Supplementary Information for "Universal structures for embedded integral control in biological adaptation"

Robyn P. Araujo[1,2*] and Lance A. Liotta[3]

[1]School of Mathematical Sciences, Queensland University of Technology, Australia.
[2]Institute of Health and Biomedical Innovation (IHBI), Brisbane, Australia.
[3]Center for Applied Proteomics and Molecular Medicine, George Mason University, USA.

[*]To whom correspondence should be addressed: r.araujo@qut.edu.au

Supplementary Text

In this Supplement, we provide full details and analysis of the key technical results summarised in the main paper, along with proofs for all supporting theorems, a range of additional illustrative examples, as well as code in the open-source software *Singular* [1] that can be adapted to any collection of chemical reactions.

This work pursues, and achieves, two interrelated goals. The first goal is to develop a definitive test for RPA capacity in any chemical reaction network (CRN) - one that can be applied to CRNs of any size or 'deficiency' [10], to both mass-conservative and 'open' CRNs [4], and one that can distinguish the special case of RPA known as Absolute Concentration Robustness (ACR). The second goal is to demonstrate the universal principles by which *all* instances of RPA-capable CRNs construct an 'internal model' of any possible perturbation or disturbance to the system, thereby allowing the CRN to implement a well-known engineering

[1]www.singular.uni-kl.de

strategy known as integral control. By solving these two problems, we are able to concretely characterise the space of all possible RPA-capable CRNs - a result of fundamental importance to understanding the basic structures of biochemical networks existing in nature, and to network design in synthetic biology. Indeed, our method for detecting RPA capacity directly calculates the 'set-point' of any RPA-capable CRN as a function of biochemical parameters; moreover, our explicit identification of integral variables by this method provides a roadmap for introducing or altering the RPA property in natural or synthetic bionetworks through design alterations or pharmacological perturbations. This definitive picture of RPA-capable CRNs also delineates the evolutionary constraints governing the creation of larger, more complex RPA-capable CRNs from simpler ones.

S1 Mathematical requirements for RPA in general CRNs

S1.1 Notation and mathematical preliminaries

We consider a collection of chemical reactions, involving n interacting molecules, $X_1, \ldots, X_n$, with corresponding concentrations $x_1, \ldots, x_n$. Each reaction is of the form

$$a_1 X_1 + \ldots + a_n X_n \to \hat{a}_1 X_1 + \ldots + \hat{a}_n X_n, \tag{1}$$

where $a_i, \hat{a}_i \in \mathbb{Z}_{\geq 0}$ are the stoichiometric coefficients of the reactants and products, respectively. In the typical case, $\sum_{i=1}^{n} a_i \leq 2$ for the *reactants* of any reaction, since individual molecules will typically collide pair-wise, sometimes gradually forming large multimolecular complexes prior to a catalytic event or post-translational modification; but in principle, more reactants could contribute to a single reaction, and we make no assumptions *a priori* as to the stoichiometry of reactions.

A standard approach in Chemical Reaction Network Theory (CRNT, see [4] for a detailed primer) is to consider a chemical reaction network (CRN) to consist of three finite sets, $\mathcal{S}$, $\mathcal{C}$,

and $\mathcal{R}$, where $\mathcal{S} = \{X_1, \ldots, X_n\}$ is the set of n interacting molecules, or *species*, of the CRN; $\mathcal{C} = \{y_1, \ldots, y_m\}$ is the set of m distinct *complexes* of the CRN, each of which is a multiset on the species, $\mathcal{S}$; and $\mathcal{R} \subset \mathcal{C} \times \mathcal{C}$ is the set of r distinct reactions of the CRN, which define a directed graph on the complexes, each reaction having the form $y_i \to y_j$.

For analytical purposes, it is convenient to introduce an (arbitrary) order on the species concentrations, setting $\mathbf{x} = [x_1, \ldots, x_n]^T$, which suggests a natural vector representation for each complex $y_k = [a_1, \ldots, a_n]^T \in \mathcal{C}$, where $a_i \in \mathbb{Z}_{\geq 0}$ is the stoichiometric coefficient of the species X_i in the complex y_k. In this way, a dynamical system $f(\mathbf{x}) = \frac{\mathrm{d}\mathbf{x}}{dt}$ can be induced by the CRN, by associating a nonnegative real-valued rate function, $\mathcal{K}_{y_i \to y_j}(\mathbf{x})$ to each reaction $y_i \to y_j \in \mathcal{R}$. In particular, a *mass action* dynamical system is induced when $\mathcal{K}_{y_i \to y_j}(\mathbf{x}) = k_{y_i \to y_j}\mathbf{x}^{y_i}$, where $k_{y_i \to y_j} \in \mathbb{R}_{>0}$ is a strictly positive *rate constant* associated to the reaction $y_i \to y_j$, and $\mathbf{x}^{y_i} = x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}$ for *reactant* complex $y_i = [a_1, \ldots, a_n]^T$.

A CRN thereby induces a dynamical system of the form

$$f(\mathbf{x}) = \frac{d\mathbf{x}}{dt} = \sum_{\mathcal{R}} \mathcal{K}_{y_i \to y_j}(\mathbf{x})(y_j - y_i), \tag{2}$$

which, under the mass action assumption, becomes a polynomial dynamical system of the form

$$f(\mathbf{x}) = \frac{d\mathbf{x}}{dt} = \sum_{\mathcal{R}} k_{y_i \to y_j}\mathbf{x}^{y_i}(y_j - y_i). \tag{3}$$

For mass-action CRNs, it is standard practice to display the rate constant $k_{y_i \to y_j}$ above the corresponding reaction arrow in the chemical reaction representation (given by (1)).

Equation (2) allows for arbitrary kinetics in a CRN, and requires only C^1 smoothness on the component functions $\{f_1, \ldots, f_n\}$. We observe that $\{f_1, \ldots, f_n\} \subset C^1(\mathbb{R}^n)$, the ring of C^1 functions in n real variables. In practice mass-action kinetics, as represented by Equation (3), provides a very general and flexible framework to represent the reaction rates of all possible molecular species in a CRN, including all transient or intermediate multimolecular complexes,

and all chemical moieties and molecular activation states. We also note that all key theorems of CRNT, including the deficiency-zero and deficiency-one theorems [4], along with the Shinar-Feinberg theorem for ACR [10], have appealed to the assumption of mass-action kinetics. Mass-action CRNs have also historically served as a framework for deriving approximations (eg. under various quasi-steady-state assumptions (QSSA)), such as the Michaelis-Menten equation and the Hill equation, to reduce the mathematical complexity of CRNs for analytical purposes. Thus, for the sake of concreteness, and with minimal loss of generality, we focus hereafter on mass-action CRNs, for which $\{f_1, \ldots, f_n\} \subset \mathbb{R}[x_1, \ldots, x_n]$, the ring of polynomials in n variables with real coefficients. In this case, the steady-states of the system constitute a real algebraic variety.

S1.2 Matrix Decomposition of the CRN Mass-Action Equations

For later use, we also recall a useful matrix representation of the mass equations of a CRN, first proposed by Horn and Jackson [8], which extricates the contribution of the CRN's graph structure (considered as a collection of vertices (complexes) and directed edges (reactions)), from the stoichiometry of the reactants and products. As we shall see in Section S4.1, this decomposition is a particularly productive starting point for many important mathematical results in CRNT (see [4]) that are relevant to ACR and RPA.

We consider an arbitrary CRN, $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$, comprising n species, m complexes, and r reactions. As noted earlier in Equation (3), the mass action equations of a CRN take the form

$$f(\mathbf{x}) = \frac{d\mathbf{x}}{dt} = \sum_{\mathcal{R}} k_{y_i \to y_j} \mathbf{x}^{y_i} (y_j - y_i),$$

which can be expressed in the equivalent matrix form

$$\frac{d\mathbf{x}}{dt} = Y.\mathcal{L}(G).\psi(\mathbf{x}). \tag{4}$$

Here $Y : \mathbb{R}^m \to \mathbb{R}^n$ is a linear map that associates to each complex $y \in \mathcal{C}$ its corresponding stiochiometry, ie. its multiset of species, whose matrix columns are thus the vector representations of the m CRN complexes: $Y = [y_1, \ldots, y_m]$.

The map $\mathcal{L}(G) : \mathbb{R}^m \to \mathbb{R}^m$ is the Laplacian of the *graph* of the CRN, with complexes $y \in \mathcal{C}$ taken as vertices, and reactions $y_i \to y_j \in \mathcal{R}$ taken as directed edges. $\mathcal{L}(G)$ thus assumes the matrix form

$$\mathcal{L}(G)_{i,j} = \begin{cases} k_{y_j \to y_i} & \text{if } i \neq j \\ \\ -\sum_{k=1}^{m} k_{y_j \to y_k} & \text{if } i = j. \end{cases}$$

Whereas the entries of Y contain only integers, the entries of $\mathcal{L}(G)$, contain the rate constants of the CRN (and sums of rate constants on its diagonals). The column vector, $\psi(\mathbf{x})$, lists the monomials $\mathbf{x}^{y_i}$ associated to the complexes of the CRN under the mass-action assumption, in an order consistent with the action of the matrix product $Y\mathcal{L}(G)$.

As a minor technicality, Equation (4) includes all complexes of the CRN, including complexes that do not act as a reactant complex for any reaction in the CRN (ie. complexes that represent products only). In this case each column of $\mathcal{L}(G)$ corresponding to such a product-only monomial in $\psi(\mathbf{x})$ contains only zero entries. If convenient, Equation (4) could also be given with these product-only complexes removed from $\psi(\mathbf{x})$, and the corresponding all-zero columns of $\mathcal{L}(G)$ also removed. With slight abuse of notation, noting that the latter version of $\mathcal{L}(G)$ is no longer technically the Laplacian of the CRN graph, either convention may be adopted wherever the explicit consideration of graph structure is not required, since both representations produce the same set of polynomial reaction equations for the network (which contain no product-only monomials).

The matrix decomposition given by Equation (4) explicitly uncouples the graph structure, G, of the CRN (as encoded by $\mathcal{L}(G)$) from the reaction stoichiometry at its vertices (as encoded by Y). In the language of CRNT [4, 10], the connected components of G are referred to as the

linkage classes of the CRN. A strongly connected component (SCC) of G, being a maximal strongly-connected subgraph of G (see [9]), are referred to as *strong-linkage classes* in CRNT. Moreover, a *terminal strong-linkage class* (or *terminal* SCC) is one in which no complex reacts to a complex in a *different* SCC. Complexes belonging to terminal SCCs are referred to as *terminal complexes*; all other complexes are *non-terminal complexes*. This terminology is important to the interpretation of the Shinar-Feinberg theroem [10] - a seminal result on ACR.

