

A cybergenetic framework for engineering intein-mediated integral feedback control systems

Received: 6 September 2022

Accepted: 21 February 2023

Published online: 11 March 2023

 Check for updates

Stanislav Anastassov ^{1,2}, Maurice Filo ^{1,2}, Ching-Hsiang Chang¹ & Mustafa Khammash ¹ 

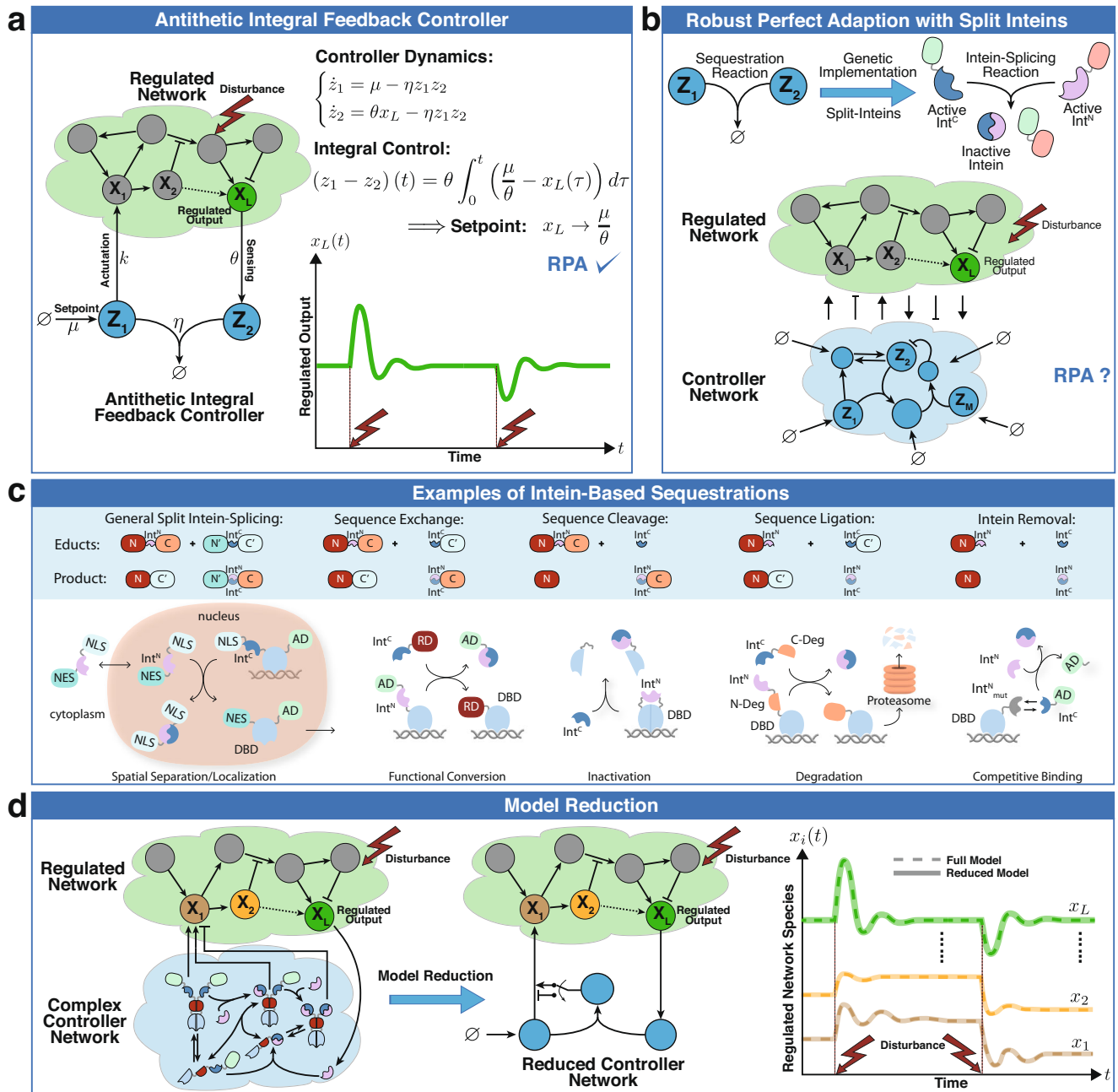
The ability of biological systems to tightly regulate targeted variables, despite external and internal disturbances, is known as Robust Perfect Adaptation (RPA). Achieved frequently through biomolecular integral feedback controllers at the cellular level, RPA has important implications for biotechnology and its various applications. In this study, we identify inteins as a versatile class of genetic components suitable for implementing these controllers and present a systematic approach for their design. We develop a theoretical foundation for screening intein-based RPA-achieving controllers and a simplified approach for modeling them. We then genetically engineer and test intein-based controllers using commonly used transcription factors in mammalian cells and demonstrate their exceptional adaptation properties over a wide dynamic range. The small size, flexibility, and applicability of inteins across life forms allow us to create a diversity of genetic RPA-achieving integral feedback control systems that can be used in various applications, including metabolic engineering and cell-based therapy.

One of the essential features of living systems is their ability to maintain a robust behavior despite disturbances coming from their external uncertain and noisy environments. This feature is referred to as homeostasis, which is typically achieved via endogenous feedback regulatory mechanisms shaped by billions of years of evolution. Pathological diseases are often linked to loss of homeostasis^{1,2}. As a result, restoring homeostasis has become a major focus of research in the emerging field of cybergenetics³, which combines control theory and synthetic biology. In particular, the rational design and implementation of biomolecular feedback controllers^{4–13} offers promising candidates that may accompany or even replace such failed mechanisms^{14–16}.

A notion, which is similar to homeostasis, but more stringent, is Robust Perfect Adaptation (RPA) (see e.g.^{17,18}) which is the biological analogue of the well-known notion of robust steady-state tracking in control theory. A controller succeeds in achieving RPA if it drives the steady state of a variable of interest to a prescribed level despite

varying initial conditions, uncertainties and/or constant disturbances. Motivated by the internal model principle¹⁹, which establishes that the designed controller must implement an integral feedback component to be able to achieve RPA, the antithetic integral feedback (AIF) controller²⁰ was brought forward. The basic antithetic integral feedback motif is depicted in Fig. 1(a). It is comprised of two species Z_1 and Z_2 whose end goal is to robustly steer the concentration of the output species of interest X_L to a prescribed level, referred to as the setpoint, in spite of disturbances and uncertainties in the regulated network – represented here as the various reactions occurring between species X_1 through X_L . RPA is achieved via four controller reaction channels. First, Z_1 is constitutively produced at a rate μ to encode for the setpoint. Second, Z_2 is catalytically produced from the output species X_L at a rate θx_L to sense its concentration. The third reaction is the annihilation or sequestration reaction between Z_1 and Z_2 occurring at a rate $\eta z_1 z_2$. The sequestration reaction encodes a comparison operation and

¹Department of Biosystems Science and Engineering, ETH Zürich 4058 Basel, Switzerland. ²These authors contributed equally: Stanislav Anastassov, Maurice Filo. ✉ e-mail: mustafa.khammash@bsse.ethz.ch



produces an inactive complex that has no function and thus its concentration need not be mathematically tracked. Finally, the feedback control action (actuation) is encrypted as a production reaction of the species X_1 , which acts as the input of the regulated network, at a rate kz_1 proportional to the concentration of controller species Z_1 . The underlying *Ordinary Differential Equations* (ODEs) governing the dynamics of the concentrations of Z_1 and Z_2 are shown in Fig. 1(a). Throughout the paper, bold uppercase letters (e.g. Z_1) denote the names of biochemical species, while their corresponding lowercase letters (e.g. z_1) denote their concentrations. By looking at the dynamics of $z_1 - z_2$, it is straightforward to reveal the integral control action where the temporal error $\mu/\theta - x_L(t)$ at time t , or deviation of the output concentration from the setpoint μ/θ , is mathematically integrated. This establishes that, as long as the closed-loop system is stable (i.e. asymptotically converges to some fixed point), the output concentration x_L will converge to the prescribed setpoint μ/θ which is independent of the regulated network and initial conditions, and thus achieves RPA. While RPA is a steady-

state property, the transient dynamic properties and tuning of the antithetic integral controller are also extensively studied as well^{21–23}.

In vivo antithetic integral feedback controllers have been previously built in both bacteria^{4,6} and mammalian cells⁷, where RPA is experimentally demonstrated. A quasi-integral controller using a slight variant of the antithetic controller was also demonstrated in *Escherichia coli*⁹. More recently, a protein based antithetic controller in mammalian cells was also recently proposed²⁴. In⁶, the controller is implemented in E. coli using sigma/anti-sigma factors as the basic parts that realize the sequestration reaction – the heart of the antithetic integral motif. In⁷, the controller is implemented in HEK293T cells using sense/anti-sense mRNAs. The sequestration reactions in both designs are achieved by the heterodimerization of Z_1 and Z_2 . In the case of sigma/anti-sigma factors, the heterodimerization reaction is reversible, a fact that may lead to reduced performance in certain operating regimes. Moreover, when present in high quantities, xenogenic sigma factors may be toxic due to their inherent property of sequestering RNA polymerases from housekeeping sigma factors²⁵.

Fig. 1 | Overview. **a** The closed-loop system is comprised of the controller network (Z_1, Z_2) connected in a feedback configuration with an arbitrary regulated network. By examining the controller dynamics, it is straightforward to uncover the integral control action that endows the closed-loop system with RPA. That is, as long as the closed-loop system is stable, the concentration of the regulated output X_1 converges to a prescribed value μ/θ , referred to as the setpoint, despite the presence of disturbances and uncertainties in the regulated network. **b** The heart of the basic AIF motif is the sequestration reaction. In this paper, we exploit the exquisite flexibility of split inteins to genetically implement a broad class of integral controllers that endow the closed-loop system with RPA. The flexibility of split inteins offers an easy-to-build biological framework at the price of (potentially) more complex mathematical models. To this end, we establish a set of simple reaction rules that enable RPA. **c** The shaded blue box schematically depicts the products of intein-splicing reactions starting from the educts. The first schematic (top left) describes the general split intein-splicing reaction where both split inteins are flanked by protein domains, labeled (N,N') and (C,C') for the N- and C-terminal protein sequences, respectively. The first product is a new protein containing the N and C' domains of the educts, while the second product is a heterodimer containing N' and C, which are held together by the two inactive split inteins (see Supplementary Information, Fig. 22). The remaining four schematics in the shaded blue box are instantiations of the general case and are labeled according to the perspective of the protein containing the Int^N segment. As the labels suggest, a part of the protein sequence is either exchanged for another one or removed through cleavage. Furthermore, it is possible to ligate another sequence to the protein of interest or make it non-responsive to future splicing reactions through intein removal. To illustrate the design modularity and flexibility, we list a selection of

intein-based implementation examples of the antithetic sequestration motif (below the shaded blue box) based on the described possible splicing reactions. In the first example (bottom left), inteins are used to shuffle proteins between the nucleus and the cytoplasm due to the NLS and the NES which flank the protein sequences. In particular, the intein-splicing reaction exchanges the NLS with a NES which leads to the export of a transcription factor (TF) out of the nucleus where it cannot initiate transcription anymore. In the second example, inteins are used to exchange an AD by a RD, which inverts the function of the TF. In the third example, a split intein is introduced within a functional domain without disturbing it. The splicing reaction results in the cleavage of the domain rendering it nonfunctional. In the fourth example, a protein is fused to the first split-intein and a split-degradation tag, while the second split-intein is fused to the other half of the split-degradation tag. The splicing reaction re-ligates the degradation tag rendering it functional and capable of degrading a POI. In the final example, a DBD can reversibly heterodimerize with an AD via its split inteins. Note that the split intein on the DBD is mutated so that it cannot perform the splicing reaction upon dimerization. A separate non-mutated intein is able to remove the intein from the AD through splicing. This renders the AD unable to heterodimerize with the DBD. AD: activation domain, RD: repressing domain, Int^C: intein C, Int^N: intein N, DBD: DNA binding domain, NLS: nuclear localization signal, NES: nuclear export signal, N-Deg: N terminus of split degradation domain, C-Deg: C terminus of split degradation domain. **d** A simple recipe is developed to reduce the otherwise mathematically complex controller models into simple motifs that resemble the basic AIF motif, but with a fundamental difference: the sequestration product is allowed to have a function that can be leveraged as a tuning knob to enhance the controller performance while maintaining RPA.

Finally, sigma factors are specific to the transcription mechanism of bacteria and cannot be easily transferred to other domains of life, which is why the sense/anti-sense RNA controller⁷ was developed for mammalian cells. At the same time, sense/anti-sense hybridizations produce double-stranded RNAs that, in high abundance, may initiate global translational repression²⁶, leading to a reduction in the effective dynamic range of operation. These constraints give rise to the need for genetic parts that are nontoxic, are transferable between different forms of life, and enjoy wider dynamic ranges of operation. Nevertheless, the suitable choice of genetic parts is a difficult task because they need to adhere to the strict design rules of the basic antithetic integral feedback motif. In this paper, we show that split inteins serve as the ideal candidate parts that are capable of doing both: adhering to the design rules and avoiding the aforementioned disadvantages. We build on the universality result of the antithetic motif⁶ to examine more complex integral controller designs for RPA as demonstrated in Fig. 1(b). While more complex mathematical topologies do not have to be necessarily more difficult to implement, they certainly broaden the biological design space. This expansion, in fact, becomes necessary due to the biological implementation constraints.

