anti-sigma module (Extended Data Fig. 10b). The $P_{BAD-RBS5000-rsiW}$ module was excised from pSKA538 between Sall and EcoRI sites by restriction digestion. The cut plasmid was then blunted using the Quick Blunting Kit from New England Biolabs before being ligated closed.

Closed- and open-loop plasmids (pSKA562 and pSKA563, Genbank accession #MK775706 and MK775707)

Final closed- and open-loop plasmids pSKA562 and pSKA563 are variants of pSKA538 and pSKA539, respectively. Briefly, a second copy of *P_{sigW-sRBS}-V5::araC::pdt#3c-FLAG::sfGFP::pdt#1-rrnBT1T2* (Module 1) was added immediately downstream of the original copy (Extended Data Fig. 10c).

Closed- and open-loop strain preparation

Host strain SKA703 was transformed with closed-loop plasmid pSKA562 and perturbation plasmid pSKA417 or open-loop plasmid pSKA563 and perturbation plasmid pSKA417.

Growth rate control plasmids pSKA570, pSKA571, pSKA571-p15A, and host strain SKA1328

Deletion of *metE* from SKA703

metE was deleted from SKA703 as previously described [98]. Briefly, a FRT-flanked chloramphenicol-resistance cassette was amplified from plasmid pSKA373 [98] and tagged with *metE* upstream and downstream sequence using primers 558 and 559 [80, 98]. The PCR product was recombineered into SKA703 using pSIM6 as previously described [99] to construct SKA703 $\Delta metE::cat$ (SKA1328).

Antithetic closed-loop and open-loop growth rate control plasmids (pSKA570, pSKA571, and pSKA571-p15A, Genbank accession #MK775708-MK775710)

Growth rate control plasmids pSKA570 and pSKA571 are variants of antithetic circuit plasmid pSKA562 and open-loop circuit plasmid pSKA563, respectively. Both copies of *FLAG::sfGFP::pdt#1* and their corresponding RBS were replaced with *metE::pdt#3c* with a weakened variant of the endogenous *metE* RBS with a translation initiation rate of 73.88 (wRBS, RBS Library Calculator v2.0, 5' CTGCAGTAGACCAAGAAAAA 3') [97] (Extended Data Fig. 10d). Reduced plasmid copy variant pSKA571-p15A was constructed by replacing the ColE1 origin of replication on pSKA571 between XbaI and SpeI for the p15A origin or replication on pSKA417 between the same restriction sites (Extended Data Fig. 10d).

Growth rate control strain preparation

Host strain SKA1328 was transformed with closed-loop growth rate control plasmid pSKA570 or open-loop growth rate control plasmids pSKA571 or pSKA571-p15A.

Wild-type growth rate control strain preparation

metE⁺ strain SKA703 was transformed with empty plasmid vector pSKA47 (unpublished, sequence available upon request) with a ColE1 origin or replication and ampicillin-resistance.

Supplementary Tables

Plasmid	Genes targeted for deletion	Description	Left flanking PCR primers	Right flanking PCR primers	OE-PCR primers
pSKA1	<i>araCBAD</i>	Arabinose metabolism and regulation	1/2	3/4	1/4
pSKA2	<i>lacZYA</i>	Lactose metabolism, regulation, and transport	7/8	9/10	7/10
pSKA21	<i>araE</i>	Arabinose transport	54/55	56/57	54/57
pSKA22	<i>araFGH</i>	Arabinose transport	60/61	62/63	60/63
pSKA101	<i>rhaTRSBADM</i>	Rhamnose metabolism, regulation, and transport	150/151	152/153	150/153

Supplementary Table S1: **Construction of suicide plasmids for allelic exchange.**