Finally, we note a key invariant of any CRN known as the *deficiency*, δ, of the CRN. As we shall see in the theorems we present in Section S4.1, the network deficiency encodes important yet subtle information on the (linear) independence of the chemical reactions in relation to the graph structure of the CRN, and will ultimately give us valuable information on polynomials existing in the *rowspan* of the CRN's reaction equations (cf. the much larger space of polynomials in the *ideal* generated by the rate equations - see Section S1.4). As shown by Feinberg [4], the deficiency of a CRN is easily calculated by the simple formula

$$\delta = m - l - s, \tag{5}$$

where m is the number of complexes in the CRN, l is the number of connected components (linkage classes) of the CRN graph, and s is the dimension of the stoichiometric subspace of the CRN. The stoichiometric subspace is the span of all reaction vectors in the CRN. Thus, s is numerically equal to the number of linearly-independent reaction vectors in the CRN (see [10] and [4]).

S1.3 RPA vs ACR, and the requirement for integral control

Historically, absolute concentration robustness (ACR) has been studied along very separate lines from robust perfect adaptation (RPA). Importantly, RPA corresponds to a special case of a defining problem in classical automatic control – namely, the robust asymptotic tracking

of a desired trajectory (the system's setpoint), while rejecting unwanted disturbances. In the 1970s, the landmark studies of Francis and Wonham [7, 6] investigated the necessary controller structures to achieve such robust tracking, and established what is now known as the internal model principle (IMP). By this principle, a control system is able to reject exogenous stimuli or disturbances by incorporating within itself a model of the dynamic structure of the stimulus or disturbance. In the face of persistent (constant) disturbances, the internal model must produce constant signals and is equivalent to the requirement for integral control [12].

The study of ACR, by contrast, has largely been studied from the CRNT viewpoint, with little consideration of its connection to control theory, although it has recently been acknowledged that ACR is equivalent to the control theoretic notion of robustness to disturbances in initial conditions [3]. Moreover, ACR has typically been considered in the context of *mass-conservative* CRNs [10, 13], where all constituent molecules are subject to a set of conservation laws, with no transfer of molecules into, or out of, the system, and no net production or degradation of molecules. Mass-conservative CRNs therefore have no *external stimuli* or *inputs*, and can only be perturbed by altering the total abundances (or concentrations) of the constituent molecules - ie. by altering the initial conditions. By contrast, RPA is typically considered in models that are subjected to some external input stimulus, or disturbance, that is distinct from the model variables; in addition, these models often incorporate synthesis and degradation processes, and may not therefore be mass-conservative.

But RPA and ACR both require at least one molecule, or species, to have a steady-state concentration that is 'fixed' in the sense of being independent of any external stimulus/disturbance (in the case of RPA) or on total concentrations of molecules (in the case of ACR). In this way, both RPA and ACR have a common mathematical definition; the two phenomena may be distinguished mathematically by the parameters that govern the system's setpoint. For the general case of RPA, the setpoint may be a function of biochemical rate constants, produc-

tion/degradation rates, and even the total abundance of constituent molecules (if this quantity is held fixed due to a conservation law). For ACR, on the other hand, the setpoint may only be a function of biochemical rate constants (see Equation (3)), not production/degradation rates, and not total abundances of any constituent molecules. Thus, ACR may be considered a special case of RPA. All classes of RPA, including ACR, thus require integral control [3] – a requirement that can readily be met in engineering design problems by incorporating special components as integral-computing controllers. But signaling networks that evolve in living systems are dynamically assembled via the physical interactions – involving collisions, binding events, and chemical modifications – among discrete entities, or molecules, which must constitute both the signals and their own controllers. Until now, integral-computing collections of molecular interactions have only been identified in exceedingly simple chemical reaction networks (CRNs), such as the antithetic integral control motif [2], where the requisite integral can be identified via a linear change of coordinates [3, 13].

S1.4 The Two-Variable Kinetic Pairing Theorem

As suggested by the discussion in the previous section, RPA (and hence ACR) is a deeply geometric concept, in the sense that the steady-state concentration of at least one species remains fixed, so that the algebraic variety associated to the system is entirely parallel to a coordinate axis or (hyper)plane for any choice of system parameters, and for any system disturbance. This being the case, the capacity for RPA may be detected through the geometric *projection* of the system onto a certain subset of the model species. But for which species subsets is such a projection possible, and which geometric projections reveal the capacity for RPA?

As a prelude to proving our first important claim as to the properties of all RPA-capable CRNs (Theorem 1) we first consider what constitutes a *variable* of the CRN.

Definition 1. Consider a mass-action CRN comprising n species with concentrations $x_1, \ldots, x_n$.

Each species contributes to a *variable* of the CRN, which can take one of two possible forms:

1. A monomial with the special property that *all* constituent factors (species) *only* appear in the CRN's mass-action equations as factors of this monomial. We refer to monomials with this property as *boundary variables* of a CRN; the presence of this class of variable in a mass-action CRN corresponds to the existence of *boundary states* (see Remark below).

2. Any species that does not contribute to a boundary variable.

Remark. Generally, we consider the individual species of the CRN to be the *variables* of the model (case (2) above). But the geometric projection of the algebraic variety associated to a mass-action system onto a subset of variables corresponds algebraically to an elimination process. This being the case, if a particular collection of species *only* appear together in a particular monomial, and in no other form, then no single such species can be eliminated without also eliminating all other species in the collection. The monomial itself must therefore be considered as a single *unit* from the point of view of elimination (and thus, geometric projection). Likewise, a non-zero value at steady-state for any one species in a boundary variable does not guarantee that the system steady-state is contained in the strictly-positive orthant; all members of boundary variables requires non-zero steady-state values to constrain the system state to the positive orthant (hence *boundary* variable). We consider a CRN containing a boundary variable in Example 2 in Section S3.

We now proceed to the central concepts underpinning Theorem 1. Let us first consider three key ideals in $\mathbb{R}[x_1, \ldots, x_n]$ of special importance to RPA:

1. I_f, the set of all polynomial consequences of the rate equations for the n interacting molecules. I_f thus contains all polynomials that can be 'reached' by the rate equations in the sense that each polynomial $p \in I_f$ can be expressed in the form $p = h_1 f_1 + \ldots + h_n f_n$, for some $h_1, \ldots, h_n \in \mathbb{R}[x_1, \ldots, x_n]$.

2. I_p, the set of all polynomials in the ring $\mathbb{R}[x_1, \ldots, x_n]$ that *vanish* at $x_i = c$, for some variable x_i of the CRN, and some $c \in \mathbb{R}_{>0}$.

3. $I_f \cap I_p$, being the ideal which contains all polynomials in the n variables $x_1, \ldots, x_n$ that vanish at $x_i = c$ *and* can be 'reached' by the rate equations $f_1, \ldots, f_n$.

Let us now introduce an additional, fourth, ideal associated with the system $\{f_1, \ldots, f_n\} \subset \mathbb{R}[x_1, \ldots, x_n]$, which will assist in finding generators of $I_f \cap I_p$, by using the generators of I_f:

4. I_e, an elimination ideal, being the set of all polynomial consequences of $f_1, \ldots, f_n$ that contain *only* the variables x_i and x_j, where x_i is the variable being tested for its RPA capacity, and where x_j is some variable that does *not* have the capacity for RPA.

We will now present a definitive test for RPA (necessary and sufficient conditions, on the assumption of system stability) that is valid for *any* collection of chemical reactions, by showing that:

1. I_e is fully contained in $I_f \cap I_p$ (see Figure S1).

2. I_e is a principal ideal, with $I_e =< g(x_i, x_j)(x_i - c) >$, where $g(x_i, x_j) \in \mathbb{R}[x_i, x_j]$ is non-vanishing on $(0, X_{itot}) \times (0, X_{jtot})$. X_{ktot} is the maximum possible concentration for the species x_k; in a mass-conservative system X_{ktot} is thus the *total* abundance of all forms of a particular molecule that include x_k as a particular form.

3. The set of polynomials $\{g(x_i, x_j)(x_i - c)$: x_j is a variable that *does not* exhibit RPA$\}$ is a generating set for $I_f \cap I_p$.

Theorem 1 (Two-variable kinetic pairing theorem)**.** Consider a chemical reaction network endowed with polynomial (eg. mass-action) kinetics comprising n species with concentrations $x_1, \ldots, x_n$ and corresponding reaction rate functions $\{f_1, \ldots, f_n\} \subset \mathbb{R}[x_1, \ldots, x_n]$.

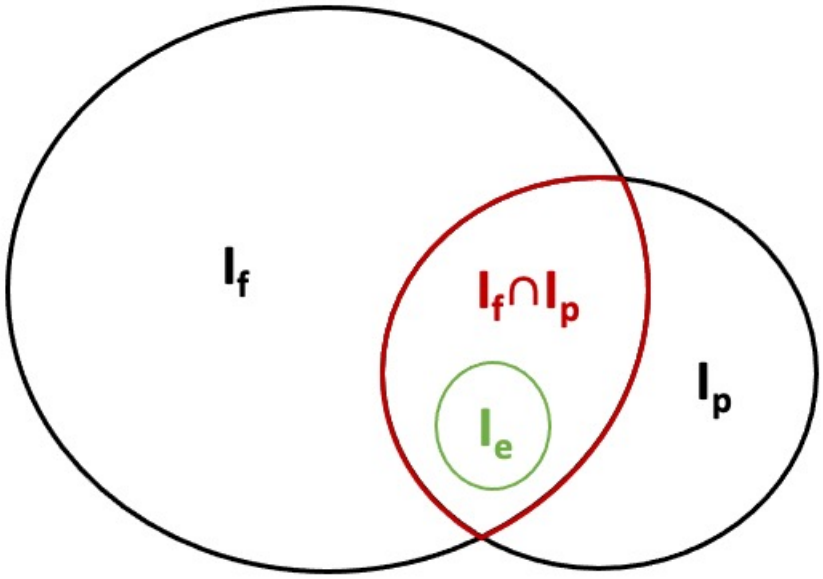

Figure S1: An RPA-capable system of polynomial reaction rates is characterised by a special elimination ideal, I_e, that is fully contained in $I_f \cap I_p$.

The system has the capacity for RPA in the variable x_i if and only if the elimination ideal $<f_1, \ldots, f_n> \cap \mathbb{R}[x_i, x_j]$ is a principal ideal generated by a polynomial of the form $g(x_i, x_j)(x_i - c)$, where x_j is any variable of the system that does *not* exhibit RPA, $g(x_i, x_j)$ is a polynomial that is non-vanishing on $(0, X_{itot}) \times (0, X_{jtot})$, and c is a rational function of system parameters (called the RPA 'setpoint'). Moreover, if c depends only on biochemical rate constants and not on total concentrations (X_{ktot}) or production/degradation rates of any of the molecules, the system also has the capacity for ACR in the variable x_i with setpoint c.

Proof. A system exhibits RPA in a variable x_i exactly when x_i maintains a constant steady-state value, say c, for all (positive) steady-states of the system. A CRN therefore has the capacity for RPA *only if* there exists some set of polynomials $\{h_1, \ldots, h_n\} \subset \mathbb{R}[x_1, \ldots, x_n]$ such that

$$h_1 f_1 + \ldots + h_n f_n = \hat{p}\,(x_i - c)$$

where $\hat{p} \in \mathbb{R}[x_1, \ldots, x_n]$ is a polynomial that does not vanish in the positive orthant, except possibly at $x_i = c$, and where c is a rational function of the parameters in the rate equations $f_1, \ldots, f_n$.