An intein is a protein segment that is capable of autocatalytically excising itself from the protein while re-ligating the remaining segments, called exteins, via forming new peptide bonds²⁷ (see Supplementary Information, Fig. 22). Inteins are universal as they can be naturally found in all domains of life spanning eukaryotes, bacteria, archaea and viruses^{28,29}. Split inteins - a subset class of inteins - are, as the name suggests, inteins split into two halves commonly referred to as Int^N and Int^C. Split inteins have been widely studied and characterized due to their extensive usage in various life science disciplines and their ability to perform fast, reliable and irreversible post-translational modifications³⁰⁻³⁴. Small split inteins like Gp41-1^C are comprised of around 40 amino acids³⁵ and are well within the size range of synthetic protein linkers³⁶. It is then possible to use them as “functional” linkers to connect different protein segments. The split inteins, when active, are capable of heterodimerizing and performing protein splicing reactions on their own where they irreversibly break and form new peptide bonds in a strict stoichiometric ratio of one to one. We shall refer to these reactions as “intein-splicing reactions” where molecules containing active Int^C segments react with molecules containing active

Int^N segments to undergo a particular splicing mechanism. When two molecules undergo an intein-splicing reaction, the Int^N and Int^C segments are permanently inactivated as they are unable to perform further splicing reactions due to the alteration of their respective biochemical structures. However the products of such a reaction may still have other functions such as activating or repressing gene expression due to the presence of other protein domains that may not be affected by the splicing reaction. Split inteins can be exploited to exchange, cleave or ligate amino acid sequences (see Fig. 1(c)). These features serve as the basis of realizing the sequestration reaction of the antithetic integral motif. A selection of antithetic “sequestrations” based on functional conversion, spatial separation, inactivation, degradation and intein removal are shown in Fig. 1(c) to emphasize the modularity and the vast flexibility of intein-based designs. Nonetheless, this high design flexibility comes with a price: simple intein-based implementations may lead to complicated network topologies very quickly as illustrated in Fig. 1(d). Here we exploit a time-scale separation argument to establish a structural model reduction result which provides an easy-to-use recipe to simplify the underlying models. This facilitates the mathematical analysis of the otherwise complicated controller network, and allows us to uncover the underlying controller structure which is not necessarily limited to integral control only.

Integral control is the fundamental building block in most controllers spanning a broad range of industrial applications in the fields of electrical, mechanical and chemical engineering; however, it is rarely used alone. In fact, Integral (I) controllers are typically augmented with Proportional (P) and/or Derivative (D) controllers to obtain PI/PID controllers that offer more flexibility in enhancing the dynamic performance while maintaining the RPA property. Recently, more advanced molecular controllers such as PI/PID controllers found their way to molecular biology^{7,37-42}. Ideally, pure proportional control is achieved via instantaneous negative feedback from the output X_1 to the input species X_1 and it is shown that it is not only capable of enhancing the transient dynamic performance, but also reducing cell-to-cell variability^{37,40}. The first biomolecular (filtered) PI controller was genetically engineered in⁷ where additional genetic parts are appended to the antithetic integral motif to realize the proportional component. Here, we establish that a filtered PI controller can be built

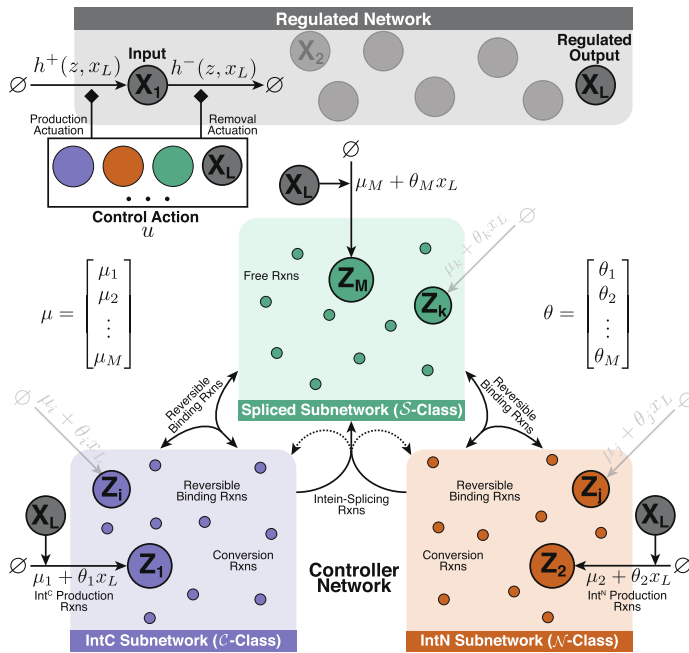


Fig. 2 | A theoretical framework for RPA-achieving intein-based integral controllers. The closed-loop network is formed of a controller network, comprised of M species Z_1, \dots, Z_M , connected in a feedback configuration with the regulated network, comprised of L species X_1, \dots, X_L . Following the general biomolecular control paradigm³⁷, it is assumed that the controller interacts with the regulated network via X_1 and X_L only, referred to as the input and regulated output species, respectively. The objective of the controller network is to steer the concentration of the regulated output X_L to a prescribed value, referred to as the setpoint, despite the presence of constant disturbances and uncertainties in the regulated network. The controller network is divided into three subnetwork classes

Species Rules	
1.	All controller species containing at least one active Int ^C belong to the \mathcal{C} -class and cannot contain active Int ^N .
2.	All controller species containing at least one active Int ^N belong to the \mathcal{N} -class and cannot contain active Int ^C .
3.	All controller species that do not contain any active inteins belong to the \mathcal{S} -class.
Reaction Rules	
1.	Productions of controller species are either constitutive or driven by the regulated output molecule X_L or both ε .
2.	Controller species in the \mathcal{C} -class can undergo an intein-splicing reaction with controller species in the \mathcal{N} -class to produce at least one product that belongs to the \mathcal{S} -class. At least one such reaction exists.
3.	The input species X_1 is either produced or degraded or both by at least one of the controller species.
4.	Within each class, controller species are allowed to convert to one another while conserving the number of active Int ^C /Int ^N segments.
5.	Within each class, controller species are allowed to reversibly bind to one another while conserving the number of active Int ^C /Int ^N segments.
6.	Controller species belonging to the \mathcal{C} -class are allowed to reversibly bind with controller species belonging to the \mathcal{S} -class to produce another controller species belonging to the \mathcal{C} -class while conserving the number of active Int ^C segments.
7.	Controller species belonging to the \mathcal{N} -class are allowed to reversibly bind with controller species belonging to the \mathcal{S} -class to produce another controller species belonging to the \mathcal{N} -class while conserving the number of active Int ^N segments.
8.	Controller species belonging to the \mathcal{S} -class are allowed to be produced, degraded, or freely react among each other to yield controller species within the \mathcal{S} -class.

according to the list of Species Rules. The allowed reactions within and between the three subnetworks are listed as Reaction Rules. The feedback controller network operates by “sensing” the abundance of the concentration of the regulated output $X_L(\theta, x_L)$, and “actuating” the input X_1 by producing it ($h^+(z, x_L)$) or removing it ($h^-(z, x_L)$). The total control action u is given by $h^+(z, x_L) - h^-(z, x_L)x_1$. Note that, throughout the paper, the diamond-shaped arrowhead denotes either an activation or repression. The setpoint and output-sensing mechanisms are jointly encoded in the vectors μ and θ to allow for multiple setpoint/sensing-encoding reactions.

without introducing additional genetic parts by harnessing the sequestration products of the split inteins³⁸.

Besides proposing intein-based implementation strategies for RPA-achieving controllers and laying down the necessary theoretical foundation, we have also selected, built and tested five structurally different controller topologies for experimental verification of RPA. All circuits were tested in HEK293T cells and range from pure I to filtered PI controllers based on the functional conversion, inactivation and intein removal strategies illustrated in Fig. 1(c).

Split inteins offer a high degree of flexibility in realizing biomolecular integral feedback controllers. This flexibility is mainly a consequence of their compatibility with essentially any transcription factor (TF). In fact, the particular structure of the expressed transcription factor including the choices of the Activation Domain (AD), Dimerization Domain (DD), DNA-Binding Domain (DBD) and insertion position of the split intein (Int^C) open the possibilities to a broad design space of controllers. Specifically, dimeric transcription factors, such as tetracycline transactivator (tTA), give rise to multiple homo- and hetero-dimerization reactions as well as multiple sequestration reactions and thus make the controller network more complex to mathematically analyze. To this end, we develop a theoretical framework tailored to mathematically analyze and simplify complex intein-based controller networks that generalize the basic antithetic integral motif which has no dimerization reactions and a single sequestration reaction. We refer the reader to Supplementary Table 1 for a list of all the abbreviations used in this paper.

Results

Achieving robust perfect adaptation using inteins

In this section, we establish a theoretical framework embodied as a set of simple rules that allows us to design biomolecular controllers

enabling RPA using split inteins. Consider the general closed-loop network depicted in Fig. 2 where an arbitrary network comprised of L species $X = \{X_1, \dots, X_L\}$, referred to as the regulated network, is in a feedback interconnection with the controller network comprised of M species $Z := \{Z_1, \dots, Z_M\}$. The overall objective of the feedback controller network is to achieve RPA of the regulated output species X_L by *automatically* actuating (producing and/or degrading) the input species X_1 . Each controller species Z_i , for $i=1, \dots, M$, belongs to one of three classes: \mathcal{C} -class, \mathcal{N} -class and \mathcal{S} -class. These classes separate the controller network into three subnetworks as depicted in Fig. 2. The classification of the controller species and the allowed reactions follow the rules that are listed in Fig. 2. In particular, the setpoint and sensing of the regulated output species X_L are encoded in the constitutive and/or catalytic production reactions following Reaction Rule 1 given by

$$\text{Setpoint/Sensing : } \varnothing \xrightarrow[\text{Rxns}]{\mu_i + \theta_i x_L} Z_i \quad (i=1, \dots, M), \quad (1)$$

with at least one μ_i and one θ_i strictly positive. The following theorem provides a guarantee for RPA of the regulated output species when controlled with intein-based controllers.

Theorem 1. Consider the closed-loop network depicted in Fig. 2 where the controller network respects the set of listed rules. Let q_i^+ and q_i^- respectively denote the number of active Int^C and Int^N segments present in controller species Z_i for $i=1, \dots, M$. Define the vector $q := [q_1^+ - q_1^- \quad \dots \quad q_M^+ - q_M^-]^T$. Then, if the closed-loop network is

stable, the controller network ensures RPA of \mathbf{X}_L with

$$\lim_{t \rightarrow \infty} x_L(t) = -\frac{q^T \mu}{q^T \theta} > 0, \quad (2)$$

where $\mu := [\mu_1 \ \dots \ \mu_M]^T$ and $\theta := [\theta_1 \ \dots \ \theta_M]^T$. Furthermore, the integrated variable is given by $z_I := q^T z$ which reveals the underlying integral controller given by

$$z_I(t) = z_I(0) + \int_0^t [\mu^{\text{eff}} - \theta^{\text{eff}} x_L(\tau)] d\tau, \quad (3)$$

with $\mu^{\text{eff}} := q^T \mu$ and $\theta^{\text{eff}} := -q^T \theta$.

The proof of Theorem 1 can be found in Supplementary Information Section 2. Before we proceed, we provide two remarks.

Remark 1.1. Theorem 1 is a special case of a more general theorem (see Supplementary Theorem 1) which can be also applied to any non-intein-based biomolecular controller with similar structure as demonstrated in the example of Box 3. This more general theorem interprets q^+ and q^- as the number of positive and negative charges (where, here, the number of inteins is an instantiation of the charge analogy) and extends the RPA sufficiency result in⁶ to the case of multiple sensing and setpoint reactions. In fact, if \mathbf{Z}_1 is the only controller species that is constitutively produced and \mathbf{Z}_2 is the only controller species that is catalytically produced by the regulated output species \mathbf{X}_L , then $\mu_i, \theta_2 > 0, \mu_i = \theta_j = 0$ for $(i,j) \neq (1,2)$ and $q = [1 \ -1 \ \dots \ *]^T$ which yields $\lim_{t \rightarrow \infty} x_L(t) = \mu_1/\theta_2$ – the RPA result in⁶.

Remark 1.2. Although the result presented in Theorem 1 is for the deterministic setting, it also holds in the stochastic setting. It is shown in Supplementary Information Section 2 that, under the assumption that the closed-loop network is ergodic – a stochastic notion of stability, the steady-state (stationary) expectation of the regulated output is also given by $\mathbb{E}_\pi[X_L] = -\frac{q^T \mu}{q^T \theta}$.