Strain	Genotype	Reference
CC118(λ pir)	$\Delta(ara\ leu)\ araD\ \Delta lacX74\ galE\ galK\ phoA20\ thi-1\ rpsE\ rpoB\ argE(Am)\ recA1\ \lambda::pir$	[76]
DH5 α λ pir	$F- supE44\ hsdR17(rK-mK+)\ recA1\ \phi80dlacZ\Delta M15\ \Delta lacU169\ gyrA96\ endA1\ thi-1\ relA1\ deoR\ \lambda::pir$	[77, 78]
MFDpir	MG1655 RP4-2 Tc::[$\Delta Mu1::aac(3)IV-\Delta aphA-\Delta nic35-\Delta Mu2::zeo$] $\Delta dapA::(erm-pir)\ \Delta recA$	[85]
MG1655 (ATCC 47076)	F- λ - <i>rph-I</i>	[100]
NEB Turbo	F' <i>proA+B+</i> <i>lacIq</i> $\Delta lacZM15/fhuA2\ \Delta(lac-proAB)\ glnV\ galK16\ galE15\ R(zgb-210::Tn10)Tet^S\ endA1\ thi-1\ \Delta(hsdS-mcrB)5$	New England Biolabs
SKA703	Host strain, MG1655 $\Delta araCBAD\ \Delta lacIZYA\ \Delta araE\ \Delta araFGH\ attB::P_{rrnBP1}(CTC-AGA)-lacYA177C\ \Delta rhaSRT\ \Delta rhaBADM\ Tn7::P_{N25-tetR}$	This study
SKA1328	Host strain, MG1655 $\Delta araCBAD\ \Delta lacIZYA\ \Delta araE\ \Delta araFGH\ attB::P_{rrnBP1}(CTC-AGA)-lacYA177C\ \Delta rhaSRT\ \Delta rhaBADM\ Tn7::P_{N25-tetR}\ \Delta metE::cat$	This study
Plasmid	Description	Reference
pBAD33	<i>araC</i> , P_{BAD} template, p15A <i>ori</i> , <i>cam</i> ^R	[92]
pCP20	FLP recombinase plasmid, temperature-sensitive, <i>amp</i> ^R , <i>cam</i> ^R	[81]
pFL503	<i>attP</i> , <i>araC</i> , $P_{BAD-pir}$, $P_{rrnBP1}(CTC-AGA)-lacYA177C$, pBR322 <i>ori</i> , <i>amp</i> ^R	[79]
pJT119b	<i>sfgfp</i> template, ColE1 <i>ori</i> , <i>cam</i> ^R	[95]
pJW27	λ Int helper plasmid, pSC101* <i>ori</i> , temperature-sensitive, <i>cam</i> ^R	[79]
pKD13	R6K γ <i>ori</i> , FRT-flanked <i>kan</i> ^R , <i>amp</i> ^R	[80]
$P_{LlacO-1}$ -mfLon	Addgene #45152, <i>lacI</i> , ColE1 <i>ori</i> , <i>cam</i> ^R	[86]
pSIM6	λ Red recombinase plasmid, pSC101* <i>ori</i> , temperature-sensitive, <i>amp</i> ^R	[99]
pSKA1	$\Delta araCBAD$ suicide plasmid, <i>sacB</i> , R6K γ <i>ori</i> , <i>cam</i> ^R	This study
pSKA2	$\Delta lacIZYA$ suicide plasmid, <i>sacB</i> , R6K γ <i>ori</i> , <i>cam</i> ^R	This study
pSKA21	$\Delta araE$ suicide plasmid, <i>sacB</i> , R6K γ <i>ori</i> , <i>cam</i> ^R	This study
pSKA22	$\Delta araFGH$ suicide plasmid, <i>sacB</i> , R6K γ <i>ori</i> , <i>cam</i> ^R	This study
pSKA27	pFL503::FRT- <i>kan</i> -FRT, <i>attP</i> , <i>araC</i> , $P_{BAD-pir}$, $P_{rrnBP1}(CTC-AGA)-lacYA177C$, pBR322 <i>ori</i> , <i>amp</i> ^R , <i>kan</i> ^R	This study
pSKA47	Empty plasmid vector, ColE1 <i>ori</i> , <i>amp</i> ^R	Unpublished
pSKA101	$\Delta rhaTRSBADM$ suicide plasmid, <i>sacB</i> , R6K γ <i>ori</i> , <i>cam</i> ^R	This study
pSKA207	pUC18R6KT-mini-Tn7T-FRT- <i>kan</i> -FRT, R6K γ <i>ori</i> , <i>oriT</i> , <i>kan</i> ^R , <i>amp</i> ^R	Unpublished
pSKA273	pUC18R6KT-mini-Tn7T-FRT- <i>kan</i> -FRT- $P_{N25-tetR}$, R6K γ <i>ori</i> , <i>oriT</i> , <i>kan</i> ^R , <i>amp</i> ^R	This study
pSKA332	$P_{TET-mcherry}::lacI::LVA$, p15A <i>ori</i> , <i>kan</i> ^R	Unpublished
pSKA353	<i>luxR</i> -T7 terminator, $P_{LUX-gfpmut3-rrnBT1}$ terminator, p15A <i>ori</i> , <i>spec</i> ^R	Unpublished