We first consider the ideal $I_f =< f_1, \ldots, f_n >= \{h_1 f_1 + \ldots + h_n f_n : h_i \in \mathbb{R}[x_1, \ldots, x_n]\}$, and the ideal $I_p =< (x_i - c) >$. If the ideal $I_f \cap I_p$ contains non-zero polynomials, it must have non-zero generators. To establish one such generator, we consider the possible subsets $\bar{x} \subset \{x_1, \ldots, x_n\}$ with the property that $I_f \cap \mathbb{R}[\bar{x}]$ is fully contained in $I_f \cap I_p$. Which subsets $\bar{x}$ can have this property? It is clear that $\bar{x}$ must include x_i, the candidate RPA variable. Suppose $\bar{x}$ contains some other variable x_k, that also exhibits RPA (at $x_k = c'$, say, where c' is a rational function of system parameters, generally distinct from c) - assuming such a variable exists in the CRN in question. In this case, $I_f \cap \mathbb{R}[\bar{x}]$ will contain polynomials in both x_i and x_k that are not contained in $I_f \cap I_p$ - eg. polynomials that vanish at $x_k = c'$. Now suppose instead that $\bar{x}$ contains x_i and more than one variable (say x_j and x_m) that *do not* have the capacity for RPA. Since a perturbation to the CRN that alters the steady state of x_j will also alter the steady-states of other non-RPA capable variables (eg. x_m), $I_f \cap \mathbb{R}[\bar{x}]$ will contain polynomials in x_j and x_m, and that are not contained in $I_f \cap I_p$.

Let us therefore choose $\bar{x}$ to contain *one* non-RPA-capable variable (x_j, say) in addition to x_i. The set $\bar{x}$ now contains two *independent* (uncoupled) variables in the sense that a perturbation to the CRN that alters the steady-state of one of the variables does not affect the steady-state of the other. It follows that $I_f \cap \mathbb{R}[x_i, x_j]$ contains only polynomials that are of the form $p.g(x_i, x_j)(x_i - c)$, where $p \in \mathbb{R}[x_i, x_j]$, and $g \in \mathbb{R}[x_i, x_j]$ is non-vanishing on $(0, X_{itot}) \times (0, X_{jtot})$, other than possibly at $x_i = c$. That is, $I_f \cap \mathbb{R}[x_i, x_j]$ is a principal ideal generated by a polynomial of the form

$$\rho = g(x_i, x_j)(x_i - c).$$

Since this argument holds for *any* variable x_j that does not have the capacity for RPA, it follows that $I_f \cap I_p$ is generated by the set of polynomials

$$\{g(x_i, x_m)(x_i - c) : x_m \text{ a non-RPA-capable variable}\}.$$

Thus a CRN has the capacity for RPA at the variable x_i only if $I_f \cap \mathbb{R}[x_i, x_j]$ is a principal ideal generated by $g(x_i, x_j)(x_i - c)$, for $g(x_i, x_j) \neq 0$ on $(0, X_{itot}) \times (0, X_{jtot})$, except possibly at $x_i = c$, for *any* variable x_j that does not have the capacity for RPA.

The converse is straightforward. If, for any variable x_j that does not have the capacity for RPA, $< f_1, \ldots, f_n > \cap \mathbb{R}[x_i, x_j]$ is generated by a single polynomial of the form $g(x_i, x_j)(x_i - c)$, it follows that $g(x_i, x_j)(x_i - c)$ vanishes at the system's steady-state. If $g(x_i, x_j) \neq 0$ for all possible steady-states of the system, other than possibly $x_i = c$, it follows that $x_i = c$ for all possible steady-states. Hence the CRN has the capacity for RPA at x_i, with setpoint c.

If c is a function only of biochemical rate constants, and does not depend on any total molecular concentrations of the system, the above argument can be applied with all instances of the term RPA replaced by the term ACR. □

Definition 2 (RPA polynomial). A polynomial in two variables, x, y, of the form

$$\rho = g(x, y)(x - c),$$

with $c \in \mathbb{R}$ and $g(x, y) \neq 0$, is called an **RPA polynomial**. Moreover, a CRN is said to *contain an RPA polynomial* exactly when an $(n-2)^{th}$ elimination ideal of the CRN's mass-action equations can be generated by an RPA polynomial (in which case the CRN has the capacity for RPA in the variable x).

Definition 3 (RPA-capacity vs RPA-exhibiting). A CRN has the capacity for RPA in the variable x, with setpoint $x = c$, if it contains an RPA polynomial, $\rho = g(x, y)(x - c)$. If,

in addition, $x = c$ is a *stable* steady-state of the CRN's mass-action equations, the CRN also *exhibits* RPA in the variable x.

Remark. 1. The fact that the key elimination ideal in Theorem 1 is generated by a polynomial in a specific number of variables, namely *two*, is key. Previous consideration of the concept of *constrained integral control* [3, 13] has posited the notion of a *constrained integrator* [3] of the form $r(t)(x - c)$, where x is the RPA variable, c its setpoint, and $r(t)$ is *some* function of system variables. But identifying that we must project the system onto exactly two variables immediately converts the question of RPA capacity to a well-defined geometric projection problem, and hence an algebraic elimination problem, for *any* CRN.

2. Although it may not be clear *a priori* which variables to select for these two special choices, x_i and x_j, in an algorithmic test for RPA capacity, we will see in the next section that Theorem 1 has important *topological* implications. Consideration of the topology of the CRN under investigation can suggest a suitable choice for the candidate RPA-capable variable x_i, as well as a judicious choice for the non-RPA-capable variable, x_j. In addition, the topological consequences of Theorem 1 impose a special structure on RPA-capable networks, as we shall explore in detail in Section S2; for this reason, the underlying elimination problem (*ut supra*) can be solved through the computation of the Gröbner basis for the (ideal of the) system's rate equations, despite the fact that the task of computing Gröbner bases (eg. via Buchberger's algorithm) *in general* (cf. for RPA-capable CRNs in particular) is well known to be NP-Hard. Failure of the Buchberger (or other Gröbner basis-computing) algorithm to terminate could thus be adduced as *prima facie* evidence that the CRN under consideration does not, in fact, have the capacity for RPA. In Section S5, we provide code in the open-source software Singular (www.singular.uni-

kl.de) to compute a suitable Gröbner basis, and thus assess RPA-capacity, for all examples considered in this study.

3. The statement of Theorem 1 also encompasses special cases where $g(x_i, x_j)$ is zero order in x_i, or x_j, or both. With judicious choice of x_i and x_j from a topological point of view, as we explore in the next section, we have most commonly obtained a polynomial $g(x_i, x_j)$ that is zero order in x_i, for a wide range of CRNs (some of which are included in the present paper as illustrative examples). Less commonly, we have also found examples of CRNs where $g(x_i, x_j)$ is zero order in *both* x_i *and* x_j (as is the case for the antithetic integral control motif [2] - see Section S1.5).

Before proceeding to an examination of the topological implications of Theorem 1 (Section S2) and its consequences for the implementation of integral control (Section S4), we first illustrate the application of Theorem 1 to the network depicted in Figure S2, and also considered in Figure 1 in the main paper.

S1.5 Example: RPA in a Two-Node Opposing Set Involving Antithetic Integral Control

In Figure S2 we present a collection of chemical reactions, which incorporates a sub-collection of reactions (noted in red) that constitute a motif known as *antithetic integral control*. With only the red reactions considered, the law of mass action would induce two rate equations, $\dot{x_1} = k_1 R - k_2 x_1 x_2$ and $\dot{x_2} = k_3 - k_2 x_1 x_2$, such that $\dot{x_1} - \dot{x_2} = k_1(R - k_3/k_1)$, which is an RPA polynomial with $g(x_1, x_2) = k_1$ (which is zero-order in both x_1 and x_2 in this particularly simple case) and setpoint $c = k_3/k_1$. But with the additional (black) reactions included, it is not immediately clear if RPA will still obtain (at R) for this more extensive collection of reactions. By computing the Gröbner basis (see Singular code provided in Section S5), and

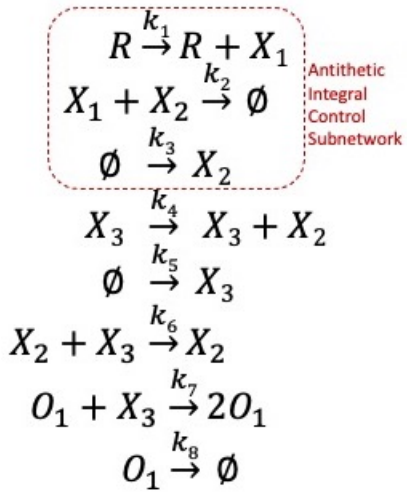

Figure S2: Chemical reaction network incorporating antithetic integral control as a sub-motif

selecting $x_i = R$ and $x_j = x_2$ for the two-variable projection, we obtain the RPA polynomial

$$\begin{aligned}\rho_1 &= k_1^2 k_6 k_7 x_2 R^2 - (2k_1 k_3 k_6 k_7 + k_1 k_4 k_6 k_8) x_2 R + (k_3^2 k_6 k_7 + k_3 k_4 k_6 k_8) x_2 \\ &\quad - k_1 k_4 k_5 k_7 R + (k_3 k_4 k_5 k_7 + k_4^2 k_5 k_8) \\ &= \left(R - \left[\frac{k_3}{k_1} + \frac{k_4 k_8}{k_1 k_7}\right]\right)\left(k_1^2 k_6 k_7 x_2 \left[R - \frac{k_3}{k_1}\right] - k_1 k_4 k_5 k_7\right),\end{aligned}$$

demonstrating that the CRN exhibits RPA at R with setpoint $c = \frac{k_3}{k_1} + \frac{k_4 k_8}{k_1 k_7}$.