Remark 1.3. The catalytic sensing terms $\theta_i x_L$ for $i=1, \dots, M$ shown in Fig. 2 and (1) do not necessarily have to be linear in the deterministic setting. In fact, these terms can be replaced by more general nonlinear functions $f_i(x_L)$ such as Hill-type functions that are allowed to be monotonically decreasing to account for repressive sensing. These sensing mechanisms will preserve RPA in the deterministic setting, but the setpoint expression will be different from (2).

Implementations using various transcription factors

So far, we have described, theoretically, how split inteins can be exploited to build a broad class of biomolecular integral controllers capable of achieving RPA. Here, we demonstrate how commonly used transcription factors can be converted into controller species that respect the rules of Fig. 2 and, as a result, enables RPA according to Theorem 1. In particular, we use three common DNA binding domains: zinc finger (ZF)^{43,44}, tetracycline repressor (TetR)⁴⁵ and Gal4 to construct four structurally different biomolecular controllers (Fig. 3), that serve as instantiations of the class of controllers described in Theorem 1. We also provide experimental proof (Fig. 4) that these intein-based controllers are indeed capable of achieving RPA and thus rejecting perturbations over a wide dynamic range.

There are a few considerations that have to be taken into account to successfully build intein-based integral controllers.

These considerations, given in Box 1, should be experimentally verified for any intein-based controller to function properly. In particular, to minimize the impairment of the protein of interest as per Building Consideration 1, we use the smaller Int^C of the fast reacting intein Gp41-1⁴⁶ for all our modified activators.

Next, we provide a detailed description of the four different controller circuits depicted in Fig. 3. We start with the ZF controller which has the simplest topological structure. It is obtained by using ZF as the DBD, and introducing the split intein in the floppy linker between the AD VP64 and the DBD ZF. This TF, denoted by \mathbf{Z}_1 , is constitutively produced at a rate μ_1 and is capable of actuating the regulated network of interest by activating the expression of the input \mathbf{X}_1 . The regulated output \mathbf{X}_L produces the second split intein Int^N, denoted by \mathbf{Z}_2 , at a rate θx_L that is proportional to the regulated output concentration. The intein-splicing reaction between \mathbf{Z}_1 and \mathbf{Z}_2 occurs at a rate η and leads to a cleavage within the TF, which separates the AD from the DBD. The resulting free floating AD is not tracked due to its inability to initiate transcription on its own. The other spliced product is the DBD, denoted by \mathbf{Z}_3 , which competes with \mathbf{Z}_1 for the promoter binding sites, and thus exerts a repressive actuation.

The second controller design, labeled as intra dimerization domain (intraDD), is based on TetR whose goal is to illustrate that it is possible to build intein-based antithetic integral controllers without functional spliced products. This controller is obtained by introducing the split intein within the DD of TetR without disrupting it. The transcription factor \mathbf{Z}_1 is generated by fusing VPR to the modified TetR. The dimer \mathbf{Z}_4 comprised of two molecules of \mathbf{Z}_1 acts as the actuating controller species. Unlike the previous controller, Int^N, denoted by \mathbf{Z}_2 , can now undergo an intein-splicing reaction with either the monomer \mathbf{Z}_1 or the dimer \mathbf{Z}_4 . The intein-splicing reaction with \mathbf{Z}_1 leads to the cleavage of the protein sequence next to the Int^C, which is acting as a linker holding the two halves of the split DD together. This results in two products: the AD VPR with part of the disabled DD, denoted by \mathbf{Z}_3 , and a monomeric TetR with the rest of the disabled DD (not tracked in Fig. 3 due to its inactivity). Neither of them are able to further interact with the controller or the regulated network. Similarly, the intein-splicing reaction with the dimer \mathbf{Z}_4 leads to the cleavage of one of the monomers within the DD. This results in the immediate falling apart of the dimer into one \mathbf{Z}_1 and one \mathbf{Z}_3 .

The third controller design is obtained by inserting an Int^C segment between TetR and the AD. The expressed TF, denoted by \mathbf{Z}_1 , has to dimerize to form \mathbf{Z}_5 to be able to actuate the regulated network. Int^N, denoted by \mathbf{Z}_2 , can undergo an intein-splicing reaction with either the monomer \mathbf{Z}_1 or the dimer \mathbf{Z}_5 . The intein-splicing reaction between \mathbf{Z}_1 and \mathbf{Z}_2 leads to the separation of the AD from the remaining DBD and DD to produce \mathbf{Z}_3 and a free floating AD which is not tracked anymore due to its inactivity. The spliced product \mathbf{Z}_3 can still heterodimerize with \mathbf{Z}_1 to yield a TetR dimer with only one AD, denoted by \mathbf{Z}_6 , which is sufficient to bind to the promoter and initiate transcription. Note that \mathbf{Z}_6 can be also obtained via the intein-splicing reaction between the fully intact dimer \mathbf{Z}_5 and \mathbf{Z}_2 . Furthermore, since \mathbf{Z}_6 still has one functional Int^C segment, it is able to perform a second intein-splicing reaction with \mathbf{Z}_2 , which removes the last AD by cleavage and hence forms a tetR dimer, denoted by \mathbf{Z}_4 . Note that, this dimer can be also obtained via the homodimerization of \mathbf{Z}_3 . The dimer \mathbf{Z}_4 can recognize and bind to the promoter, but can not initiate transcription unlike the other two dimers \mathbf{Z}_5 and \mathbf{Z}_6 . It therefore, competes with \mathbf{Z}_5 and \mathbf{Z}_6 for the promoter binding sites and, as a result, acts as a repressor.

The last controller design of Fig. 3 is based on the yeast derived DBD Gal4 and is thus labeled as the Gal4 controller. Here, we introduced an Int^C segment between the DBD and the DD. Similar to TetR, Gal4 needs to be a dimer (\mathbf{Z}_5) in order to bind to the promoter and actuate the regulated network. Once again, Int^N, denote by \mathbf{Z}_2 , can undergo an intein-splicing reaction with either \mathbf{Z}_5 or \mathbf{Z}_1 . The intein-splicing reaction with \mathbf{Z}_1 leads to the separation of the DBD from the remaining DD and the AD to produce \mathbf{Z}_3 . As already mentioned, Gal4 cannot bind to the promoter as a monomer, and so we do not track this species due to its inactivity. Furthermore, the intein-splicing reaction with \mathbf{Z}_5 leads to the removal of one DBD from the dimer through cleavage, which renders the entire complex unable of binding to the

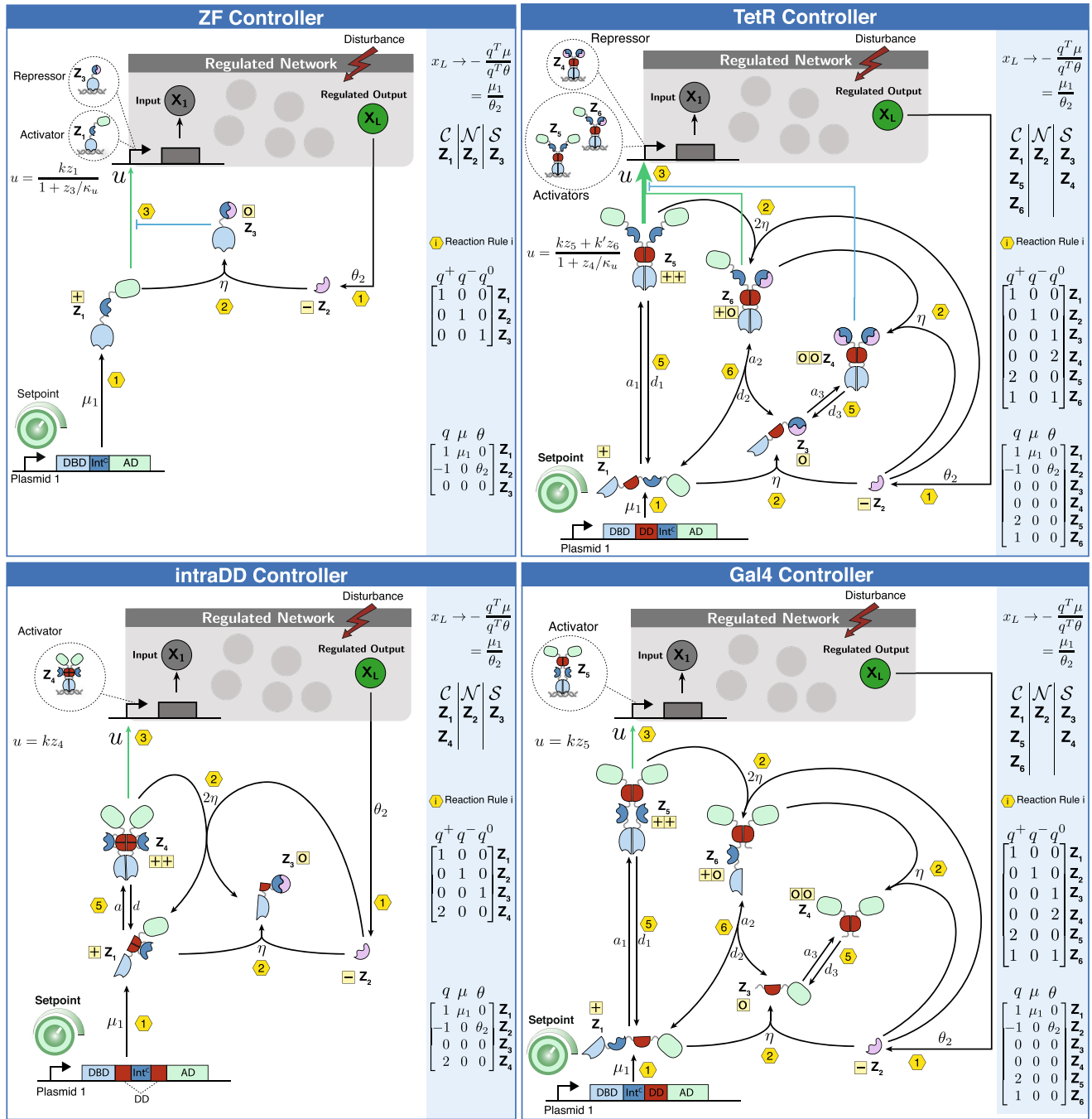


Fig. 3 | InTEIN-based implementation of RPA-achieving integral controllers using ZF, TetR and Gal4 as DBDs. In all four circuits, the protein Z_1 is constitutively expressed at a rate μ_1 from Plasmid 1 to encode for the desired setpoint. One of the two main tasks of Z_1 , which contains one Int^C within a TF, is to either directly actuate the regulated network by producing the input species X_1 , or to dimerize first and then actuate. The second task of Z_1 is to undergo an inTEIN-splicing reaction with the second split inTEIN Int^N , denoted by Z_2 , whose production is driven by the regulated output X_L at a rate $\theta_2 X_L$ to encode for the “sensing” reaction. Different positions of the Int^C segment and different TF structures yield different control topologies. Controller species containing DDs undergo reversible homo- or hetero-dimerization reactions with association and dissociation rates of a_i and d_i ; whereas, controller species containing active Int^C and Int^N segments undergo irreversible inTEIN-splicing reactions with rate η multiplied by an integer that depends on the

number of participating inTEINs. Note that inactive splicing products are omitted here for simplicity. The control action u is mathematically expressed as a (Hill-type) function of the repressors and activators depicted in the dashed bubbles. Every reaction is labeled from 1 to 6 according to the permitted reaction rules stated in Fig. 2. Furthermore, every monomer, independent of its dimerization status, is labeled in the yellow boxes with one of the following “charges”: +, -, 0, according to Theorems 1 and 2. This is also repeated in the charge vectors q^+ , q^- and q^0 that encode, for each controller species, the number of active Int^C , Int^N , and monomers with no active inTEINs. Furthermore, in the blue boxes, all controller species are grouped into C , N and S classes according to the split inTEINs they contain (see Species Rules in Fig. 2). Since all the Species and Reaction Rules of Fig. 2 are respected, then by Theorem 1, we conclude that all four controllers ensure RPA with a setpoint of μ_1/θ_2 .