pSKA357	<i>luxR</i> -T7 terminator, P _{LUX} - <i>mflon-rrnB</i> T1 terminator, p15A <i>ori</i> , spec ^R	Unpublished
pSKA373	R6K γ <i>ori</i> , FRT-flanked cam ^R , amp ^R	[98]
pSKA417	Negative disturbance plasmid, P _{TET-sRBS} - <i>mflon</i> , p15A <i>ori</i> , spec ^R	This study
pSKA538	Closed-loop antithetic circuit precursor plasmid, ColE1 <i>ori</i> , amp ^R (pZE1-P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-FLAG::sfgfp::pdt#1-rrnBT1T2-luxR-T7</i> -P _{LUX-RBS5000-sigW-rrnBT1} -P _{BAD-RBS5000-rsiW-rrnBT1T2})	This study
pSKA539	Open-loop control precursor plasmid, ColE1 <i>ori</i> , amp ^R (pZE1-P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-FLAG::sfgfp::pdt#1-rrnBT1T2-luxR-T7</i> -P _{LUX-RBS5000-sigW-rrnBT1})	This study
pSKA562	Closed-loop antithetic circuit final plasmid, ColE1 <i>ori</i> , amp ^R (pZE1-P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-FLAG::sfgfp::pdt#1-rrnBT1T2</i> -P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-FLAG::sfgfp::pdt#1-rrnBT1T2-luxR-T7</i> -P _{LUX-RBS5000-sigW-rrnBT1} -P _{BAD-RBS5000-rsiW-rrnBT1T2})	This study
pSKA563	Open-loop control final plasmid, ColE1 <i>ori</i> , amp ^R (pZE1-P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-FLAG::sfgfp::pdt#1-rrnBT1T2</i> -P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-FLAG::sfgfp::pdt#1-rrnBT1T2-luxR-T7</i> -P _{LUX-RBS5000-sigW-rrnBT1})	This study
pSKA570	Closed-loop antithetic growth rate control plasmid, ColE1 <i>ori</i> , amp ^R (pZE1-P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-wRBS-metE::pdt#3c-rrnBT1T2</i> -P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-wRBS-metE::pdt#3c-rrnBT1T2-luxR-T7</i> -P _{LUX-RBS5000-sigW-rrnBT1} -P _{BAD-RBS5000-rsiW-rrnBT1T2})	This study
pSKA571	Open-loop growth rate control plasmid, ColE1 <i>ori</i> , amp ^R (pZE1-P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-wRBS-metE::pdt#3c-rrnBT1T2</i> -P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-wRBS-metE::pdt#3c-rrnBT1T2-luxR-T7</i> -P _{LUX-RBS5000-sigW-rrnBT1})	This study
pSKA571-p15A	Open-loop growth rate control plasmid, p15A <i>ori</i> , amp ^R (pZE1-P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-wRBS-metE::pdt#3c-rrnBT1T2</i> -P _{sigW-sRBS} -V5:: <i>araC</i> :: <i>pdt#3c-wRBS-metE::pdt#3c-rrnBT1T2-luxR-T7</i> -P _{LUX-RBS5000-sigW-rrnBT1})	This study
pTNS3	Tn7 helper plasmid, R6K γ <i>ori</i> , <i>oriT</i> , amp ^R	[84]
pUC18R6KT-mini-Tn7T	mini-Tn7, R6K γ <i>ori</i> , <i>oriT</i> , amp ^R	[82]
pVRc11_987	Addgene #49733, <i>luxR</i> and P _{LUX} template, p15A <i>ori</i> , cam ^R	[87]
pZE1R-gfp	pZE1 plasmid backbone template, ColE1 <i>ori</i> , amp ^R	[83]
pZEtetR21-gfp	P _{LtetO-1} -gfp, P _{N25-tetR} , ColE1 <i>ori</i> , kan ^R	[83]
Primer	Sequence (5' to 3')	Description
1	CAACAAGAGCTCCTGGCAGGAAGTA-AAACG	SacI-araC-F