To illustrate that we can project onto *any* non-RPA-capable molecule to test for RPA, not just x_2, we repeat the test with $x_j = x_1$ which yields

$$\begin{aligned}\rho_2 &= k_1 k_2 k_4 k_5 k_7 x_1 R - (k_2 k_3 k_4 k_5 k_7 + k_2 k_4^2 k_5 k_8) x_1 - k_1^3 k_6 k_7 R^3 + (2k_1^2 k_3 k_6 k_7 + k_1^2 k_4 k_6 k_8) R^2 \\ &\quad - (k_1 k_3^2 k_6 k_7 + k_1 k_3 k_4 k_6 k_8) R \\ &= k_1 k_7 \left(R - \left[\frac{k_3}{k_1} + \frac{k_4 k_8}{k_1 k_7}\right]\right)\left(k_2 k_4 k_5 x_1 - k_1^2 k_6 R \left[R - \frac{k_3}{k_1}\right]\right),\end{aligned}$$

and with $x_j = O_1$, which yields

$$\begin{aligned}\rho_3 &= k_1 k_7 O_1 R - (k_3 k_7 + k_4 k_8) O_1 \\ &= k_1 k_7 \left(R - \left[\frac{k_3}{k_1} + \frac{k_4 k_8}{k_1 k_7} \right] \right).\end{aligned}$$

With the choice of $x_j = x_3$ (and $x_i = R$), we find that an RPA polynomial no longer obtains, demonstrating that this CRN has the capacity for RPA at x_3 as well as R. Indeed, repeating the test with $x_i = x_3$, and $x_j \in \{x_2, x_1, O_1\}$, we obtain, respectively,

$$\begin{aligned}\rho_4 &= k_6 k_7 x_2 x_3^2 - k_6 k_8 x_2 x_3 - k_5 k_7 x_3 + k_5 k_8 \\ &= k_7 \left(k_6 x_2 x_3 - k_5 \right) \left(x_3 - \frac{k_8}{k_7} \right),\end{aligned}$$

$$\begin{aligned}\rho_5 &= k_2 k_5 k_7 x_1 x_3 - k_2 k_5 k_8 x_1 - k_4 k_6 k_7 x_3^3 + \left(k_4 k_6 k_8 - k_3 k_6 k_7 \right) x_3^2 + k_3 k_6 k_8 x_3 \\ &= k_7 \left(x_3 - \frac{k_8}{k_7} \right) \left(k_2 k_5 x_1 - k_6 x_3 (k_4 x_3 + k_3) \right),\end{aligned}$$

and

$$\begin{aligned}\rho_6 &= k_7 O_1 x_3 - k_8 O_1 \\ &= k_7 O_1 \left(x_3 - \frac{k_8}{k_7} \right).\end{aligned}$$

Each of these three RPA polynomials demonstrates that the setpoint for x_3 is $\frac{k_8}{k_7}$. Thus, for RPA variable R, the ideal $I_f \cap I_p =< \rho_1, \rho_2, \rho_3 >$. For RPA variable x_3, $I_f \cap I_p =< \rho_4, \rho_5, \rho_6 >$.

S2 The RPA Polynomial and the Topological Hierarchy of Invariants

In the previous section we proved that all CRNs that are capable of exhibiting RPA at a particular variable (x_i) are characterised by an elimination ideal in *exactly two* variables (x_i and x_j) that

can be generated by a single polynomial, called an RPA polynomial, with the particular form $\rho = g(x_i, x_j)(x_i - c)$. Moreover, we proved that if a CRN has the capacity for RPA at variable x_i, a corresponding RPA polynomial exists for *every* variable x_j that does *not* have the capacity for RPA.

We now consider the consequences of Theorem 1 for the underlying network structure, or topology, of RPA-capable CRNs, along with the critical question of making suitable choices for x_i and x_j in an algorithmic test for RPA-capacty.

First, since $\rho = 0$ for any network steady-state, the existence of an RPA polynomial in a CRN's mass-action equations implies that the activating (or *upregulating*), influence of the non-RPA-capable variable (x_j) is exactly matched, or *paired*, with its inhibitory (or *downregulating*), influence on the RPA-variable, x_i, via the function g. We refer to this feature of the RPA polynomial, and the regulation of the underlying CRN, as *kinetic pairing*. With this observation in mind, we note that there are two network (ie. topological) interpretations of ρ: (i) x_j is a regulator of x_i; or (ii) x_j is regulated by x_i. See Figure S3.

From this, the topological significance of the RPA polynomial is clear: If x_j is a *regulator* of x_i, and its upregulating and downregulating influences on the regulated variable x_i are exactly *paired* as described above, this implies an overarching network structure that corresponds to a Balancer module (see Figure (S3B), where x_j plays the role of a *diverter variable*, and x_i (the RPA variable) is a *connector variable*). See [1] for a comprehensive explanation of Balancer modules. Likewise, if x_j is *regulated* by x_i, and the upregulating and downregulating influences on the regulating (RPA) variable x_i are exactly *paired* as described above, this implies an overarching network structure that corresponds to a Opposer module (see Figure (S3A), where x_j plays the role of an *opposer variable*, and x_i (the RPA variable) is the unique regulator of the Opposer module). See [1] for a comprehensive explanation of Opposer modules. Indeed, we note the striking similarity between an RPA polynomial, ρ, and both *connector kinetics* (which

are required in 'Balancer Modules' [1]) and *opposer kinetics* (which are required in 'Opposer Modules') as presented in our previous general solution to the RPA problem at the network macroscale (see [1]).

While, in practice, some trial and error may be necessary in the selection of elimination variables for highly complex CRNs, these topological observations suggest natural choices for x_i and x_j. In particular, if the CRN is topologically of Balancer-type, any molecule with the function of a diverter variable is a judicious choice for x_j, the non-RPA-capable variable, while a molecule regulated by x_j via a number of parallel pathways (ie. a connector molecule) should be assigned to x_i as the candidate RPA-capable variable. By contrast, if the CRN is topologically of Opposer type, a molecule acting as the sole regulator of the collection of opposer molecules is a suitable choice for x_i, while any member of a collection of opposer molecules makes a suitable choice for x_j.

These principles will be clarified through the analysis of a range of additional examples in the remainder of this Supplement. In particular, we will show that the elimination ideal associated to a CRN's RPA polynomial has a very special structure, which reveals a universal implementation of *integral control* for all RPA-capable CRNs.

S3 RPA Polynomials in the Rowspan of a CRN

Given that all RPA-capable CRNs have a two-variable RPA-polynomial, ρ, in the ideal associated to the system's rate equations - ie. $\rho = h_1 f_1 + \ldots + h_n f_n$, for some $\{h_1, \ldots, h_n\} \subset \mathbb{R}[x_1, \ldots, x_n]$ - we now wish to clarify the algebraic, and ultimately topological, significance of a set of h_i's that can perform the elimination task required for the system to 'reach' ρ. We first consider a preliminary question: Under what circumstances are $h_i \in \mathbb{R}$? That is, when does an RPA-capable CRN possess an RPA polynomial in the rowspan of the system, where ρ is an $\mathbb{R}$-linear combination of the polynomial rate equations?

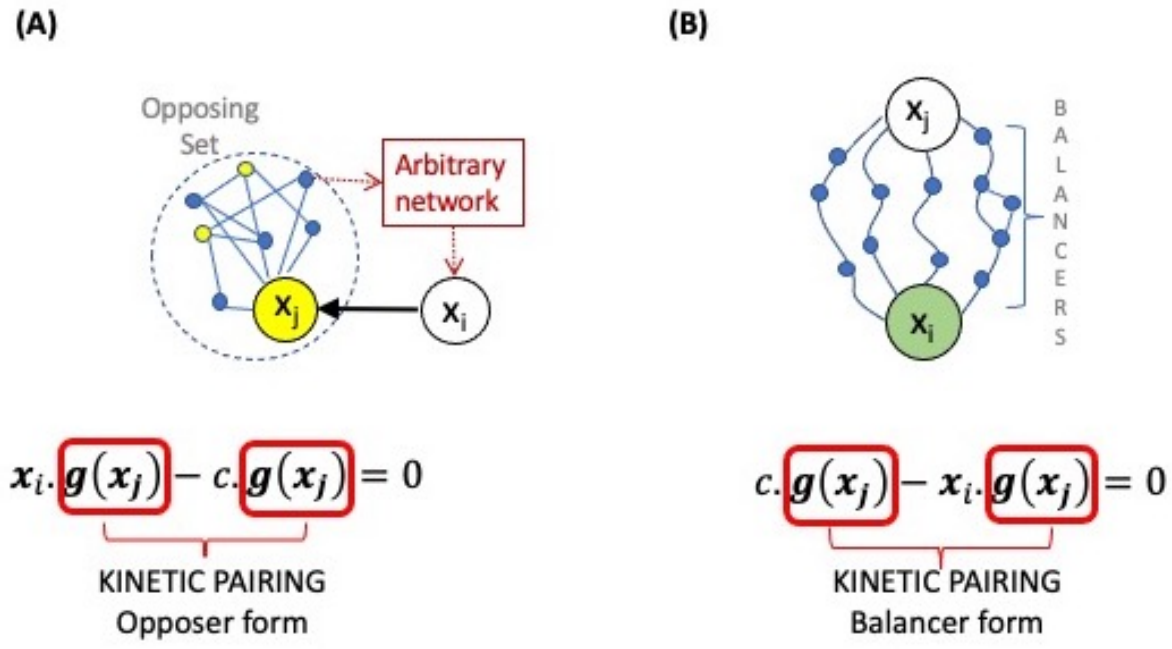

Figure S3: (Presented as Figure 3 in the main paper). The principle of **kinetic pairing**, via a *pairing function* $g(x_i, x_j)$. For clarity, we depict a functional form for g that is zero order in x_i, since this is the most common form obtained in practice (i.e. provided the RPA variable x_i is not autoregulatory). (a) Kinetic pairing in an Opposer module: Here the *regulator* molecule, x_i, has the capacity for RPA, while the influence of the *regulated* molecule, x_j, which does not exhibit RPA, is balanced (paired) in both its upregulatory and downregulatory contributions. (b) Kinetic pairing in a Balancer module: In this case, the *regulated* molecule, x_i, has the capacity for RPA, while the influence of the *regulating* molecule, x_j, which does not exhibit RPA, is paired in both its upregulatory and downregulatory contributions. Full details on Opposer and Balancer modules, at the level of the network macroscale, are given in [1].

To address this question, we first examine two different RPA-capable CRNs whose RPA polynomial is *not* in the rowspan of the system. In each case, we consider a simple deficiency-preserving modification to the network to allow the same RPA polynomial to exist in the system's rowspan. These examples are sufficiently simple for the elimination process to be executed manually, but are sufficiently involved to point to deep and general properties of the elimination problem that exists in all RPA-capable CRNs.

S3.1 Example 1

Cappelletti et al. [3] discuss an ACR-capable toy model comprising biochemical interactions among five molecules, A, B, C, D and E, for which it was shown that no $\mathbb{R}$-linear combination of the system's rate equations can take the form $r.(A^\gamma - q)$, for some polynomial $r \in \mathbb{R}[A, B, C, D, E]$ and some $\gamma, q \in \mathbb{Q}_{>0}$, even though the system could be shown to exhibit ACR in the molecule A.

We depict this toy model in Figure S4, and observe from the set of chemical reactions that the species B inhibits A by direct binding and conversion to C (reaction at rate k_1). In addition, B regulates the production of C (reactions at rates k_2 and k_3), while C in turn produces A (reaction at rate k_7). Thus B both upregulates A (via C) and downregulates A, such that the collection of chemical reactions is topologically a Balancer Module. We depict this relationship schematically in Figure S4 (center). We further note that B also regulates E, which in turn regulates D, as indicated in Figure S4; from our prior comprehensive analysis of RPA-permissive network topologies [1], it is clear that these additional interactions involving D and E do not play any role in the RPA capacity of the system, but are still entirely compatible with the topology of a Balancer Module (and hence do not interfere with the RPA capacity of the system).

In Figure S4 (right) we present a modified version of the Cappelletti model [3] - one which preserves the topological structure of the CRN as depicted in Figure S4 (center), but is constructed from a slightly different collection of chemical reactions. We note that both the original (Figure S4 left) and modified (Figure S4 right) versions of the model have a deficiency of two. Thus, their capacity for ACR cannot be ascertained from the Shinar-Feinberg theorem (see Theorem 2 in Section S4.1), which is applicable *only* to CRNs of deficiency *one*.