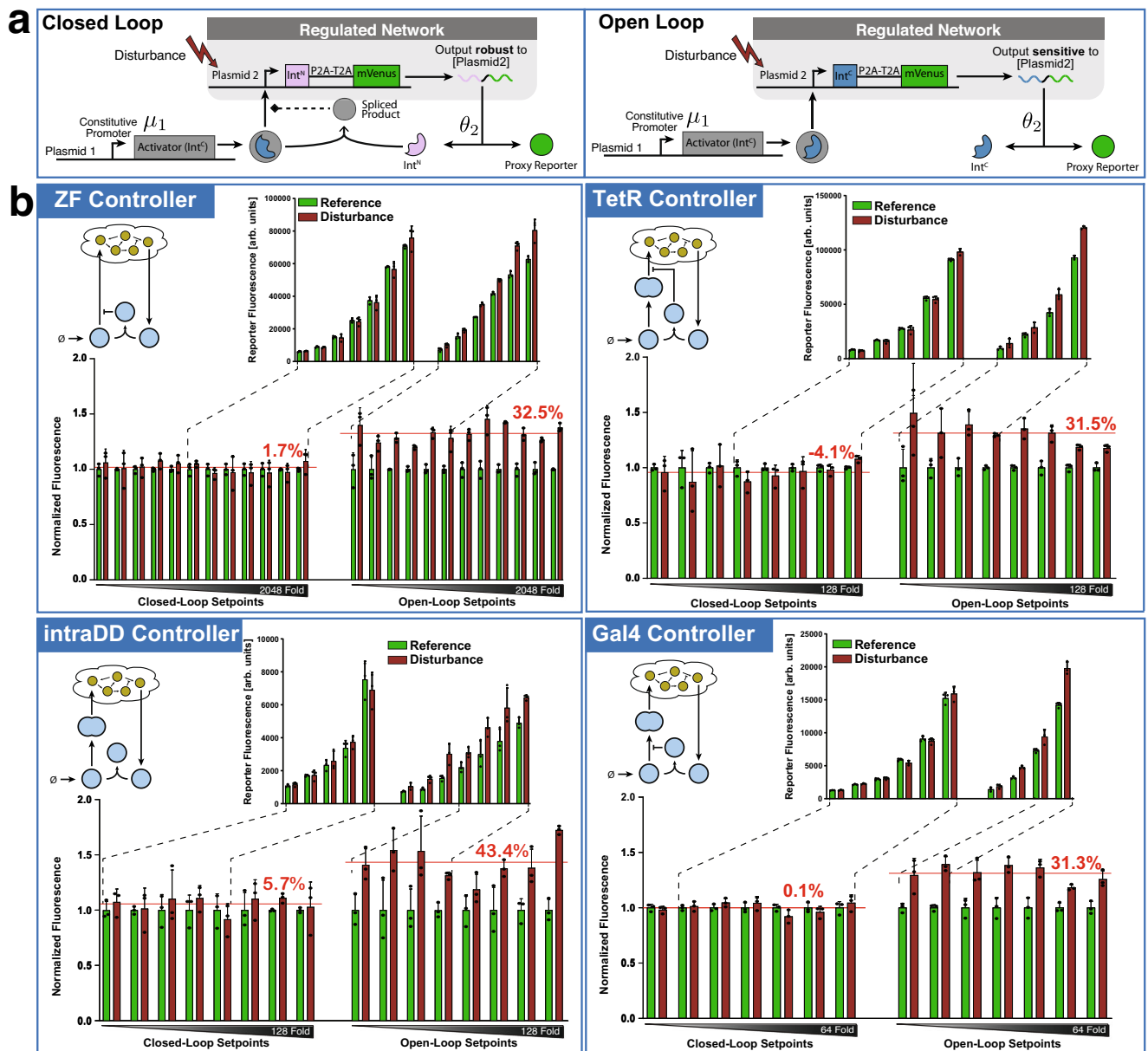


Fig. 4 | Experimental demonstration of RPA. **a** Illustration of the experimental setup. All four controllers of Fig. 3 were tested in a two-plasmid system. Plasmid 1 encodes an Int^C segment incorporated in an activator driven by a constitutive promoter. Plasmid 2 encodes Int^N (resp. Int^C) in the closed-loop (resp. open-loop) setting, which is fused to the fluorophore mVenus via a P2A-T2A linker and is driven by the activator expressed from Plasmid 1. The fluorescent protein mVenus is used as a proxy reporter for its own mRNA which is the regulated output expected to exhibit RPA in closed loop. In open loop, there is no interaction between the two Int^C segments; whereas, in closed loop, the inteins pair can perform the intein-splicing reaction to produce (possibly) functional products. The setpoint is tuned by changing the constitutive production rate μ_1 via the transfected copy numbers of Plasmid 1. The reference, or undisturbed output level, is obtained by fixing the copy numbers of the transfected Plasmid 2 across all setpoints; whereas, the disturbed output level is obtained by repeating the same experiment for all setpoints

but with a higher copy number of Plasmid 2. **b** Steady-state errors. For each controller, we show a simplified schematic (top left) and two bar graphs. The bottom bar graphs show the normalized fluorescence of the proxy reporter with (red) and without (green) disturbance and for the closed-loop (left) and open-loop (right) settings. The disturbed and reference triplicate measurements were normalized to the mean fluorescence of the reference data for each setpoint. The x-axis follows a log₂-scale and shows the amount of Plasmid 1 transfected within every well. The red horizontal lines give the normalized output averaged over all the setpoints, and the numbers above the lines indicate the averaged error of the disturbed output relative to the reference. The top bar graphs show the non-normalized data over a selected subset of setpoints (pointed out by the dashed lines) that match in absolute fluorescence between the open- and the closed-loop circuits. For all the data, the HEK293T cells were measured using flow cytometry 48 h after transfection, and the normalized data are shown as mean + SD for $n = 3$ technical replicates.

DNA. This truncated dimer, denoted by Z_6 , can perform a second intein splicing reaction with Z_2 to remove the second DBD and form a new dimer denoted by Z_4 which is also incapable of acting directly on the regulated network. However, it is able to disassociate into its monomers, Z_3 , which are able to reversibly sequester Z_1 through a heterodimerization reaction yielding the non-functional dimer Z_6 .

It is fairly straight forward to verify that all the reaction rules listed in Fig. 2 are respected by all of the proposed four controllers. As a result, by applying Theorem 1, we conclude that all four proposed controllers achieve RPA (as long as the closed-loop network is stable) such that the concentration of the regulated output x_I converges to μ_1/θ_2 at steady state. Next, we provide an experimental verification to back up our

BOX 1**Building considerations**

1. Fusion of the split intein to or into a protein of interest should not completely impair the function of the target.
2. The split intein has to be able to fold properly to efficiently perform the intein-splicing reaction.
3. The split intein has to be sterically accessible by its other split intein to be capable of efficiently performing the intein-splicing reaction.

developed theory. To do so, all of the four proposed controller circuits were first tested for the three Building Considerations. In fact, to test for Building Consideration 1, we expressed all of the modified activators constitutively and compared their ability to transcribe a fluorophore. We observed a drop in activity for all modified ZF, tetR and GAL4 based TFs ranging from significant to minor (see Supplementary Information, Fig. 24). To this end, strong impairments were partially compensated by using stronger activation domains like VPR. Intein insertions within floppy linkers were relatively straight forward; however, insertions within functional protein domains, as was the case for the intraDD-Circuit (see Fig. 3), required some screening (see Supplementary Information, Fig. 23). Next we tested for Building Considerations 2 and 3 by constitutively expressing the modified activator carrying Int^C together with the second split intein (Int^N) and observed the levels of a fluorescent reporter. If the Building Considerations are satisfied the fluorescent output will decrease with increased levels of Int^N. We were able to reach background levels for every controller type upon a high expression of the second split intein (see Supplementary Information, Fig. 25). This indicates that the intein-splicing reaction is indeed happening as expected.

After making sure that all Building Considerations were fulfilled, we proceeded with characterizing the controllers in the closed-loop setting. We opted for a simple two-plasmid, closed-loop system for testing the controller performance as demonstrated in Fig. 4(a). This allowed us to focus on the controller behavior without having to worry about potential cross-talks⁴⁷, resource burden⁴⁸ or saturation⁴⁹ which might appear in larger circuits. The first plasmid encodes for the modified transcription factor Z_1 and the other one encodes for either Int^C for the open-loop circuit or Int^N for the closed-loop circuit. In both cases, the split intein was encoded with a P2A-T2A linker and the fluorophore mVenus. Note that the P2A-T2A linker leads to the translation of two separate proteins (Int^N and mVenus) in a fixed ratio from a common mRNA due to ribosome skipping⁵⁰. The fluorophore is used as a proxy for its own mRNA, which is the regulated species expected to exhibit RPA. The advantage of this setup is that changing the copy numbers of the two transfected plasmids can be conveniently used to characterize the controllers. More precisely, μ_1 and hence the setpoint can be easily tuned by altering the amount of the plasmid encoding for the activator. Furthermore, the translation rate θ_2 of the mRNA is independent from the plasmid copy numbers in the cell. Perturbing the copy numbers of plasmid 2 only leads to an increase in the transcription rate of the output mRNA and should be rejected if the integral controller works as expected. Hence, to experimentally test the four controllers for RPA, we perturb the regulated network by increasing the copy number of plasmid 2 as it does not affect the setpoint parameters μ_1 and θ_2 . The experimental results, depicted in Fig. 4, detail the steady-state measurements of the reporter, serving as a proxy for the regulated output (mRNA) for all four controllers. The measurements were taken for all the circuits operating in both open and closed loop, with and without disturbance. All four circuits were able to reject the disturbance over a wide titration of plasmid 1, which

defines the output setpoint through tuning μ_1 . The best performance was observed with the ZF circuit, which succeeded in rejecting the disturbances over the entire range from the detection limit to the onset of burden (see Supplementary Information, Fig. 27).

We have used so far only the split inteins of Gp41-1, and we have successfully shown the implementation of intein-based RPA-achieving integral controllers using different TFs. Many split intein pairs with different properties have been described in literature with some of them being orthogonal to each other⁵¹. To demonstrate that intein-based integral controllers are not limited to Gp41-1, and that it is possible to have multiple orthogonal intein-based integral controllers within the same cell, we have modified our ZF controller accordingly. In particular, we exchanged the Gp41-1, for NrdJ-1 Int^C, one of the many orthogonal split inteins characterized by Pinto et al.⁵¹ and closed the loop with the corresponding Int^N of NrdJ-1. However, instead of using Int^C for the open-loop circuit, we used the Int^N that corresponds to Gp41-1. Finally, we performed the experiment with the same plasmid ratios, which was deemed suitable for the previous Gp41-1 ZF experiment. The disturbance rejection was only visible for the compatible intein pair, and the dynamic range was similar to the experiment performed with the Gp41-1 containing ZF (see Supplementary Information, Fig. 26).

Model reduction

The broad class of intein-based, RPA-achieving controllers introduced in Theorem 1 gives rise to a high degree of design flexibility and thus allows topologies that may possibly involve a large number of controller species Z_i . Furthermore, these species are allowed to react among each other via multiple binding, conversion and intein-splicing reactions according to the Reaction Rules listed in Fig. 2. This possible large number of control species and reactions may lead to complex mathematical models of high dimensions whose dynamics are not easy to understand. In this section, we consider a subset of the general RPA-achieving controllers of Theorem 1 to provide a model reduction result that makes the otherwise complex dynamics more transparent and easy to analyze. Our model reduction result is structural in the sense that its validity is independent of the particular values of the rate parameters.

Consider the Species and Reaction Rules of Fig. 2 and replace Reaction Rule 8 with five additional rules given in Box 2.

Note that Rule 9 makes Rule 2 stricter in the sense that the intein-splicing reactions are not optional anymore so that any two active intein pairs have a strictly positive propensity to undergo an intein-splicing reaction. Rule 12 takes into account the more realistic situation where $\delta > 0$ which implies that RPA is not exact anymore; however, robust adaptation remains practically satisfactory as long as the dilution rate is small compared to the other rates in the network (see^{6,52}). Finally, Rule 13 relates the intein-splicing rate to the number of participating active inteins. The following theorem provides a recipe for model reduction of (possibly complex) intein-based controllers. The

BOX 2**Additional reaction rules**

9. All controller species in the \mathcal{C} -class undergo an intein-splicing reaction with all controller species in the \mathcal{N} -class.
10. Binding and conversion reactions are reversible and conserve the number of inactive (bound) Int^C-Int^N complexes.
11. All controller species belonging to the \mathcal{S} -class degrade at a rate $\delta_0 \geq 0$.
12. All controller species dilute at a rate $\delta \geq 0$.
13. The intein-splicing rate between Z_i and Z_j is $\eta_i \max(q_i^+, q_j^-)$ with $\eta > 0$.

model reduction result is valid in both the ideal ($\delta = 0$) and non-ideal ($\delta > 0$) settings and for any rate-parameter regimes.

Theorem 2. Consider the closed-loop network depicted in Fig. 2 where the controller network respects Species Rules 1-3 and Reaction Rules 1-7,9-13. Let q_i^+ and q_i^- respectively denote the number of active Int^C and Int^N segments present in controller species Z_i for $i=1, \dots, M$. Let q_i^0 denote the number of monomers in species Z_i with no active inteins, and construct the three vectors $q^k := [q_1^k \dots q_M^k]^T$, for $k \in \{+, -, 0\}$. Furthermore, let (S_B, S_C) and $(\lambda_B(z), \lambda_C(z))$ respectively denote the stoichiometry matrices and total propensity functions associated with the reversible binding and conversion reactions that are assumed to be fast enough. If the following conditions are satisfied:

- S_B is full-column rank.
- The columns of S_C are linearly independent from those of S_B .
- $p + \text{rank}(S_C) = M - 3$,

where p is the number of reversible binding reactions, then all controller networks respecting the structure described in Fig. 2 reduce to the simple motif, depicted in Fig. 5, which is governed by only three effective species Z^+, Z^- and Z^0 whose concentrations are linear combinations of the controller species Z_i for $i=1, \dots, M$.