2	GTTTATCCAGCAGCGTTTGAAAGTG-AATGATGTAGCCG	Δ araCBAD-R
3	CGGCTACATCATTCACTTTCAAACG-CTGCTGGATAAAC	Δ araCBAD-F
4	CAACAAGGTACCTAGAAACCCACTT-GAGCG	KpnI-araD-R
7	CAACAAGAGCTCGGATGTTGATGCA-ATGGTG	SacI-lacI-F
8	GAATAACCCGACAAGGAACGCACCA-CCCTGAATTGACTC	Δ lacIZYA-R
9	GAGTCAATTCAGGGTGGTGCGTTCC-TTGTCGGGTATTTC	Δ lacIZYA-F
10	CAACAAGGTACCGTTCTGCGGCTTT-ATCTG	KpnI-lacA-R
54	CAACAAGAGCTCGTCAATGCTCTCC-TTCCAG	SacI-araE-F
55	CAAGTGCCCAATCTCTATTAACGCA-TGTCGTTATGTGATGGATATTCC	Δ araE-R
56	GGAATATCCATCACATAACGACATG-CGTTAATAGAGATTGGGCACTTG	Δ araE-F
57	CAACAAGGTACCGACGTTTCGGTAA-ATTGCC	KpnI-araE-R
60	CAACAAGAGCTCGTGAAGGTAGAGA-TTGGATGC	SacI-araF-F
61	GCATCAGACAGTGCGTTTCGCATTA-CCTGGCGTTGAG	Δ araFGH-R
62	CTCAACGCCAGGTAATGCGAAACGC-ACTGTCTGATGC	Δ araFGH-F
63	CAACAAGGTACCGGATACTGAGCGA-TGACTG	KpnI-araH-R
84	CAACAAGCTAGCGCGGCCGCATTCC-GGGGATCCGTCGACC	NheI-NotI-pKD13-F
85	CAACAAGCTAGCGTGTAGGCTGGAG-CTGCTTC	NheI-pKD13-R
150	CAACAAGAGCTCGGTTTTCTTGTCT-GCTGGC	SacI-rhaT-F
151	CATTGGGCAATTACGGCAGGGCGTT-ACTCATCTTCTTATCCTCATC	Δ rhaTRSBADM-R
152	GATGAGGATAAGAAGATGAGTAACG-CCCTGCCGTAATTGCCCAATG	Δ rhaTRSBADM-F
153	CAACAAGGTACCGCTTTCCAGTCAA-TAGCCTGC	KpnI-rhaM-R
268	GTGAAAAGTTCTTCTCCTTTACGC	gfpmut3-R
342	GATAGCCATGTTATCCTCCTCG	mcherry-R
349	CAACAATCTAGAGGCATCAAATAAA-ACGAAAGGC	XbaI-T1 term-F
382	CAACAACCTCGAGAATCATAAAAAAT-TTATTTGCTTTCAGG	XhoI-PN25-F
383	CAACAACTAGTTTAAGACCCACTT-TCACATTTAAGTTG	SpeI-tetR-R
441	CCAGCTGGCAATTCCGAC	Ptet*-F
504	GTCTCTTGATCAGATATTGATCC	BglII mut-spec-F
505	GGATCAATATCTGATCAAGAGAC	BglII mut-spec-R
510	CAACAAAGATCTATGAGTAAAAAAA-TAAAACTGCCTATTTTCC	BglII-PLlacO-1-F
511	CAACAATCTAGATTATTTTGTTTTA-AAAACTATTGCAAATACCTC	XbaI-mflon-R