Under the law of mass action, both the original and modified models have identical reaction

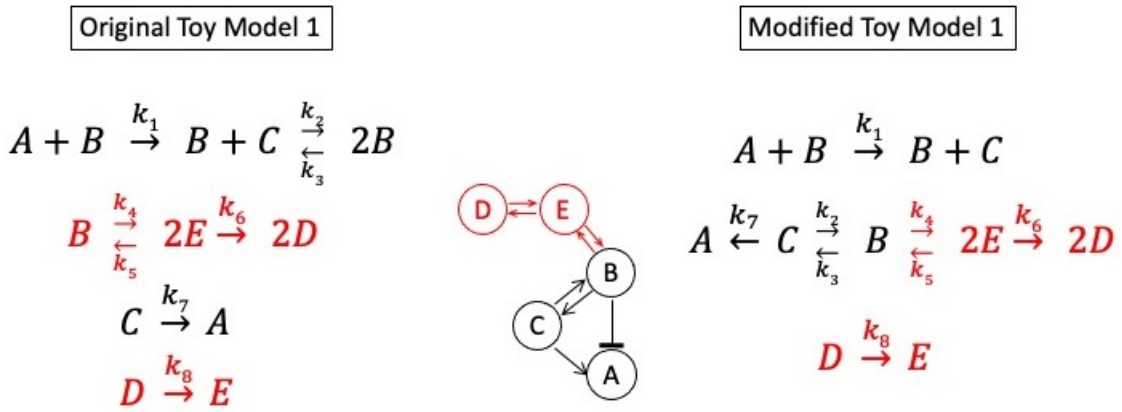

Figure S4: A toy CRN proposed by [3] (left) along with a modified version (right). Both CRNs have the underlying topology of a Balancer Module, and thus have the potential for RPA capacity at A. Both CRNs have a deficiency of two, and thus elude the Shinar-Feinberg theorem [10].

rates for species A, D, and E, namely

$$\dot{A} = k_7 C - k_1 AB, \tag{6}$$

$$\dot{D} = 2k_6 E^2 - k_8 D, \tag{7}$$

$$\dot{E} = k_8 D - 2(k_5 + k_6)E^2 + 2k_4 B. \tag{8}$$

For the **original model** (Figure (S4, left)), the reaction rates for B and C are

$$\dot{B} = k_2 BC - k_3 B^2 + k_5 E^2 - k_4 B, \tag{9}$$

$$\dot{C} = k_1 AB - k_7 C - k_2 BC + k_3 B^2, \tag{10}$$

while for the **modified model** ((Figure (S4, right)), the reaction rates for B and C are, instead,

$$\dot{B} = k_2 C - k_3 B + k_5 E^2 - k_4 B, \tag{11}$$

$$\dot{C} = k_1 AB - k_7 C + k_3 B - k_2 C. \tag{12}$$

Since both CRN models are topologically of balancer type, with A as the connector species [1],

they have the *potential* to be RPA-capable at species A. Whether or not they *do*, in fact, have the capacity for RPA depends on the intricate molecular details of the CRN, leading to the presence or absence of an RPA polynomial in the ideal generated by the system equations. Since B is topologically a diverter species [1], it is natural to project the system onto the two species A and B, to look for an RPA polynomial of the form

$$\rho = g(B)(A - \kappa). \tag{13}$$

The balancing nature of both models, whereby B regulates A directly, and indirectly via C, is captured by the equation for $\dot{A}$ (Equation (6)), from which it is clear that if C is proportional to B at steady-state, then the system will contain an RPA polynomial in its steady-state ideal.

That both models are indeed characterised by a steady-state concentration of C in direct proportion to that of B is revealed by the polynomial $(\dot{A} + \dot{C})$. Indeed, for the modified model,

$$\dot{A} + \dot{C} = k_3 B - k_2 C. \tag{14}$$

From this it is clear that an RPA polynomial exists *in the rowspan* of the modified model, since the variable to be eliminated (C) exists as a linear monomial in both (6) and (14), which allows these two equations to be combined through linear combination to yield the needed RPA polynomial, ie.

$$k_2\dot{A} + k_7(\dot{A} + \dot{C}) = k_2 k_7 C - k_1 k_2 AB + k_7 k_3 B - k_2 k_7 C \tag{15}$$

$$= -k_1 k_2 B \left(A - \frac{k_7 k_3}{k_1 k_2} \right), \tag{16}$$

indicating that, for $B^* \neq 0$, $A^* = (k_7 k_3)/(k_1 k_2)$. (Here and elsewhere in this Supplement, we indicate steady-states by a superposed asterisk.) In this way, we 'pass' the invariant given by (14) to the invariant given by (6) through the operation of linear combination. The RPA polynomial and its two component invariants are linearly dependent. In fact, if we simply

remove the 'unnecessary' (from an RPA viewpoint) reactions from this CRN (indicated in red in Figure (S4, right), the remaining CRN has deficiency one, and can be shown to exhibit ACR (at A) by the Shinar-Feinberg theorem.

For the original model, on the other hand,

$$\dot{A} + \dot{C} = -k_2BC + k_3B^2 = -k_2B\left(C - \frac{k_3}{k_2}B\right), \tag{17}$$

so $C^* = (k_3/k_2)B^*$ - a relationship of direct proportion - so that an RPA polynomial of the form (13) can exist in the ideal of the system. But in this case, no *linear* combination of $\dot{A}$ and $(\dot{A}+\dot{C})$ can eliminate C to construct the needed RPA polynomial. This is because in Equation (17), C is contained only in the monomial BC, whereas in Equation (6), C exists only as a linear monomial. Reconciling the two invariants, then, to construct the RPA polynomial requires a 'concatenating monomial': Equation (6) must first be multiplied by the linear monomial B to produce

$$B\dot{A} = -k_1AB^2 + k_7BC. \tag{18}$$

Equation (18) clearly resides in the ideal generated by the CRN's reaction equations, but not in the rowspan, and now incorporates C by the monomial BC as in the invariant (17).

Now (17) and (18) can be combined through linear combination to produce the requisite RPA polynomial,

$$\begin{aligned} k_2B\dot{A} + k_7(\dot{A} + \dot{C}) &= -k_1k_2AB^2 + k_2k_7BC - k_2k_7BC + k_7k_3B^2, \\ &= -k_1k_2AB^2 + k_7k_3B^2, \\ &= -k_1k_2B^2\left(A - \frac{k_7k_3}{k_1k_2}\right), \end{aligned}$$

Thus, for the original model, we 'pass' the balancer invariant $(\dot{A}+\dot{C})$ to the connector invariant $\dot{A}$ by first applying a concatenating monomial (B) to $\dot{A}$ in order for the elimination step to be

executed.

The key distinction between the two models, from the point of view of the RPA-promoting structure of their reaction graphs, and thus of their rate equations, is one of stoichiometry. For the original model, the balancer polynomial (Equation (17)) incorporates the species C from the reaction $B + C \to 2B$, whereas for the modified model, the balancer polynomial (Equation (14)) incorporates C from the reaction $C \to B$. Thus, since C appears in the connector polynomial (Equation (6)) from the reaction $C \to A$, the original model requires a concatenating monomial (B, applied to the connector polynomial) to reconcile the stoichiometries of the reactions contributing to its balancer and connector polynomials. Thus, we say that the balancer and connector polynomials for the original model are *stoichiometrically independent*. No such concatenating monomial is required for the modified model, since the reactions contributing to its balancer and connector polynomials are already stoichiometrically compatible from an elimination standpoint.

S3.2 Example 2

In the Supplementary Information to [10], Shinar and Feinberg present a CRN of an EnvZ-OmpR motif of deficiency two (see Figure S5 (left), and Figure 2b in our main article) which, unlike its deficiency-one counterparts in the main paper of [10], fall outside the scope of the Shinar-Feinberg theorem. Nevertheless, the model is known to exhibit ACR in the species Y_p.

Shinar and Feinberg's deficiency-two model of the EnvZ-OmpR CRN is represented in Figure S5 (left), with each of the nine species $\{XT, Y_p, XD, X, X_p, Y, X_pY, XTY_p, XDY_p\}$ denoted by $\{x_1, \ldots, x_9\}$, respectively, in order to simplify the rate equations. As depicted schematically in Figure S5 (center), this ACR-capable network is, topologically-speaking, a Balancer Module. In particular, XD and XT (and hence X) are, topologically, diverter species [1]. For the sake of concreteness, we select XT as the diverter species: XT directly down-

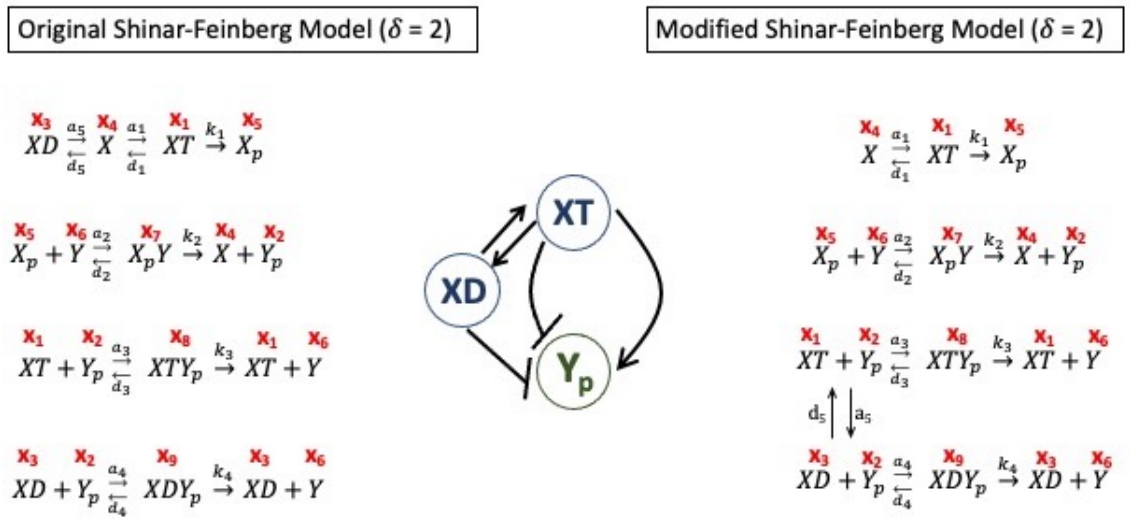

Figure S5: A deficiency-two CRN of the EnvZ(X)-OmpR(Y) osmoregulation system studied by Shinar and Feinberg [10] (left). As shown (centre) this CRN is, topologically, a Balancer Module. A modified (hypoethetical) version of the CRN is also presented (right), which preserves both topology and deficiency.

regulates Y_p through direct binding and enzymatic conversion back to Y. XT also indirectly downregulates Y_p via XD (a balancer species) through the binding of XD to Y_p and enzymatic conversion to Y. In addition, X_T upregulates Y_p via X_p (a balancer species), with the binding of X_p to Y catalysing the transfer of a phosphate group to produce Y_p. A similar argument holds when selecting XD or X as a candidate diverter species instead, in which case XT will function as a balancer species. In any case, there are three parallel pathways leading from the diverter species to Y_p (the connector species [1]), conferring the potential for RPA (and in this case ACR, more specifically) to this collection of reactions.