The proof of Theorem 2 can be found in Supplementary Information Section 2. Before we proceed, we provide five remarks.

Remark 2.1. Once again, Theorem 2 is a special case of a more general theorem (see Supplementary Theorem 2) which can be also applied to any non-intein-based biomolecular controller with similar structure as demonstrated in Box 3. The proof essentially invokes the deficiency-zero theorem⁵³ and singular perturbation theory⁵⁴.

Remark 2.2. The dynamics of the reduced model are depicted in the box of Fig. 5, in general, as a set of *Differential Algebraic Equations* (DAEs) comprised of only three differential equations (describing the basic effective motif) and a set of $M - 3$ algebraic equations that should be solved for $\tilde{z} \geq 0$. In certain cases, these algebraic equations can be explicitly solved and thus further reducing the dynamics to a set of

three ODEs (see Fig. 6). Otherwise, the algebraic equations can be left in their implicit form.

Remark 2.3. Unlike the effective species Z^+ and Z^- , Z^0 has an extra production term, in general, that is equal to $\delta_0 [\mathbb{1}(q^+ + q^-) \circ q^0]^T \psi(z^{\text{tot}})$, where $\mathbb{1}(\cdot)$ is the indicator function, \circ is the Hadamard (elementwise) product and $\psi(z^{\text{tot}})$ is given implicitly in Fig. 5. This production term is zero in two cases: (1) if there are no degradation reactions ($\delta_0 = 0$), or (2) if no controller species simultaneously hold both an active intein and a monomer with no active inteins ($\mathbb{1}(q^+ + q^-) \circ q^0 = 0$). Intuitively, this extra production term can be explained as follows. Controller species holding both an active intein and a monomer with no active inteins belong to either the C- or N-class (Species Rules), and are thus not allowed to degrade (Reaction Rules). Nevertheless, these species are still represented within Z^0 since they hold monomers with no active inteins. As a result, the extra production term compensates for those species that do not degrade yet are represented by Z^0 which degrades at a rate δ_0 .

Remark 2.4. Observe that no matter what the original controller network in Fig. 1 is and as long as it satisfies the conditions of Theorem 2, the underlying effective motif is the same and is dictated by the three effective species Z^+, Z^- and Z^0 as depicted in Fig. 5. However, different controller networks give rise to different actuation functions U^\pm and production functions ψ . The forms of these functions lead to different control designs that may offer different tuning knobs capable of enhancing the overall performance.

Remark 2.5. Unlike Theorem 1, it is unclear whether Theorem 2 can be extended to the stochastic setting. While a mathematically rigorous approach is left for future work, we have conducted a simulation-based case study which revealed that the reduced model was capable of accurately capturing the stochastic dynamics of the full model. See Supplementary Information Section 6 for more details.

Next, we apply Theorem 2 to the four controller circuits of Fig. 3 to obtain a reduced mathematical model for each. Here, we consider

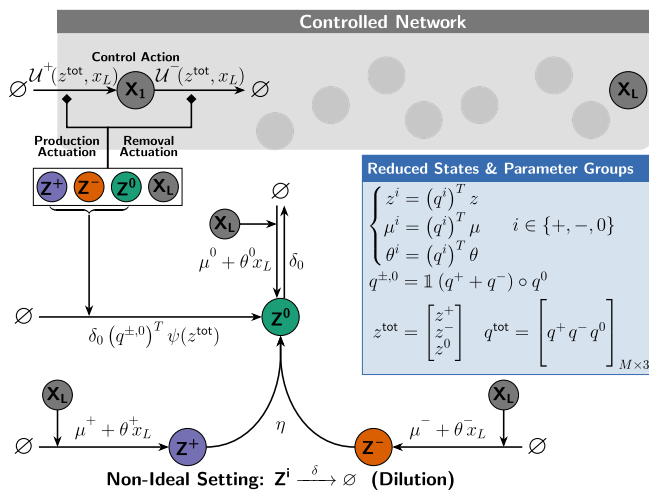


Fig. 5 | A model reduction recipe for Intein-based controllers. Under the conditions of Theorem 2, all controllers comprised of M species (where M can be large) that respect the flexible structure depicted in Fig. 2, reduce to the simple motif shown here. The reduced model is shown schematically as a motif comprised of only three effective species (Z^+, Z^-, Z^0), and mathematically as a set of *Differential Algebraic Equations* (DAEs) comprised of only three differential equations in (z^+, z^-, z^0) and $M - 3$ algebraic equations in \tilde{z} . Note that (S_B, λ_B) and (S_C, λ_C) denote the stoichiometry matrices and total propensity functions (forward minus backward) of the reversible binding and conversion reactions, respectively. Furthermore

Reduced Controller Dynamics

State Dynamics:
$$\begin{cases} \dot{z}^+ = \mu^+ + \theta^+ x_L - \eta z^+ z^- - \delta z^+ \\ \dot{z}^- = \mu^- + \theta^- x_L - \eta z^+ z^- - \delta z^- \\ \dot{z}^0 = \mu^0 + \theta^0 x_L + \eta z^+ z^- - (\delta + \delta_0) z^0 + \delta_0 (q^{\pm,0})^T \psi(z^{\text{tot}}) \end{cases}$$

Control Action:
$$U^\pm(z^{\text{tot}}, x_L) = h^\pm(\psi(z^{\text{tot}}), x_L)$$

$$\psi(z^{\text{tot}}) = W_1^T z^{\text{tot}} + W_2^T \tilde{z}$$

where $\tilde{z} \geq 0$ solves:

$$\tilde{C} S_B \lambda_B (W_1^T z^{\text{tot}} + W_2^T \tilde{z}) + \tilde{C} S_C \lambda_C (W_1^T z^{\text{tot}} + W_2^T \tilde{z}) = 0,$$

with $I_M = \begin{bmatrix} \tilde{C} \\ \tilde{C} \end{bmatrix} \rightarrow \begin{matrix} 3 \times M \\ (M-3) \times M \end{matrix}$ and $\begin{cases} W_1 := (\tilde{C} q^{\text{tot}})^{-1} \tilde{C} \\ W_2 := \tilde{C} (I_M - q^{\text{tot}} (\tilde{C} q^{\text{tot}})^{-1} \tilde{C}) \end{cases}$

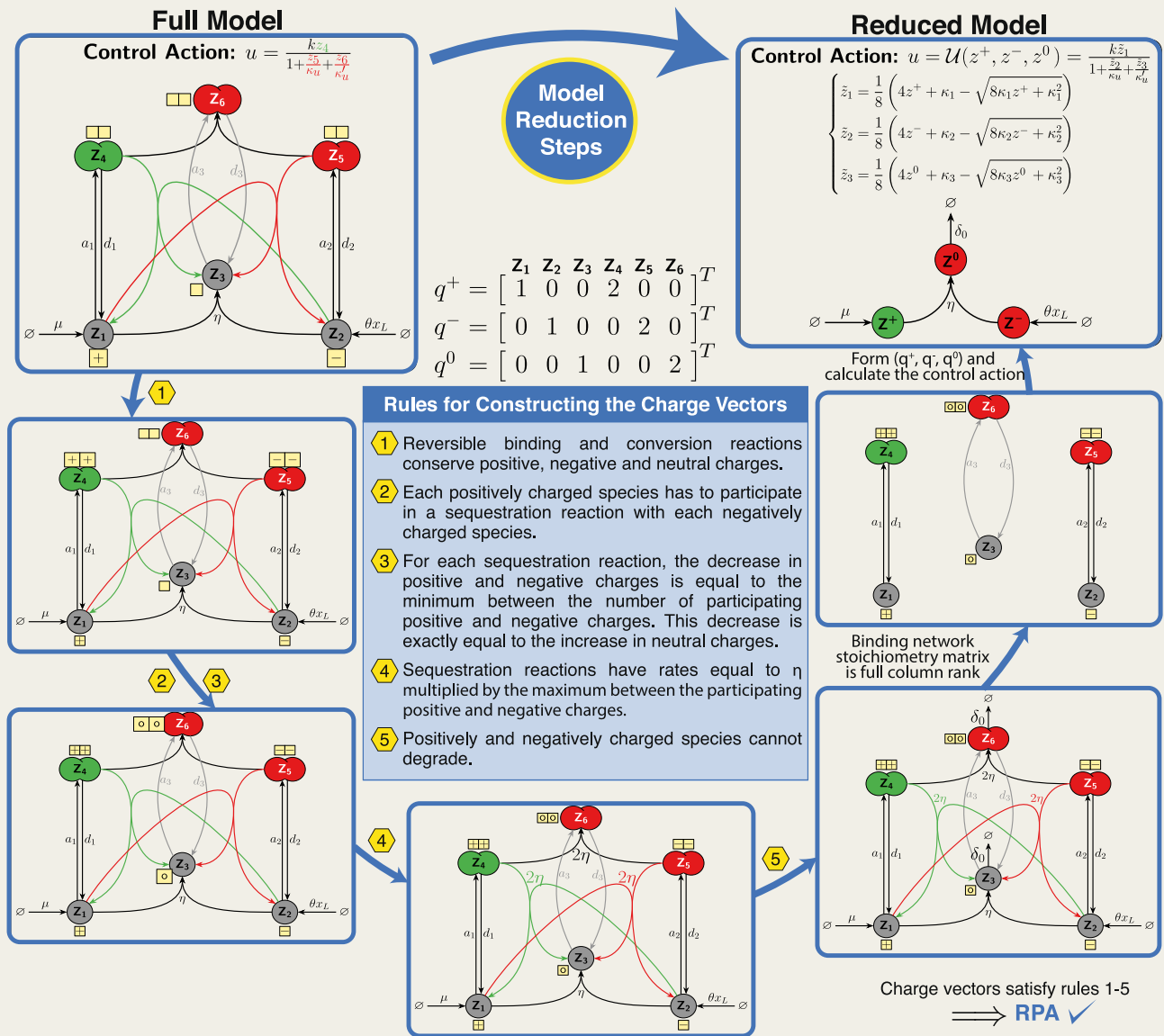
$\mathbb{1}(\cdot)$, \circ , I_M and $(\cdot)^T$ denote the indicator function, the Hadamard (element-wise) product, the identity matrix of size M and the transpose of a matrix, respectively. In certain scenarios (see Fig. 6), the algebraic equations can be solved explicitly and thus further simplifying the model to only three *Ordinary Differential Equations* (ODEs). Observe that the schematic of the simple motif is fully determined once the three vectors q^+, q^-, q^0 and the function $\psi(z^{\text{tot}})$ are calculated. The vectors q^+, q^- and q^0 are easily calculated by counting active split inteins (see Theorem 2); whereas, $\psi(z^{\text{tot}})$ can be calculated by solving the algebraic equations for $\tilde{z} \geq 0$ as a function of z^{tot} .

BOX 3

Example application of the RPA and model reduction theorems to a general controller network

The goal of this illustrative example is to reduce the full model of the controller network comprised of $M = 6$ species Z_1, Z_2, \dots, Z_6 to the simple motif comprised of Z^+, Z^- and Z^0 which is generally depicted in Fig. 5. The full model here is chosen to be purely mathematical, i.e. with no relevance to inteins, in order to demonstrate that the theorems are not restricted to intein-based controllers only. Essentially, the q and (q^+, q^-, q^0) vectors in Theorem 1 and 2 that were related to the number of inteins present in the controller species, are now more generally referred to as charge vectors inspired by⁶. For establishing RPA, computing $q := q^+ - q^-$ is enough; whereas to carry out the model reduction technique, all individual positive, negative and neutral charge vectors (q^+, q^-, q^0) need to be computed. Unlike the intein-based controller networks where the charge vectors are simply computed by counting intein segments, here we lay down general rules for

constructing these charge vectors. Once these charge vectors are available, then one can easily apply Theorems 1 and 2 for establishing RPA and performing the model reduction, respectively. Note that in this example, we include three reversibly binding reactions but no conversion reactions for simplicity, and thus the conditions of Theorem 2 boil down to requiring the stoichiometry matrix of the binding network to be full-column rank and equal to $M - 3 = 3$. Furthermore, for simplicity, we start by assuming that Z_1 and Z_2 carries only one positive and negative charge, respectively. This assumption can however be relaxed to a more general number of charges (see Supplementary Information, Theorem 2). Two other mathematical controller topologies are also proposed and similarly analyzed in Supplementary Information, Section 4: one controller involves conversion reactions, and the other involves a trimer.