518	CCAGCTGGCAATTCCGACGTCCAAA- AAACCCCTCAAGACCC	T7 term-atII-spec-F
519	GGGTCTTGAGGGGTTTTTTGGACGT- CGGAATTGCCAGCTGG	spec-AatII-T7 term-R
oGL110	GAGTGCTATCATTGACTTACCGCCA- GAGTTTGTAGAAACGCAA	Module2- Module1-R
oGL115	TACTACATCACAGATGACAGGCAAA- AGGCCAGCAAAAGGC	Module3-XbaI- Backbone-F
oGL116	TTCATAACTGAACAGACAGCCGTCA- GGTGGCACTTTTCGG	Module1-AatII- Backbone-R
oGL308	GACGAAGCAACATTACACTAGTCGA- CTAGCATTTTTATCCATAAGATTAG- CGGATCCTACC	Module2- Module3-F
oGL309	CTGTCATCTGTGATGTAGTATCTAG- AGAATTCAGAGTTTGTAGAAACGCA-AAAAGGC	Backbone-XbaI- Module3-R
oGL310	TAGTGTAATGTTGCTTCGTCATTTG- TCCTACTCAGGAGAGCGTTC	Module3- Module2-R
oGL311	GTAAGTCAATGATAGCACTCGGTAC- CTCCAAAAAACCCCTCAAGACCCGT	Module1- Module2-F
oGL312	GCTGTCTGTTTCAGTTATGAACGAAA- AGTGCCACCTGACGTC	Backbone-AatII- Module1-F

Supplementary Table S2: **Strains, plasmids and primers used in this study.**

Supplementary References

- [34] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**, 2340–2361 (1977).
- [35] Kurtz, T. G. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics* **57**, 2976–2978 (1972).
- [36] Kurtz, T. G. Limit theorems and diffusion approximations for density dependent markov chains. In *Stochastic Systems: Modeling, Identification and Optimization, I*, 67–78 (Springer, 1976).
- [37] Cameron, D. E. & Collins, J. J. Tunable protein degradation in bacteria. *Nature Biotechnology* **32**, 1276–1281 (2014).
- [38] Sharma, U. K. & Chatterji, D. Differential mechanisms of binding of anti-sigma factors Escherichia coli Rsd and bacteriophage T4 AsiA to E. coli RNA polymerase lead to diverse physiological consequences. *Journal of Bacteriology* **190**, 3434–3443 (2008).
- [39] Mauri, M. & Klumpp, S. A model for sigma factor competition in bacterial cells. *PLoS Computational Biology* **10**, e1003845 (2014).
- [40] Buchler, N. & Louis, M. Molecular titration and ultrasensitivity in regulatory networks. *Journal of Molecular Biology* **384**, 1106 (2008).
- [41] Davidi, D. *et al.* Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro kcat measurements. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 3401–3406 (2016).
- [42] García-Contreras, R., Vos, P., Westerhoff, H. V. & Boogerd, F. C. Why in vivo may not equal in vitro—new effectors revealed by measurement of enzymatic activities under the same in vivo-like assay conditions. *The FEBS Journal* **279**, 4145–4159 (2012).
- [43] Gencoglu, M., Schmidt, A. & Becskei, A. Measurement of in vivo protein binding affinities in a signaling network with mass spectrometry. *ACS Synthetic Biology* **6**, 1305–1314 (2017).
- [44] Kalisky, T., Dekel, E. & Alon, U. Cost-benefit theory and optimal design of gene regulation functions. *Physical Biology* **4**, 229 (2007).
- [45] Santillán, M. & Mackey, M. C. Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon. *Biophysical Journal* **86**, 1282–1292 (2004).
- [46] Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the lacR/O, the tetR/O and arac/I1-I2 regulatory elements. *Nucleic Acids Research* **25**, 1203–1210 (1997).
- [47] Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–6 (2002).
- [48] McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 814–819 (1997).
- [49] Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8614–8619 (2001).
- [50] Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173 (2010).