Under the law of mass action, this chemical reaction graph structure induces the following

nine rate equations:

$$f_1 = \dot{XT} = \dot{x_1} = a_1 x_4 - (d_1 + k_1)x_1 - a_3 x_1 x_2 + (d_3 + k_3)x_8, \tag{19}$$

$$f_2 = \dot{Y_p} = \dot{x_2} = k_2 x_7 - a_3 x_1 x_2 + d_3 x_8 - a_4 x_3 x_2 + (k_4 + d_4)x_9, \tag{20}$$

$$f_3 = \dot{XD} = \dot{x_3} = d_5 x_4 - a_5 x_3 - a_4 x_3 x_2 + (k_4 + d_4)x_9, \tag{21}$$

$$f_4 = \dot{X} = \dot{x_4} = a_5 x_3 + d_1 x_1 - (a_1 + d_5)x_4 + k_2 x_7, \tag{22}$$

$$f_5 = \dot{X_p} = \dot{x_5} = k_1 x_1 - a_2 x_5 x_6 + d_2 x_7, \tag{23}$$

$$f_6 = \dot{Y} = \dot{x_6} = d_2 x_7 - a_2 x_5 x_6 + k_3 x_8 + k_4 x_9, \tag{24}$$

$$f_7 = \dot{X_p Y} = \dot{x_7} = a_2 x_5 x_6 - (d_2 + k_2)x_7, \tag{25}$$

$$f_8 = \dot{XTY_p} = \dot{x_8} = a_3 x_1 x_2 - (d_3 + k_3)x_8, \tag{26}$$

$$f_9 = \dot{XDY_p} = \dot{x_9} = a_4 x_2 x_3 - (d_4 + k_4)x_9. \tag{27}$$

Notice that both x_5 and x_6 *only* appear in this system of equations as components of the monomial $x_5 x_6$, which thereby acts as a single 'unit', and is therefore a boundary variable of the model. There are thus eight variables in this model: $x_1, x_2, x_3, x_4, x_5 x_6, x_7, x_8, x_9$.

Now consider our **modified version** of Shinar and Feinberg's model in Figure S5 (right). Like Shinar and Feinberg's original model (Figure S5 (left)), this is topologically a Balancer Module, with XT, XD and X all potential diverter species, and Y_p the connector species. Under the law of mass action, f_2, f_5, f_6, f_7, f_8 and f_9 for the modified system are identical to their counterparts in the original model. By contrast, f_1, f_3 and f_4 are now altered in the modified system, and take the forms

$$f_1 = \dot{XT} = \dot{x_1} = a_1 x_4 - (d_1 + k_1)x_1 - (a_3 + d_5)x_1 x_2 + (d_3 + k_3)x_8 + a_5 x_3 x_2, \tag{28}$$

$$f_3 = \dot{XD} = \dot{x_3} = d_5 x_1 x_2 - (a_5 + a_4)x_2 x_3 + (k_4 + d_4)x_9, \tag{29}$$

$$f_4 = \dot{X} = \dot{x_4} = d_1 x_1 - a_1 x_4 + k_2 x_7. \tag{30}$$

An examination of the rate equation f_2 reveals the balancing nature of this motif in both the original and modified forms:

$$f_2 = \underbrace{k_2 x_7}_{\substack{\text{upregulation by } XT \\ \text{via } X_p \text{ and } X_p Y}} \quad \underbrace{-a_3 x_1 x_2 + d_3 x_8}_{\substack{\text{downregulation by} \\ \text{conversion to } Y \text{ by } XT}} \quad \underbrace{-a_4 x_3 x_2 + (k_4 + d_4) x_9}_{\substack{\text{downregulation by} \\ \text{conversion to } Y \text{ by } XD}}. \tag{31}$$

We now consider the additional row operations on the system equations that bring Equation (31) closer to the form of an RPA polynomial as follows:

First, for both the original and modified CRNs, the rate equation f_8 indicates that x_8 is directly proportional to the monomial $x_1 x_2$ at steady state. The second grouping of terms in Equation (31) can therefore be absorbed into a single term via the identity

$$\frac{d_3}{d_3 + k_3} f_8 = \frac{a_3 d_3}{d_3 + k_3} x_1 x_2 - d_3 x_8.$$

Likewise, the rate equation f_9 indicates that x_9 is directly proportional to the monomial $x_2 x_3$ at steady state, enabling the third grouping of terms in Equation (31) to be absorbed into a single term via the identity

$$\frac{d_4}{d_4 + k_4} f_8 = \frac{a_4 d_4}{d_4 + k_4} x_2 x_3 - d_4 x_9.$$

In addition, it is clear that x_1 is proportional to x_7 at steady state, since

$$f_5 + f_7 = k_1 x_1 - k_2 x_7.$$

Therefore, the rowspan of the system contains the polynomial

$$\hat{p} = f_2 + \frac{d_3}{d_3 + k_3} f_8 + \frac{d_4}{d_4 + k_4} f_9 + f_5 + f_7 = k_1 x_1 - \frac{a_3 k_3}{d_3 + k_3} x_1 x_2 - \frac{a_4 k_4}{d_4 + k_4} x_2 x_3. \tag{32}$$

The crucial feature of Equation (32) is that if x_1 is proportional to x_3 at steady-state, then $\hat{p}$ will assume the form of an RPA polynomial.

Now, for the modified system, we observe that x_1 and x_3 are indeed proportional at steady-

state, provided $x_2 \neq 0$, since

$$f_3 + f_9 = d_5 x_1 x_2 - a_5 x_2 x_3 \tag{33}$$

- a balancer condition [1]. Moreover, Equation (33) can be directly incorporated (by linear combination) into Equation (32) to yield the RPA polynomial

$$\begin{aligned} a_5\hat{p} - \frac{a_4 k_4}{d_4 + k_4}\left(f_3 + f_9\right) &= a_5 k_1 x_1 - \frac{a_5 a_3 k_3}{d_3 + k_3} x_1 x_2 - \frac{a_4 k_4 d_5}{d_4 + k_4} x_1 x_2, \\ &= a_5 k_1 x_1 - \gamma x_1 x_2, \\ &= -\gamma x_1 \left(x_2 - \frac{a_5 k_1}{\gamma}\right), \end{aligned}$$

where $\gamma = \left(\frac{a_5 a_3 k_3}{d_3 + k_3} + \frac{d_5 a_4 k_4}{d_4 + k_4}\right)$.

For the original system, on the other hand, we note that x_1 and x_3 are proportional at steady state from the identities

$$f_3 + f_9 = d_5 x_4 - a_5 x_3$$

and

$$f_1 + f_8 = a_1 x_4 - \left(d_1 + k_1\right) x_1,$$

which gives

$$a_1(f_3 + f_9) - d_5(f_1 + f_8) = d_5(d_1 + k_1)x_1 - (a_1 a_5)x_3. \tag{34}$$

Thus, we now observe a critical difference between the original and modified systems from the point of view of their RPA capacity: both systems are RPA-capable (indeed ACR capable) since they both contain the polynomial $\hat{p}$ in their rowspan, and each system contains a 'balancer polynomial' in its rowspan that constrains x_1 to be proportional to x_3 at steady state, thereby allowing $\hat{p}$ to assume the form of an RPA polynomial. But the two systems differ in how this additional balancer condition is realized *stoichiometrically*. And, as a consequence, the two systems differ in how the two conditions (the connector polynomial, $\hat{p}$, and the respective

balancer polynomial) can be combined *algebraically* to produce the RPA polynomial.

Importantly, in both Example 1 and Example 2 (Sections S3.1 and S3.2), the balancer polynomials and connector polynomials required to construct the respective RPA polynomials were in the *rowspan* of the CRN's mass action equations. In the next section, we will see that this is *always* true of RPA-capable CRNs of balancer type: the connector polynomial, and any requisite balancer polynomials, are *always* in the rowspan of the CRN's mass action equations. A similar result obtains for CRNs of opposer type, whose mass-action equations *always* contain opposer polynomial(s) in the rowspan.

S4 RPA Invariants and a Universal Implementation of Integral Control for all RPA-Capable CRNs

In this section, we explore the significance of a CRN's *deficiency* (see main article) in the context of RPA capacity, and show that for *all* RPA-capable CRNs, the rowspan of the system contains a connector polynomial along with all required balancer polynomials (if the CRN is topologically of Balancer-type), or one or more opposer polynomials (if the CRN is topologically of Opposer-type).

The standard control theory viewpoint is to identify a (generally nonlinear) coordinate change that enables the system to be partitioned into two components [12] - one being the 'internal model', and the other being the remainder of the system. In the viewpoint we propose, by contrast, we partition the system into (generally) *more than two* components. In particular, the internal model is distributed over a number of elements (invariants) in our formulation, each of these invariants being obtained by a linear coordinate change (i.e. via a linear combination of the CRN's rate equations). These individual invariants may then be combined nonlinearly through 'monomial concatenation' - that is, through multiplication by a suitable monomial in order to reconcile each such invariant with the 'next' invariant in the CRN's topological struc-

ture - to obtain the all-important RPA polynomial (see previous Section).

We note that, although there should always exist some single nonlinear coordinate change that yields the RPA polynomial *in principle*, for any mass action system, identifying such a coordinate change *in practice* may be extremely difficult in all but the very simplest RPA-capable CRNs. (See, for instance [11] for an example of identifying a single nonlinear coordinate change for a very simple RPA-capable motif). Of greater concern is the fact that, even where a suitable nonlinear coordinate change can be identified, this transformation, in and of itself, cannot reveal a universal picture of the essential structural principles by which RPA is transacted through CRNs *in general*.

By contrast, the integral control formulation we present here provides a definitive picture of *all possible* RPA-capable CRNs. Moreover, the fact that the transformation required to obtain an RPA polynomial is *almost linear*, in the sense that a limited number of special polynomial invariants - each of which is obtained through *linear* transformation - are combined to yield the RPA polynomial, means that the existence of an RPA polynomial can be quickly checked through direct computation of a suitable Gröbner basis with an elimination (eg. lexicographic) monomial ordering. Importantly, this simple and direct test for RPA capacity also yields the setpoint of the system, along with a candidate set of elimination polynomials $(h_1, \ldots, h_n)$ which make it clear how many (nonlinear) concatenation steps are required for the CRN in question.

We now provide a mathematical justification for our claim that any RPA-capable CRN has connector and balancer polynomials (for Balancer-type CRNs) or opposer polynomials (for Opposer-type CRNs) in the *rowspan* of its mass-action equations. We first briefly review a number of key known theorems, as well as several new theorems, relevant to our claim that a linear change of coordinates is always sufficient to identify the key subsidiary invariants of any RPA-capable CRN.