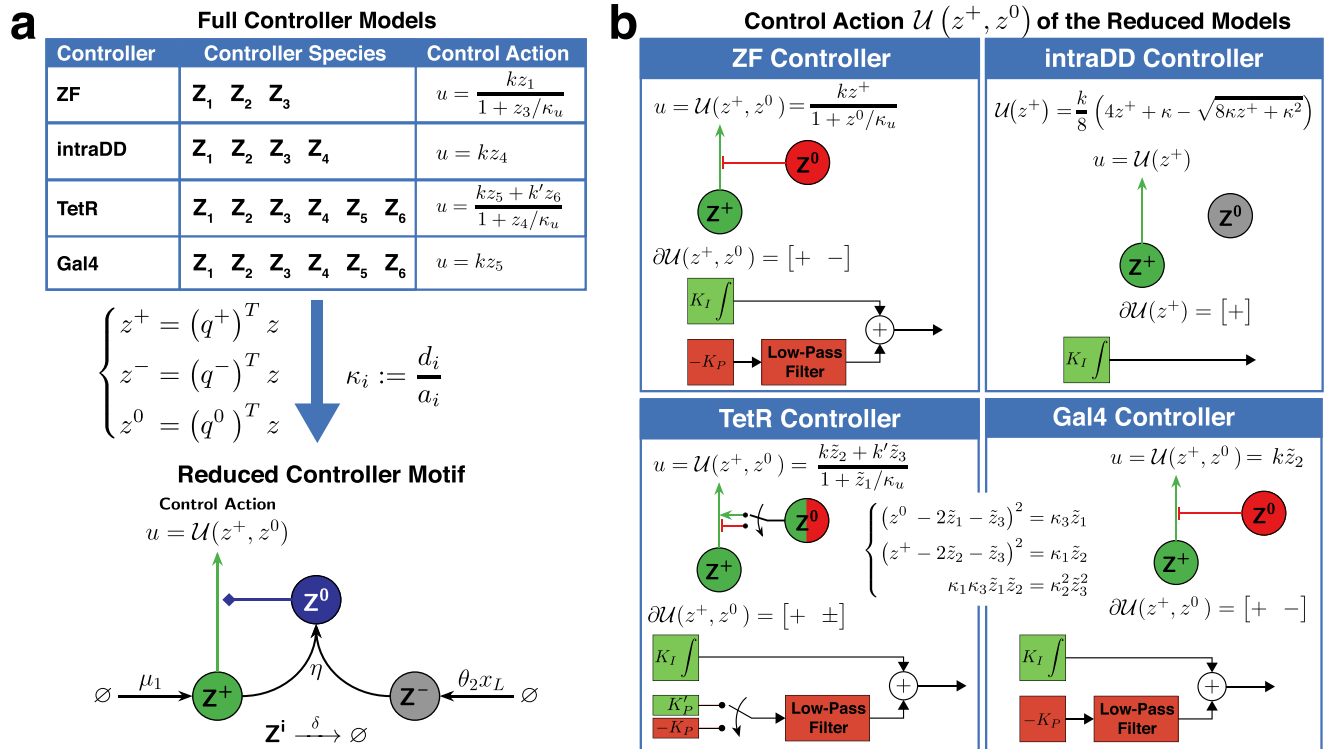


Fig. 6 | Reduced models for the ZF, intraDD, TetR and Gal4 controllers. **a** Reduced Motif. For simplicity, we assume that the protein degradation rates are negligible compared to the dilution rate δ ; however, this assumption can be easily relaxed (see Supplementary Information Section 3). Note that δ is assumed to be non-zero here to capture the more realistic scenario. The model reduction recipe presented in Theorem 2 and Fig. 5 can be straightforwardly applied to all of the four controller topologies in Fig. 3, where the “charge” vectors q^+ , q^- and q^0 are shown explicitly. Observe that all four controllers reduce to the same motif comprised of the three effective species Z^+ , Z^- and Z^0 . The difference between them appears only in the effective control action $u = U(z^+, z^0)$. **b** Effective control actions. The control actions $u = U(z^+, z^0)$ is given separately for each controller as a function of the effective species concentrations. For the intraDD controller, the control action u is

a strictly monotonically increasing function of z^+ only, and hence the control structure is a standalone integrator. In contrast, for the ZF and Gal4 controllers, it is shown (see Supplementary Information Sections 3.D and 3.B) that the control action u is strictly monotonically increasing (resp. decreasing) in z^+ (resp. z^0). This gives rise to a filtered proportional-integral (PI) control structure³⁸. Finally, for the TetR controller, it is shown that the control action u is strictly monotonically increasing in z^+ ; whereas, its monotonicity switches from increasing to decreasing as the levels of (z^+, z^0) rise (see Supplementary Information Section 3.A). This gives rise to a filtered PI control structure where the P-component switches sign. Note that the algebraic equations presented in Fig. 5 are solved explicitly for the ZF and intraDD controllers; however, they are kept in their implicit form for the TetR and Gal4 controllers.

the more practical scenario where all controller species dilute at a rate $\delta > 0$. Furthermore, we assume, for simplicity, that the degradation of the various proteins are negligible compared to the dilution rate; however, this assumption can be easily relaxed (see Supplementary Information Section 3). The model reduction results are compactly depicted in Fig. 6 for all four controllers. The underlying reduced motif, as illustrated in Fig. 6, is the same for all four controller circuits and is comprised of only three effective species Z^+ , Z^- and Z^0 whose concentrations are linear combinations of the biological species Z_i . The differences between the reduced models of each controller circuit is encrypted in the effective control action $u = U(z^+, z^0)$ which is a function of the concentrations of Z^+ and Z^0 . Observe that the control action is given in an explicit form for the ZF and intraDD controllers; whereas, for the TetR and Gal4 controllers, it is given implicitly as a set of three algebraic equations. Once these algebraic equations are solved for $(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3) \geq 0$, the control actions can be directly computed as a function of z^+ and z^0 . The topology of the reduced models is clearly simpler to analyze compared to the full models described in Fig. 3, and thus the underlying control architecture can be uncovered more easily. In fact, the intraDD controller realizes a standalone antithetic integral controller since the control action $u = U(z^+)$ depends (monotonically) on Z^+ only. On the other hand, it is shown in Supplementary Information Sections 3.D and 3.B that the control action $u = U(z^+, z^0)$ of the ZF- and Gal4-Circuits depends on both Z^+ and Z^0 , such that U is monotonically increasing (resp. decreasing) in z^+ (resp. z^0). This

particular topology can be shown to realize a filtered Proportional-Integral (PI) controller, where the proportional component can be used as an additional knob to enhance the dynamic performance (see³⁸ for a thorough analysis). Finally, it is shown in Supplementary Information Section 3.A that the control action $u = U(z^+, z^0)$ of the TetR controller also depends on both Z^+ and Z^0 . Nevertheless, U is a monotonically increasing function of z^+ , but its monotonicity switches from increasing (at low levels of z^0 and z^+) to decreasing (at higher levels of z^0 and z^+). We refer the reader to Supplementary Information Section 3.A for more details on the exact monotonicity analysis of U . Interestingly, this architecture realizes a filtered PI controller whose proportional component switches from positive to negative gain. This gives rise to a nice feature that initially speeds up the response when the concentrations of the controller species are low, and then switches to negative feedback as the concentrations rise and thus favoring closed-loop stability. The various reduced models are validated via simulations that demonstrate the highly accurate matching between the dynamics of the full and reduced models in Supplementary Information Section 4.

Integral controllers with competing sequestrations

In this section, we demonstrate that Theorem 1 can be applied to controller circuits that are more general compared to those of Theorem 2. That is, there are certain intein-based controllers that can be easily tested for RPA using Theorem 1; however, their model reduction cannot be carried out by applying Theorem 2. We do so by considering

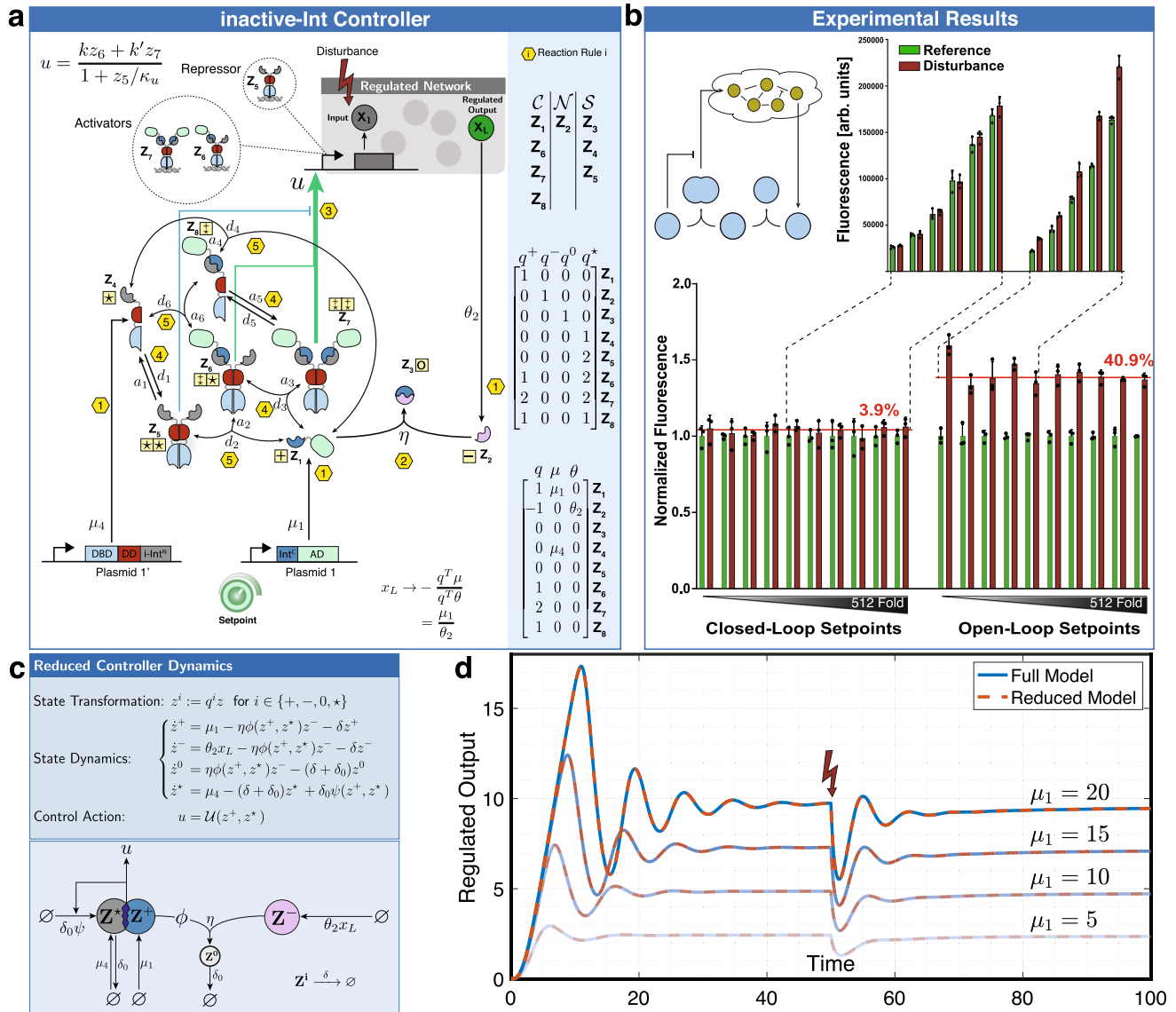


Fig. 7 | Inactive-intein controller: theoretical and experimental analysis. **a** A schematic of the inactive-intein controller. This controller consists of two genes, realized on separate plasmids. The gene in Plasmid 1 encodes for a protein (Z_1) comprised of Int^C-AD; whereas, the gene in Plasmid 1' encodes for a protein (Z_4) comprised of TetR-inactive Int^N. Both genes are driven by a strong constitutive promoter (EF-1 α), and their expression rates are denoted by μ_1 and μ_4 , respectively. Z_1 and Z_4 can reversibly bind to form a heterodimeric transcription factor, which positively actuates the regulated network via the production of the input species X_1 . The production of the second split intein Int^N, denoted by Z_2 , is driven by the regulated output X_1 at a rate $\theta_2 x_L$ to encode for the “sensing” reaction. Controller species containing DDs undergo reversible homo- or hetero-dimerization reactions with association and dissociation rates of a_i and d_i . Here, only Z_1 and Z_2 can directly undergo the intein-splicing reaction (at a rate η), because Z_1 is the only species that contains an active Int^C segment not bound to the inactive Int^N segment. The control action u is mathematically expressed as a (Hill-type) function of the repressors and activators depicted in the dashed bubbles. Every reaction is labeled from 1 to 6 according to the permitted reaction rules stated in Fig. 2. The entire charge matrix can be viewed in the blue shaded box where, additionally, all controller species have been grouped into the three classes, C -class, \mathcal{N} -class and S -class, according to

the species rules of Fig. 2. Since all the Species and Reaction Rules of Fig. 2 are respected, then by Theorem 1, we conclude that this controller ensures RPA with a setpoint of μ_1/θ_2 and is thus interestingly independent of μ_4 . **b** Experimental demonstration of RPA. The performance was tested using the same setup as in Fig. 4(a). The only difference here is that the Int^N segment (Gp41-1) is replaced by an orthogonal Int^N (NrdJ-1) for the open-loop setting, and thus no intein-splicing reaction can occur. The results are demonstrated in a fashion similar to that of Fig. 4. **c** Reduced model. Unlike the three-dimensional reduced models in Fig. 6 that are obtained by directly applying Theorem 2, the reduced model here is four dimensional because it was necessary to introduce the dynamics of an additional state variable z^* . The functions ψ and ϕ are given implicitly in Supplementary Information Section 3.E. Note that the cartoon describing the reduced model is non-physical because the mathematical equations do not satisfy the structure of a simple motif like the models that satisfy Theorem 2. **d** Simulation Results. A closed-loop system is simulated for four increasing setpoints, where a model of a gene expression network is controlled by the inactive-intein controller. The simulation results demonstrate that the reduced model indeed accurately captures the dynamics of the original full model. The numerical values are provided in Supplementary Information Section 3.E.