- [51] Fraser, D. & Kaern, M. A chance at survival: gene expression noise and phenotypic diversification strategies. *Molecular Microbiology* **71**, 1333–1340 (2009).
- [52] Anderson, D. & Kurtz, T. Continuous time Markov chain models for chemical reaction networks. In Koeppl, H., Setti, G., di Bernardo, M. & Densmore, D. (eds.) *Design and Analysis of Biomolecular Circuits* (Springer-Verlag, 2011).
- [53] Gupta, A. & Khammash, M. Computational identification of irreducible state-spaces for stochastic reaction networks. *SIAM Journal on Applied Dynamical Systems* **17**, 1213–1266 (2018).
- [54] Ethier, S. N. & Kurtz, T. G. *Markov processes : Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics (John Wiley & Sons Inc., New York, 1986).
- [55] Meyn, S. & Tweedie, R. L. *Markov chains and stochastic stability* (Cambridge University Press, Cambridge, 2009), second edn. With a prologue by Peter W. Glynn.
- [56] Meyn, S. P. & Tweedie, R. L. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Advances in Applied Probability* **25**, 518–548 (1993).
- [57] Gupta, A., Briat, C. & Khammash, M. A scalable computational framework for establishing long-term behavior of stochastic reaction networks. *PLoS Computational Biology* **10**, e1003669 (2014).
- [58] Meyn, S. *Control techniques for complex networks* (Cambridge University Press, Cambridge, 2008).
- [59] Tang, Z. F. & McMillen, D. R. Design principles for the analysis and construction of robustly homeostatic biological networks. *Journal of Theoretical Biology* **408**, 274–289 (2016).
- [60] Araujo, R. P. & Liotta, L. A. The topological requirements for robust perfect adaptation in networks of any size. *Nature Communications* **9**, 1757 (2018).
- [61] Gupta, A. & Khammash, M. Sensitivity analysis for stochastic chemical reaction networks with multiple time-scales. *Electronic Journal of Probability* **19**, 1–53 (2014).
- [62] Dürrenberger, P., Gupta, A. & Khammash, M. A finite state projection method for steady-state sensitivity analysis of stochastic reaction networks. *The Journal of Chemical Physics* **150**, 134101 (2019).
- [63] Briat, C., Gupta, A. & Khammash, M. Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks. *Cell Systems* **2**, 15–26 (2016).
- [64] Ma, W., Trusina, A., El-Samad, H., Lim, W. A. & Tang, C. Defining network topologies that can achieve biochemical adaptation. *Cell* **138**, 760–773 (2009).
- [65] Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology—identifying causes of failure of synthetic biological systems. *Biotechnology Journal* **7**, 856–866 (2012).
- [66] Briat, C., Zechner, C. & Khammash, M. Design of a synthetic integral feedback circuit: dynamic analysis and DNA implementation. *ACS Synthetic Biology* **5**, 1108–1116 (2016).
- [67] Goentoro, L., Shoval, O., Kirschner, M. W. & Alon, U. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular Cell* **36**, 894–899 (2009).
- [68] Heidary, Z., Ghaisari, J., Moein, S., Naderi, M. & Gheisari, Y. Stochastic Petri net modeling of hypoxia pathway predicts a novel incoherent feed-forward loop controlling SDF-1 expression in acute kidney injury. *IEEE Transactions on Nanobioscience* **15**, 19–26 (2016).