S4.1 Known Results

Proposition 1. For any mass-action CRN, with reaction rates given by $f(\mathbf{x}) = Y.\mathcal{L}(G).\psi(\mathbf{x})$, the following holds

$$\dim(\ker \mathrm{Y}\mathcal{L}(\mathrm{G})) - \dim(\ker \mathcal{L}(\mathrm{G})) \leq \delta.$$

Proof. See Section 6 of [5]. □

Corollary 1. Consider a mass-action CRN with either δ = 0, or $\delta = 1$. Then,

$$\dim(\ker \mathrm{Y}\mathcal{L}(\mathrm{G})) - \dim(\ker \mathcal{L}(\mathrm{G})) = \delta.$$

Proof. See Proof for Lemma S3.20 in Supplementary Information of [10]. □

Remark. In other words, it follows that $\dim(\ker \mathrm{Y}\mathcal{L}(\mathrm{G})) = \dim(\ker \mathcal{L}(\mathrm{G}))$ for deficiency-zero networks, and that $\dim(\ker \mathrm{Y}\mathcal{L}(\mathrm{G})) = \dim(\ker \mathcal{L}(\mathrm{G})) + 1$ for deficiency-one networks.

The next theorem, originally proved in [5], appeals to the standard terminology that the support of a vector $\mathbf{z}$, $\mathrm{supp}(\mathbf{z})$, is the collection of non-zero elements of $\mathbf{z}$.

Proposition 2. Consider a CRN with t terminal SCCs,$\{\Lambda^1, \ldots, \Lambda^t\}$. The kernel of $\mathcal{L}(G)$ contains a basis $\{b^1, \ldots, b^t\}$ with the property that supp(b^i) = Λ^i, for $i \in \{1, \ldots, t\}$.

Proof. See Appendix of [5]. □

Remark. The dimension of the kernel of $\mathcal{L}(G)$ is numerically equal to the number of terminal SCCs of the CRN. For each terminal SCC, Λ^i, the associated basis vector, b^i, has non-zero entries corresponding to the complexes contained in Λ^i, and zero entries elsewhere. In other words, each basis vector has its support on the complexes of a single terminal SCC of the CRN graph.

Lemma 1. Consider an n-species mass-action CRN of deficiency one with at least one non-terminal complex, and a steady state $\mathbf{c} \in \mathbb{R}^n_{>0}$ in the positive orthant. Then,

$$\mathrm{ker}\mathrm{Y}\mathcal{L}(\mathrm{G}) = \mathrm{ker}\mathcal{L}(\mathrm{G}) \oplus \eta,$$

where $\eta = \{\alpha\psi(\mathbf{c}) : \alpha \in \mathbb{R}\}$.

Proof. This follows straightforwardly from Corollary 1 and Proposition 2 above. In particular, since $\mathbf{c}$ is a steady-state of the system, the set η forms a one-dimensional vector space contained in $\mathrm{ker}\mathrm{Y}\mathcal{L}(\mathrm{G})$ that is linearly independent of $\mathrm{ker}\mathcal{L}(\mathrm{G})$, since all vectors in $\mathrm{ker}\mathcal{L}(\mathrm{G})$ have support in the terminal complexes (only), whereas $\psi(\mathbf{c})$ has support on all complexes (which include at least one non-terminal one, by supposition), since $\mathbf{c}$ is in the positive orthant. □

Theorem 2 (Shinar-Feinberg)**.** Consider a mass-action system of deficiency one that admits a positive steady state. If, in the network, there are two non-terminal complexes that differ only in species S, then the system has absolute concentration robustness (ACR) in S.

Proof. See Supplementary Information for [10] □

S4.2 Polynomials in the Rowspan of a Mass-Action CRN

We now provide an alternative version of Shinar-Feinberg's theorem, with a new proof, to give our first key result on an important class of polynomials (more specifically, *binomials*) that always reside in the rowspan of certain mass-action CRNs.

Theorem 3. Consider a deficiency-one mass-action CRN with state vector $\mathbf{x} \in \mathbb{R}^n$, that admits a steady-state, $\mathbf{c} \in \mathbb{R}^n$, in the positive orthant. For any two non-terminal complexes, with corresponding monomials $\psi_i(\mathbf{x})$ and $\psi_j(\mathbf{x})$, there exists some $\mathbf{a} \in \mathbb{R}^n$ such that

$$\mathbf{a}^T \frac{d\mathbf{x}}{dt} = \alpha_1\psi_i(\mathbf{x}) - \alpha_2\psi_j(\mathbf{x}),$$

for some pair of rational functions (of system parameters), α_1 and α_2.

Proof. We consider the vector of monomials $\psi(\mathbf{x})$, whose component $\psi_k(\mathbf{x})$ corresponds to complex k ($k \in \{1, \ldots, m\}$). Now, consider any two distinct non-terminal complexes of the CRN, say y_i and y_j, with corresponding monomials $\psi_i(\mathbf{x})$ and $\psi_j(\mathbf{x})$.

Consider a vector $\rho \in \mathbb{R}^m$ with exactly two non-zero entries, namely $\rho_i = \psi_j(\mathbf{c})$ and $\rho_j = \psi_i(\mathbf{c})$. Recall that for deficiency-one CRNs, $\mathrm{ker}\mathrm{Y}\mathcal{L}(\mathrm{G}) = \mathrm{ker}\mathcal{L}(\mathrm{G}) \oplus \eta$ (where η is as defined in Lemma 1). It is clear that ρ is orthogonal to all vectors in η, i.e $\rho \in \eta^{\perp}$. Moreover, since any vector $b \in \mathrm{ker}\mathcal{L}(\mathrm{G})$ can be expressed in the form $a_1 b_1 + \ldots + a_t b_t$, where the support of each $b_i \in \{b_1, \ldots, b_t\}$ is on terminal complexes (see Proposition 2), it is clear that $\rho \in (\mathrm{ker}\mathcal{L}(\mathrm{G}))^{\perp}$. Thus, since $\rho \in \eta^{\perp}$ and $\rho \in (\mathrm{ker}\mathcal{L}(\mathrm{G}))^{\perp}$, it follows from Lemma 1 that $\rho \in (\mathrm{ker}\mathrm{Y}\mathcal{L}(\mathrm{G}))^{\perp}$. Moreover, since ρ lies in the orthogonal complement to the kernel (nullspace) of $Y\mathcal{L}(G)$, it follows that ρ^T is in the rowspan of $Y\mathcal{L}(G)$. Thus, there exists some $\mathbf{a} \in \mathbb{R}^n$ such that $\mathbf{a}^T \frac{d\mathbf{x}}{dt} = \alpha_1 \psi_i(\mathbf{x}) - \alpha_2 \psi_j(\mathbf{x})$, with $\alpha_1 = \rho_i = \psi_j(\mathbf{c})$ and $\alpha_2 = \rho_j = \psi_i(\mathbf{c})$. Moreover, since the matrix entries of $Y\mathcal{L}(G)$ contain linear combinations of the system rate constants, it follows that $\alpha_1 = \rho_i$ and $\alpha_2 = \rho_j$ are rational functions of system parameters. □

Corollary 2. For a deficiency-one mass-action CRN that admits a positive steady-state, whose RPA polynomial is its connector polynomial (if the CRN is topologically of balancer type) or its (single) opposer polynomial (if the CRN is topologically of opposer type), then the RPA polynomial is guaranteed to be the rowspan of the system, and its RPA capacity may be determined by the Shinar-Feinberg theorem [10].

Because of Corollary 1 and Lemma 1, the contents of the rowspan of a CRN's mass-action equations, and the RPA-capacity of the CRN, may easily be determined for deficiency-one CRNs. But once $\delta > 1$, no such simple conclusions may be drawn in complete generality since the dimension of the kernel of $Y\mathcal{L}(G)$ now exceeds the dimension of the kernel of $\mathcal{L}(G)$ by *at*

least δ (as opposed to *exactly* δ, as was the case when $\delta = 1$ or 0). Nevertheless, it is clear that RPA is possible for CRNs with $\delta > 1$, and there are many known examples, including Examples 1 and 2 examined in the preceding section, and Figure 1 in the main paper.

To proceed further to the general principles of RPA-capable CRNs of arbitrary deficiency (δ), we recognise that there are two distinct ways for a network to increase its deficiency while maintaining RPA:

1. Through additional reactions that lie (topologically) outside of the network structures through which RPA is implemented. These additional deficiency-increasing reactions do not contribute to RPA, but are nevertheless compatible with RPA. In this case, the additional deficiency-increasing reactions may be discarded from the analysis, and RPA capacity may be determined from the remaining ($\delta = 1$) collection of reactions;

2. Through the embedding of deficiency-increasing reactions into the RPA-conferring reaction structures of the CRN; in contrast to the previous case *ut supra*, the added deficiency-increasing reactions cannot simply be discarded since they are now part of the RPA mechanism itself.

We now consider each of these possibilities in turn, through the analysis of some illustrative examples.

S4.2.1 Deficiency-increasing reactions that do not contribute to RPA

Consider, for example, the well-established antithetical integral control mechanism (Figure S6), which is topologically of Opposer type:

This reaction mechanism comprises five complexes ($m = 5$), two linkage classes ($l = 2$) and has a rank of two ($s = 2$), so that the deficiency of this portion of the CRN - excluding any additional chemical reactions that may be added into the overarching feedback loop as

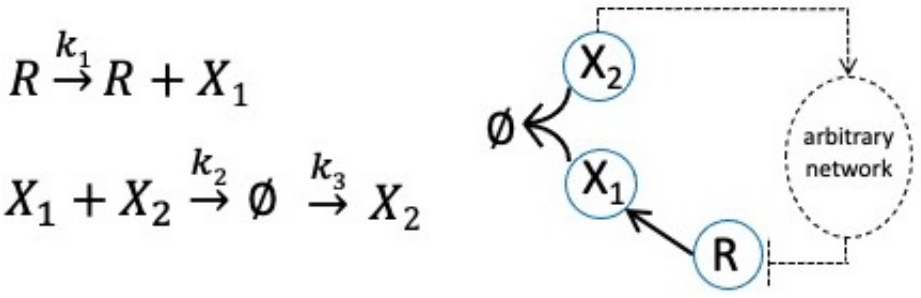

Figure S6: Antithetic integral control, as proposed by Briat et al. [2].

indicated in Figure S6 - is one. From this it is clear from our Theorem 3 that, considering the two non-terminal complexes R and ϕ (with corresponding monomials R and 1, respectively), the rowspan of the mass-action system will contain a polynomial of the form $\alpha_1 R - \alpha_2$ - which is an opposer polynomial as well as the RPA polynomial of the system, which indicates that the CRN has the capacity for RPA at R. Indeed, the induced mass-action equations readily confirm this property, since

$$\frac{dX_1}{dt} = k_1 R - k_2 X_1 X_2, \tag{35}$$

$$\frac{dX_2}{dt} = k_3 - k_2 X_1 X_2, \tag{36}$$

and thus, it is clear by inspection that $\frac{dX_1}{dt} - \frac{dX_2}{dt} = k_1 R - k_3$. (So, $\alpha_1 = k_1$ and $\alpha_2 = k_3$ in this case.)

But note that *any* collection of chemical reactions could be embedded into the network at the indicated position in Figure S6 - ie. outside of the opposer mechanism. These reactions could (and generally will) increase the deficiency of the network, possibly even making the deficiency very large ($\delta \gg 1$). In addition, these (potentially) deficiency-increasing chemical reactions do not change the setpoint ($R^* = \frac{k_3}{k_1}$), which is computed (as given above) by the deficiency-one opposer sub-network.