the circuit depicted in Fig. 7(a), where two independent controller species (active Int^N denoted by Z_2 versus inactive Int^N denoted by Z_4 in Fig. 7(a)) stoichiometrically compete to sequester another controller species (Z_1 in Fig. 7(a)). In this circuit, we constructed two genes encoding for an AD fused to an active Int^C (expressing Z_1) and a DBD-

DD fused to an inactive Int^N (expressing Z_4). Although the inactive Int^N lacks essential amino acids to undergo the intein-splicing reaction⁵⁵, Z_4 can still reversibly bind to Z_1 to form a heterodimeric transcription factor. In this controller design, the intein-splicing reaction can occur only between the expressed Int^N, denoted by Z_2 , and Z_1 is

the only controller species that contains an active Int^c segment in its unbound state. In fact, although the other controller species containing active Int^c segments (\mathbf{Z}_6 , \mathbf{Z}_7 and \mathbf{Z}_8) belong to the \mathcal{C} -class, they cannot directly undergo intein-splicing reactions since they are bound to the inactive Int^N. This results in a violation of Reaction Rule 9 rendering the model reduction recipe of Theorem 2 inapplicable. Nonetheless, it is straightforward to check that the conditions of Theorem 1 still apply and, as a result, RPA is still guaranteed as long as the closed-loop system is stable. Furthermore, applying (2), by noting that $q = [1 \ -1 \ 0 \ 0 \ 0 \ 1 \ 2 \ 1]^T$, yields the setpoint expression given by μ_1/θ_2 (see Fig. 7(a)). Observe that the rate of expression μ_4 of \mathbf{Z}_4 does not affect the setpoint – a result that is not immediate without resorting to (2). Similar to Fig. 4, the experimental results depicted in Fig. 7(b) demonstrate that the controller indeed ensures RPA yielding an average steady-state error of 3.9% over a wide dynamic range of setpoints compared to an error of 40.9% when operating in open loop.

Although the model reduction recipe provided in Theorem 2 cannot be applied here, one can still invoke singular perturbation theory to this particular controller circuit to obtain the reduced mathematical model depicted in Fig. 7(c). The model reduction here assumes, once again, that the reversible binding reactions are fast. Observe that, unlike the previous controllers, the reduced model is four dimensional. Intuitively, this is a result of an additional conservation law imposed by the inactive inteins which introduce an additional (fourth) vector q^* required to carry out the state transformation. Hence, the reduced mathematical model is described by the set of four ODEs for (z^+, z^-, z^0, z^*) shown in Fig. 7(c) where the functions ϕ and ψ are implicitly given in Supplementary Information Section 3.E. A “fictitious network” describing the ODEs is also depicted in Fig. 7(c) to emphasize that the reduced model is mainly mathematical and cannot be easily translated to a simple motif. This highlights that controller circuits not adhering to the conditions of Theorem 2 fail to reduce to the simple motif given in Fig. 2. The reduced model is validated by the simulation results shown in Fig. 7(d) for four different setpoints and by applying a disturbance.

Discussion

In this paper, we introduced a theoretical and experimental framework to design, build and analyze a broad class of biomolecular integral feedback controllers that achieve RPA. The framework is based on custom-built split inteins that are shown to be capable of realizing the sequestration reaction – the heart of the basic antithetic integral feedback motif – via protein splicing. The sequestration reaction in previously proposed^{20,37,39,40,52} and built^{6,7,9,13} integral controllers, whether in vivo, or in vitro, relies on the complete stoichiometric annihilation of two controller species (see \mathbf{Z}_1 and \mathbf{Z}_2 in Fig. 1(a)). Here, we relax this requirement by establishing that the sequestration reaction does not have to fully annihilate the participating controller species, and, in fact, it suffices to stoichiometrically annihilate sub-components within these two controller species. Indeed, this is precisely what intein-splicing reactions do: active split inteins inserted in two target proteins are *inactivated* by undergoing the splicing reaction. While the function of the active split inteins is indeed *annihilated*, the spliced target proteins are still allowed to have specific functions. In fact, we showed that one can harness the function of the spliced proteins to augment the standalone integral controller with a filtered proportional component to yield a PI controller. We previously computationally demonstrated (see³⁸) that the resulting filtered PI controller adds an extra degree of freedom which enables the enhancement of the transient performance and the reduction of cell-to-cell variability while maintaining the RPA property. However, it is left for future work to back up this theory with experimental demonstrations. It is worth to mention that the realization of a molecular PI controller in mammalian cells is not new. Ideally, a proportional component can be theoretically achieved by appending the integral controller with an instantaneous

negative feedback from the regulated output species \mathbf{X}_1 onto the input species \mathbf{X}_1 (see e.g.^{37,40}). This requires the output species \mathbf{X}_1 to have multiple functions including the production of \mathbf{Z}_2 for sensing and the inhibition of the input species \mathbf{X}_1 to realize the proportional component. In practice this might not be possible as the output species is determined by the biological application. In⁷, this was circumvented by introducing additional genetic parts to express a proxy to the regulated output upon which the proportional control action is based on. Here, in contrast, the design flexibility and modularity offered by inteins allowed us to implement PI controllers by simply choosing an actuator and a suitable insertion site of the split-intein (see Fig. 3 and 6) without adding additional genetic parts and without requiring the regulated output \mathbf{X}_1 to have multiple functions.

The simple antithetic integral feedback control topology was first introduced in²⁰, and more recently a generalized antithetic topology was introduced in⁶ which characterizes all RPA-achieving controllers involving exactly one sensing and one setpoint-encoding reaction. This characterization lead to simple algebraic conditions that enable RPA and are expressed in terms of quantities that are referred to as “charges”. The general charge analogy borrowed from electronics was made due to the lack of biological parts capable of respecting the algebraic conditions. This is exactly where inteins came in, because they naturally satisfy the RPA algebraic conditions and act as “charges” neutralizing each other via the intein-splicing reactions. Indeed, split inteins are typically charged at the locations where they interact⁵⁶. This makes the charge analogy biologically suitable. In fact, Theorem 1, which is a direct application of Supplementary Theorem 1 (tailored towards inteins), is a generalization of the RPA sufficiency result of⁶ such that multiple sensing and setpoint-encoding reactions are now allowed. Theorem 1 facilitates the screening of controller circuit designs for RPA. Furthermore, we went one step further here, beyond establishing RPA, to provide an easy-to-apply recipe for model reduction. The recipe is given in Theorem 2 which is, once again, a direct application of Supplementary Theorem 2 tailored towards inteins (see Box 3 for an application example of these theorems in a purely mathematical and more general context, that is, without an intein-based interpretation). The model reduction result presented here exploits the time-scale separation imposed by fast reversible binding and conversion reactions and is established by invoking singular perturbation theory⁵⁴ and the deficiency zero theorem⁵³ to prove structural (rate-independent) stability of the slow manifold.

The five controller circuit implementations presented in this paper (see Fig. 3 and 7) are based on the widely used DNA binding domains TetR, ZF and Gal4. For the experimental verification of RPA, we used a simple regulated network (see Fig. 4(a)) that resulted in a two (resp. three) plasmid closed-loop system depicted in Fig. 4 (resp. 7). The regulated network was intentionally chosen to be simple here, in order to minimize possible cross talks which might emerge from larger networks (e.g. burden)⁴⁸. This allowed us to focus our study on the controllers themselves instead of possible undesirable behaviors incurred by larger networks – an important topic that is not within the scope of the current study and is left for future work. Note that with this experimental setup, we were not able to directly detect the regulated output which is an mRNA (see Fig. 4(a)). To circumvent this, we used a fluorescent reporter which, unlike the regulated (mRNA) species, is not robust to translational burden. This implies that although RPA is not observed at high setpoints by the reporter, it may actually be achieved by the mRNA.

The controller circuits that are designed, built and analyzed in this paper are all based on controller species generated using TFs. However, split inteins can also be introduced in other protein classes such as proteases (Supplementary Information, Fig. 18) and receptors (Supplementary Information, Fig. 19). Split inteins can be even introduced in endogenous proteins to convert them into controller species. This has an attractive advantage of exploiting parts of the regulated

network to realize the controller and, as a result, requiring less to no additional genes. From a protein engineering point of view, such designs may be more challenging than designs based on the well-characterized TFs used in this study. Besides tinkering with insertion sites, linker lengths and split-intein pairs, it is also possible to use more systematic approaches like transposon screens with inteins as performed by Ho et al.⁵⁷ or computationally-guided optimizations by Dolber et al.⁵⁸

The remarkable flexibility offered by inteins for building integral controllers opens the doors to many possible future research directions. For instance, it is easy to think of regulated networks with negative gain, in other words, producing more input species X_1 leads to a lower concentration of the regulated output species X_L . For example, producing more insulin leads to a lower concentration of glucose in the blood. As a result, to realize an overall negative feedback, the actuation direction of the controller species Z_1 would have to be flipped, that is instead of having Z_1 upregulating X_1 (like in Fig. 1(a) and, in fact, all previously built antithetic integral controllers), Z_1 would have to downregulate X_1 (see Supplementary Information, Fig. 18). Intein-based realizations of such “negative actuation” mechanisms can be easily carried out using repressors or proteases. Furthermore, n inteins (with $n=1, 2, 3, \dots$) can be embedded sequentially in a single controller species leading to the scaling of the setpoint by an integer n (see Supplementary Information, Fig. 11 and 12). Note that other functional domains can be placed between inteins to alter the functionality of the various spliced products (see Supplementary Information, Fig. 13 and 14). The flexibility offered by inteins also allows us to freely design the (multi)functionality of the spliced products as activators and/or repressors (e.g. Supplementary Information, Fig. 15, 16 and 17).

Another possible future direction is intein-based implementations of more advanced controllers. For example, one can easily add functional domains to the controller species Z_2 , which was comprised of a standalone Int^N segment in all the controller circuits proposed here. These added domains enable the implementation of the rein controller introduced in⁵⁹ which is capable of enhancing the overall performance. Another example is the implementations of more advanced biomolecular Proportional-Integral-Derivative (PID) controllers³⁷ that are capable of shaping the transient response and reducing cell-to-cell variability. In particular, the wide library of orthogonal split inteins⁵¹ allows one to implement the fourth order PID controller³⁷ that is comprised of two antithetic motifs: antithetic integrator and antithetic differentiator.

In conclusion, rather than providing another way of implementing antithetic integral controllers, we propose here a systematic (theoretical and experimental) approach of designing, building and analyzing a broad class of biomolecular integral controllers that are capable of achieving RPA. The key of our approach is the exploitation of the splicing reactions that occur between split inteins. Due to their simplicity, modularity, irreversibility, lack of side effects and applicability across species, we believe that inteins will revolutionize biomolecular controllers and partake in filling the gap between theory and experiments.

Methods

Plasmid construction

All plasmids were generated with a mammalian adaptation of the modular cloning (MoClo) yeast toolkit standard⁶⁰. All individual parts were generated by PCR amplification (Phusion Flash High-Fidelity PCR Master Mix; Thermo Scientific) or synthesized with Twist Bioscience. PCR primers were obtained from Sigma-Aldrich and Integrated DNA Technologies. The parts were then assembled with golden gate assembly. All enzymes for plasmid construction were obtained from New England Biolabs (NEB). Constructs were chemically transformed into *E. coli* Top10 strains (Invitrogen). The plasmid list and protein

sequences can be found in Supplementary Information Section 9. DNA and oligo sequences can be found in the Data Source file.

Cell culture

All experiments were performed with HEK293T cells (ATCC, strain number CRL-3216, LGC standards). The cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% FBS (Sigma-Aldrich), 1x GlutaMAX (Gibco), 1 mM Sodium Pyruvate (Gibco), penicillin (100U/ μ L), and streptomycin (100 μ g/mL) (Gibco) at 37° with 5% CO₂. The cell culture was passaged into a fresh T25 flask (Axon Lab) every 2 to 3 days. Upon detachment some part of the cell suspension was used for the transfection.

Transfection

All plasmids were isolated using ZR Plasmid Miniprep-Classic (Zymo Research). The plasmids were introduced to the HEK293T cells via suspension transfection. A transfection solution in Opti-MEM I (Gibco) was prepared using Polyethylenimine (PEI) “MAX” (MW 40000; Polysciences, Inc.) at a 1:3 (μ g DNA to μ g PEI) ratio while the culture was detached with Trypsin-EDTA (Gibco). The cell density was assessed with the automated cell counter Countess II FL (Invitrogen). 100 μ L of culture with 26'000 cells was transferred in each well of the plate Nunc Edge 96-well plate (Thermo Scientific). The transfection mixture was added to the cells once it has incubated for approximately 30 min. All transfection tables can be found in Supplementary Information Section 9.