- [69] Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11980–11985 (2003).
- [70] Xiao, F. & Doyle, J. C. Robust perfect adaptation in biomolecular reaction networks. *bioRxiv* 299057 (2018).
- [71] Francis, B. A. & Wonham, W. M. The internal model principle for linear multivariable regulators. *Applied Mathematics and Optimization* **2**, 170–194 (1975).
- [72] Francis, B. A. & Wonham, W. M. The internal model principle of control theory. *Automatica* **12**, 457–465 (1976).
- [73] Yi, T.-M., Huang, Y., Simon, M. I. & Doyle, J. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 4649–4653 (2000).
- [74] Sontag, E. D. Adaptation and regulation with signal detection implies internal model. *Systems & Control Letters* **50**, 119–126 (2003).
- [75] Edwards, R. A., Keller, L. H. & Schifferli, D. M. Improved allelic exchange vectors and their use to analyze 987P fimbria gene expression. *Gene* **207**, 149–157 (1998).
- [76] Herrero, M., de Lorenzo, V. & Timmis, K. N. Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. *Journal of Bacteriology* **172**, 6557–6567 (1990).
- [77] Kolter, R., Inuzuka, M. & Helinski, D. R. Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. *Cell* **15**, 1199–1208 (1978).
- [78] Grant, S. G., Jessee, J., Bloom, F. R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 4645–4649 (1990).
- [79] Bowers, L. M., Lapoint, K., Anthony, L., Pluciennik, A. & Filutowicz, M. Bacterial expression system with tightly regulated gene expression and plasmid copy number. *Gene* **340**, 11–18 (2004).
- [80] Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6640–6645 (2000).
- [81] Cherepanov, P. P. & Wackernagel, W. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9–14 (1995).
- [82] Choi, K.-H., Gaynor, J. B., White, K. G., Lopez, C., Bosio, C. M., Karkhoff-Schweizer, R. R., & Schweizer, H. P. A Tn7-based broad-range bacterial cloning and expression system. *Nature Methods* **2**, 443–448 (2005).
- [83] Da Re, S., Le Quéré, B., Ghigo, J.-M. & Beloin, C. Tight modulation of Escherichia coli bacterial biofilm formation through controlled expression of adhesion factors. *Applied and Environmental Microbiology* **73**, 3391–3403 (2007).
- [84] Choi, K.-H., Mima, T., Casart, Y., Rholl, D., Kumar, A., Beacham, I. R. & Schweizer, H. P. Genetic tools for select-agent-compliant manipulation of Burkholderia pseudomallei. *Applied and Environmental Microbiology* **74**, 1064–1075 (2008).

- [85] Ferrière, L., Hémery, G., Nham, T., Guèrout, A. M., Mazel, D., Beloin, C. & Ghigo, J. M. Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *Journal of Bacteriology* **192**, 6418–6427 (2010).
- [86] Huang, D., Holtz, W. J. & Maharbiz, M. M. A genetic bistable switch utilizing nonlinear protein degradation. *Journal of Biological Engineering* **6**, 9 (2012).
- [87] Rhodius, V. A., Segall-Shapiro, T. H., Sharon, B. D., Ghodasara, A., Orlova, E., Tabakh, H., Burkhardt, D. H., Clancy, K., Peterson, T. C., Gross, C. A. & Voigt, C. A. Design of orthogonal genetic switches based on a crosstalk map of σ s, anti- σ s, and promoters. *Molecular Systems Biology* **9**, 702 (2013).
- [88] Miertens, N., Remaut, E. & Fiers, W. Increased stability of phage T7g10 mRNA is mediated by either a 5'- or a 3'-terminal stem-loop structure. *Biological Chemistry* **377**, 811–818 (1996).
- [89] Dunn, J. J. & Studier, F. W. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of Molecular Biology* **166**, 477–535 (1983).
- [90] de Boer, H. A., Comstock, L. J. & Vasser, M. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 21–25 (1983).
- [91] Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. 3rd & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345 (2009).
- [92] Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology* **177**, 4121–4130 (1995).
- [93] Southern, J. A., Young, D. F., Heaney, F., Baumgartner, W. K. & Randall, R. E. Identification of an epitope on the P and V proteins of simian virus 5 that distinguishes between two isolates with different biological characteristics. *Journal of General Virology* **72**, 1551–1557 (1991).
- [94] Chen, D. & Arkin, A. P. Sequestration-based bistability enables tuning of the switching boundaries and design of a latch. *Molecular Systems Biology* **8**, 620 (2012).
- [95] Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R. & Tabor, J. J. Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nature Methods* **11**, 449–455 (2014).
- [96] Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Cerretti, D. P., Urdal, D. L. & Conlon, P. J. A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/Technology* **6**, 1204–1210 (1988).
- [97] Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M. & Salis, H. M. Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Molecular Systems Biology* **10**, 731 (2014).
- [98] Miliadis-Argeitis, A., Rullan, M., Aoki, S. K., Buchmann, P. & Khammash, M. Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nature Communications* **7** (2016).
- [99] Datta, S., Costantino, N. & Court, D. L. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109–115 (2006).

- [100] Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F. & Gregor, J. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).