Flow cytometry

The cells were detached approximately 48 h after transfection on the Eppendorf ThermoMixer C at 25 °C at 700 rpm with 53 μ L Accutase solution (Sigma-Aldrich) per well for 20 min. The fluorescence data was collected on the Beckman Coulter CytoFLEX S flow cytometer with the 488 nm excitation with a 525/40+OD1 bandpass filter and the 638 nm excitation with a 660/10 bandpass filter. All data was processed with the CytExpert 2.3 software. A representative example of the gating strategy can be found in Supplementary Information, Fig. 28. The data was visualized with GraphPad 8.2.0.

Numerical simulations and visualizations

All simulations are carried out in MATLAB R2021a (academic use). Stochastic simulations shown in the supplementary information file are carried out on the Euler cluster (<https://scicomp.ethz.ch/wiki/Euler>). Manuscript figures were structured and formatted on Illustrator (2022 26.5), MATLAB and TexStudio (v3.1.1, open source).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The fluorescence measurements data of Fig. 4 and 7 are provided in the Supplementary Information File under Section 9. Plasmid and oligo sequences can be found in the Source Data file. Plasmids are available from the corresponding author upon request. Source data are provided with this paper.

Code availability

The MATLAB code for generating all the figures is available at a dedicated GitHub repository: <https://github.com/Maurice-Filo/Inteins-in-the-Loop>⁶¹.

References

1. Kotas, M. E. & Medzhitov, R. Homeostasis, inflammation, and disease susceptibility. *Cell* **160**, 816–827 (2015).
2. Mullur, R., Liu, Yan-Yun & Brent, G. A. Thyroid hormone regulation of metabolism. *Physiol. Rev.* **94**, 355–382 (2014).

3. M, Khammash, Mario Di, Bernardo, and Diego Di, Bernardo Cybergenetics: Theory and methods for genetic control system. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 916–926. IEEE, (2019).
4. Lillacci G., Aoki S., Schweingruber D., & Khammash M. A synthetic integral feedback controller for robust tunable regulation in bacteria. *BioRxiv*. 170951, <https://doi.org/10.1101/170951> (2017).
5. Lillacci, G., Benenson, Y. & Khammash, M. Synthetic control systems for high performance gene expression in mammalian cells. *Nucl. Acids Res.* **46**, 9855–9863 (2018).
6. Aoki, S. K. et al. A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature* **570**, 533–537 (2019).
7. Frei, T., Chang, Ching-Hsiang, Filo, M., Arampatzis, A. & Khammash, M. A genetic mammalian proportional–integral feedback control circuit for robust and precise gene regulation. *Proc. Nat. Acad. Sci.* **119**, e2122132119 (2022).
8. Franco, E., Giordano, G., Forsberg, Per-Ola & Murray, R. M. Negative autoregulation matches production and demand in synthetic transcriptional networks. *ACS Synthetic Biol.* **3**, 589–599 (2014).
9. Huang, Hsin-Ho, Qian, Y. & Del Vecchio, D. A quasi-integral controller for adaptation of genetic modules to variable ribosome demand. *Nat. Commun.* **9**, 1–12 (2018).
10. Ng, A. H. et al. Modular and tunable biological feedback control using a de novo protein switch. *Nature* **572**, 265–269 (2019).
11. Hsiao, V., De Los Santos, Emmanuël C., Whitaker, W. R., Dueber, J. E. & Murray, R. M. Design and implementation of a biomolecular concentration tracker. *ACS Synthetic Biol.* **4**, 150–161 (2015).
12. Kelly, Ciarán L. et al. Synthetic negative feedback circuits using engineered small rnas. *Nucl. Acids Res.* **46**, 9875–9889 (2018).
13. Agrawal, D. K., Marshall, R., Noireaux, V. & Sontag, E. D. In vitro implementation of robust gene regulation in a synthetic biomolecular integral controller. *Nat. Commun.* **10**, 1–12 (2019).
14. Kemmer, C. et al. Self-sufficient control of urate homeostasis in mice by a synthetic circuit. *Nat. Biotechnol.* **28**, 355–360 (2010).
15. Rössger, K., Charpin-El-Hamri, G. & Fussenegger, M. A closed-loop synthetic gene circuit for the treatment of diet-induced obesity in mice. *Nat. Commun.* **4**, 1–9 (2013).
16. Xie, M. et al. β -cell-mimetic designer cells provide closed-loop glycemic control. *Science* **354**, 1296–1301 (2016).
17. F., Xiao and J. C., Doyle. Robust perfect adaptation in biomolecular reaction networks. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 4345–4352. IEEE, (2018).
18. Khammash, M. H. Perfect adaptation in biology. *Cell Sys.* **12**, 509–521 (2021).
19. Francis, B. A. & Wonham, Walter Murray The internal model principle of control theory. *Automatica* **12**, 457–465 (1976).
20. Briat, C., Gupta, A. & Khammash, M. Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks. *Cell Sys.* **2**, 15–26 (2016).
21. Olsman, N., Xiao, F. & Doyle, J. C. Architectural principles for characterizing the performance of antithetic integral feedback networks. *Isience* **14**, 277–291 (2019).
22. Olsman, N. et al. Hard limits and performance tradeoffs in a class of antithetic integral feedback networks. *Cell Sys.* **9**, 49–63 (2019).
23. M., Filo and M., Khammash. Optimal parameter tuning of feedback controllers with application to biomolecular antithetic integral control. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 951–957. IEEE, (2019).
24. Bernardo D. di. Engineering the next generation of mammalian whole cell biosensors for advanced biotechnological applications. 2022 International Mammalian Synthetic Biology Workshop, (2022).
25. Yudkin, M. D. The sigma-like product of sporulation gene spoiac of bacillus subtilis is toxic to escherichia coli. *Mol. General Genetics* **MGG**. **202**, 55–57 (1986).
26. Nejepinska, J., Malik, R., Wagner, S. & Svoboda, P. Reporters transiently transfected into mammalian cells are highly sensitive to translational repression induced by dsrna expression. *PLoS One*. **9**, e87517 (2014).
27. Sarmiento, C. & Camarero, J. A. Biotechnological applications of protein splicing. *Current Protein Peptide Sci.* **20**, 408–424 (2019).
28. Lennon, C. W. & Belfort, M. Inteins. *Current Biol.* **27**, R204–R206 (2017).
29. Shah, N. H. & Muir, T. W. Inteins: nature’s gift to protein chemists. *Chem. Sci.* **5**, 446–461 (2014).
30. Muldoon, J. J. et al. Model-guided design of mammalian genetic programs. *Sci. Adv.* **7**, eabe9375 (2021).
31. Palanisamy, N. et al. Split intein-mediated selection of cells containing two plasmids using a single antibiotic. *Nat. Commun.* **10**, 1–15 (2019).
32. Jillette, N., Du, M., Zhu, JacquelineJufen, Cardoz, P. & Wu Cheng, A. Split selectable markers. *Nat. Commun.* **10**, 1–8 (2019).
33. Chen, Hao-Shan et al. An intein-split transactivator for intersectorial neural imaging and optogenetic manipulation. *Nat. Commun.* **13**, 1–14 (2022).
34. Wang H., Wang L., Zhong B., & Dai, Z. Protein splicing of inteins: A powerful tool in synthetic biology. *Frontiers in Bioengineering and Biotechnology. Front Bioeng. Biotechnol.* **10**, 810180 (2022).
35. Dassa, B., London, N., Stoddard, B. L., Schueler-Furman, O. & Pietrokovski, S. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucl. Acids Res.* **37**, 2560–2573 (2009).
36. Klein, J. S., Jiang, S., Galimidi, R. P., Keeffe, J. R. & Bjorkman, P. J. Design and characterization of structured protein linkers with differing flexibilities. *Protein Eng. Des. Sel.* **27**, 325–330 (2014).
37. Filo, M., Kumar, S. & Khammash, M. A hierarchy of biomolecular proportional-integral-derivative feedback controllers for robust perfect adaptation and dynamic performance. *Nat. Commun.* **13**, 1–19 (2022).
38. Filo, M., Kumar, S., Anastassov, S. & Khammash, M. Exploiting the nonlinear structure of the antithetic integral controller to enhance dynamic performance. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, 1294–1299 (IEEE, 2022).
39. Chevalier, M., Gómez-Schiavon, M., Ng, A. H. & El-Samad, H. Design and analysis of a proportional-integral-derivative controller with biological molecules. *Cell Sys.* **9**, 338–353 (2019).
40. Briat, C., Gupta, A. & Khammash, M. Antithetic proportional-integral feedback for reduced variance and improved control performance of stochastic reaction networks. *J. Royal Society Interface.* **15**, 20180079 (2018).
41. Samaniego, ChristianCuba & Franco, E. Ultrasensitive molecular controllers for quasi-integral feedback. *Cell Sys.* **12**, 272–288 (2021).
42. V., Martinelli, D., Salzano, D., Fiore, and M., di Bernardo. Multi-cellular pi control for gene regulation in microbial consortia. *bioRxiv* (2022).
43. Donahue, P. S. et al. The comet toolkit for composing customizable genetic programs in mammalian cells. *Nat. Commun.* **11**, 1–19 (2020).
44. Khalil, A. S. et al. A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647–658 (2012).
45. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci.* **89**, 5547–5551 (1992).
46. Carvajal-Vallejos, P., Pallissé, R., Mootz, H. D. & Schmidt, S. R. Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J. Biol. Chem.* **287**, 28686–28696 (2012).

47. E Müller, I. et al. Gene networks that compensate for crosstalk with crosstalk. *Nat. Commun.* **10**, 1–8 (2019).
 48. Frei, T. et al. Characterization and mitigation of gene expression burden in mammalian cells. *Nat. Commun.* **11**, 1–14 (2020).
 49. Nöbel, F., De Battista, H., Nuñez, S. & Picó, J. Reference conditioning anti-windup for the biomolecular antithetic controller. *IFAC-PapersOnLine*. **52**, 156–162 (2019).
 50. Sharma, P. et al. 2a peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucl. Acids Res.* **40**, 3143–3151 (2012).
 51. Pinto, F., Thornton, EllaLucille & Wang, B. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. *Nat. Commun.* **11**, 1–15 (2020).
 52. Qian, Y. & Del Vecchio, D. Realizing ‘integral control’ in living cells: how to overcome leaky integration due to dilution? *J. Royal Society Inter.* **15**, 20170902 (2018).
 53. Feinberg, M. Chemical reaction network structure and the stability of complex isothermal reactors-i. the deficiency zero and deficiency one theorems. *Chem. Eng. sci.* **42**, 2229–2268 (1987).
 54. H. K., Khalil. Nonlinear systems third edition. *Patience Hall* 115, (2002).
 55. Stevens, A. J. et al. Design of a split intein with exceptional protein splicing activity. *J. American Chem. Society.* **138**, 2162–2165 (2016).
 56. Beyer, H. M., Mikula, K. M., Li, M., Wlodawer, A. & Iwai, H. The crystal structure of the naturally split gp41-1 intein guides the engineering of orthogonal split inteins from cis-splicing inteins. *FEBS J.* **287**, 1886–1898 (2020).
 57. Ho, TrevorY. H. et al. A systematic approach to inserting split inteins for boolean logic gate engineering and basal activity reduction. *Nat. Commun.* **12**, 1–12 (2021).
 58. Dolberg, T. B. et al. Computation-guided optimization of split protein systems. *Nat. Chem. Biol.* **17**, 531–539 (2021).
 59. A., Gupta and M., Khammash. An antithetic integral rein controller for bio-molecular networks. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 2808–2813. IEEE, (2019).
 60. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synthetic Biol.* **4**, 975–986 (2015).
 61. M., Filo. Inteins-in-the-Loop: A Framework for Engineering Advanced Biomolecular Controllers for Robust Perfect Adaptation. Github repository: Inteins-in-the-Loop, (2022). <https://doi.org/10.5281/zenodo.7048552>
- and assistance. We would also like to thank Drs. Zhou Fang, Timothy Frei and Stephanie Aoki for the insightful discussions.

Author contributions

S.A. and M.F. conceived the study. M.F. developed the theorems and carried out the mathematical analyses and simulations. S.A. carried out the parts characterizations. C.-H.C. designed the two-plasmid testing framework for RPA. S.A. and C.-H.C designed, built and experimentally validated the controller circuits and performed the data analysis; all authors wrote the paper. M.K. supervised the project and secured funding.

Competing interests

ETH Zurich has filed a patent application on behalf of the inventors S.A., M.F., C.-H.C. and M.K. on the genetic circuit designs described (application no. EP22186956.3).

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-36863-9>.

Correspondence and requests for materials should be addressed to Mustafa Khammash.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (CyberGenetics; grant agreement 743269). We acknowledge the ETH Zurich D-BSS Single Cell Facility for its excellent support