Another clear example of deficiency-increasing chemical reactions that lie outside of the RPA-conferring subnetwork of a CRN was considered in Example 1 of the previous section (see Figure S7 below).

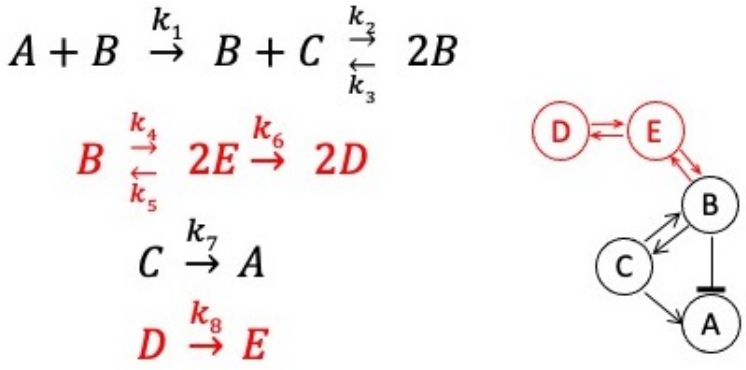

Figure S7: Chemical reactions and topological structure for the toy model proposed by Cappelletti et al. [3], as examined in detail in Section S3.1 of the present Supplement.

This CRN comprises ten complexes ($m = 10$), four linkage classes ($l = 4$), and has a rank of four ($s = 4$), giving rise to a deficiency of two ($\delta = 2$). As explained in Section S3.1, B both downregulates A (through direct binding and subsequent catalytic conversion to C) *and* upregulates A (indirectly, via C) - a topological arrangement depicted schematically in Figure S7(right). For this reason, B acts as a diverter species, and A acts as a connector species, with C as the intermediate balancer. A thus has the potential for RPA due to this topological structure. In addition, B regulates E, which in turn regulates D. Note that these additional reactions emanating from B do not affect the diverter role of B; these reactions are not required to form the balancer module, but are nevertheless compatible with a balancer topology (see [1] for an exhaustive explanation of viable balancer and opposer module topologies). Thus, from the viewpoint of assessing the RPA capacity of this CRN, these additional reactions, indicated in red in Figure S7, may be removed. This leaves five remaining complexes ($m = 5$), two linkage classes ($l = 2$), and a rank of two ($s = 2$), producing a deficiency of one ($\delta = 1$). From this it is

clear that the CRN has the capacity for RPA at A, since polynomials of the form $\alpha_1 AB - \alpha_2 C$ and $\alpha_3 BC - \alpha_4 B^2$ reside in the rowspan of the CRN's mass-action equations as a consequence of Theorem 3. Indeed, as demonstrated in Section S3.1, these polynomials are readily shown to be $-k_1 AB + k_7 C$ and $-k_2 BC + k_3 B^2$, respectively, which can be combined (via a concatenating monomial) and re-arranged to produce the RPA polynomial $k_1 k_2 B^2 \left(A - \frac{k_7 k_3}{k_1 k_2}\right)$. In addition, the setpoint $A^* = \frac{k_7 k_3}{k_1 k_2}$ is unaffected by the presence or absence of the 'unnecessary' (red) deficiency-increasing reactions.

In both of these examples (the antithetic integral control motif [2], and the toy model proposed by Cappelletti [3]), RPA is implemented by a deficiency-one subnetwork of the system. Any additional deficiency-increasing reactions have no bearing on the RPA capacity of the CRN, or the value of its setpoint, due to the topological relationship of the additional reactions to the deficiency-one RPA-implementing reaction structures. This principle is also captured algebraically by the concept of *independent subnetworks* of a CRN (see Appendix 6.A of [4] for a detailed description); with this algebraic viewpoint in mind, we note that the *span* of the reactions noted in red in Figure S7 ($s = 2$) may be added to the *span* of the remaining reactions noted in black ($s = 2$) to give the *span* of the full CRN ($s = 4$). From elementary linear algebra, it follows that any steady-state of the full CRN must also be a steady-state of any independent subnetwork (see [4], Appendix 6.A.)

S4.2.2 Deficiency-increasing reactions that contribute to RPA

A second possibility is for any deficiency exceeding unity to be embedded into the RPA-conferring reaction structures themselves, thereby contributing to the RPA mechanism. In contrast to the cases considered in the preceding section (S4.2.1), the deficiency-increasing reactions can no longer simply be discarded from analysis. Topologically, there are two distinct ways this scenario can arise - one pertaining to Balancer modules, and one pertaining to

Opposer modules.

1. **Balancer modules**. If a CRN is topologically a Balancer module, it comprises at least two parallel pathways - at least one of which has an inhibitory (downregulating) effect on the connector variable, and at least one of which has a promoting (upregulating) effect. A balancer-type CRN with exactly two such parallel counteracting pathways (ie. the minimum), and without any additional deficiency-increasing reactions outside the parallel pathways (as was the case for the Cappelletti toy model in the previous section, for instance), will have a deficiency of *one*[2]. But an RPA-capable CRN with more than two parallel pathways in its balancer structure will increase its deficiency by one for each extra parallel pathway.

   We demonstrate these principles through a detailed analysis of the deficiency-two EnvZ-OmpR motif studied by Shinar and Feinberg (see Supplementary Information to [10]), which we considered in Example 2 in Section S3.2. We reproduce the four linkage-class model of this deficiency-two CRN of the EnvZ-OmpR osmoregulation system in Figure S8 below.

   It is clear that the first two linkage classes, taken together, have defiency zero. But adding in *either* the third linkage class *or* the fourth linkage class increases the deficiency to one. The deficiency-one subnetwork formed by the first three linkage classes, for example, satisfies the Shinar-Feinberg theorem since there are now two non-terminal complexes - XT and $XT + Y_p$ - that differ in a single species (Y_p). Thus, these three linkage classes, taken together as a subnetwork, have the capacity for RPA (ACR) at Y_p. Likewise, the deficiency-one subnetwork formed from linkage classes 1, 2 and 4 also satisfies the Shinar-Feinberg theorem since there are two non-terminal complexes - this time, XD

[2]The only exception to this is a 'trivial' Balancer module, which is simply a single connector node (no diverter or balancers), which therefore has deficiency-zero, Eg. $B+C \rightleftharpoons B$. It is easy to show that no such deficiency-zero RPA-capable CRN can be mass-conservative.

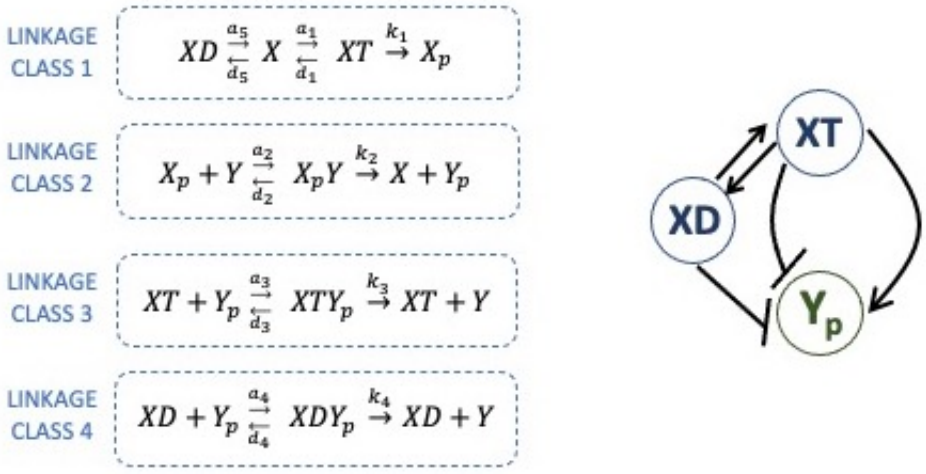

Figure S8: Deficiency-two CRN of the EnvZ-OmpR osmoregulation system first studied by Shinar and Feinberg (see Supplementary Information to [10]).

and $XD + Y_p$ - that differ in the species Y_p, thereby conferring RPA (ACR) on Y_p for this three linkage-class subnetwork.

It is clear from the topological structure of this CRN how additional linear dependences of reactions (and hence higher deficiencies) arise as we successively add linkage classes to form larger subnetworks. Indeed, the first two linkage classes, taken together, contain the reactions (as differences between complexes, see [10]) $XT - X$, $X_p - XT$, $X_pY - Y - X_p$, and $X + Y_p - X_pY$. Adding these together gives $Y_p - Y$ in the span of the reactions, highlighting the fact that Y is converted to Y_p via these first two linkage classes. Furthermore, the third linkage class contains the reactions $XTY_p - XT - Y_p$ and $XT + Y - XTY_p$. Adding these together gives $Y - Y_p$, highlighting the fact that Y_p is converted back to Y by the reactions in this third linkage class. Likewise, the fourth linkage class contains the reactions $XDY_p - XD - Y_p$ and $XD + Y - XDY_p$, the sum of which produces $Y - Y_p$. Thus, the fourth linkage class has the same basic function as the third: to convert Y_p to Y.

To form a Balancer module, then, we require the first two linkage classes to be paired with *at least one* of linkage classes three or four. These latter two linkage classes provide two parallel pathways that counteract the pathway constructed by the first two linkage classes taken together. Indeed, for the deficiency-one subnetwork formed by the first three linkage classes, Theorem 3 guarantees that the following polynomials exist in the rowspan of the mass-action system corresponding to these three linkage classes in isolation:

$$\alpha_1 XT - \alpha_2 XT.Y_p, \tag{37}$$

$$\alpha_3 XT - \alpha_4 XD. \tag{38}$$

By contrast, for the deficiency-one subnetwork constructed from linkage classes 1, 2 and 4, Theorem 3 guarantees that the rowspan of the mass-action system corresponding to these three linkage classes will include the following polynomials:

$$\alpha_5 XT - \alpha_6 XD.Y_p, \tag{39}$$

$$\alpha_3 XT - \alpha_4 XD.$$

Thus, if all four linkage classes are now taken together, it follows that

$$\alpha_7 XT - \alpha_8 XT.Y_p - \alpha_9 XD.Y_p, \tag{40}$$

from polynomials (37) and (39) above, and

$$\alpha_3 XT - \alpha_4 XD,$$

will both reside in the rowspan of the full system. Polynomial (40) is the connector polynomial for the full system, whereas polynomial (38) is a balancer polynomial. As shown in Section S3.2, these two key polynomial invariants may be combined, via a

concatenating monomial, to produce an RPA polynomial.

The connector polynomial (40) is a special case of *connector kinetics*, first identified in [1]. The general form of a connector polynomial is thus

$$m_1 X_c^{\beta} \pm \ldots \pm m_t X_c^{\beta} - m_{t+1} X_c^{\alpha} \ldots \pm m_T X_c^{\alpha}, \tag{41}$$

where $\beta \in \mathbb{Z}_{\geq 0}$, $\alpha \in \mathbb{Z}_{>0}$, and $\beta < \alpha$, and where $m_1 \ldots m_T$ are monomials, each of which involving one of the T *terminal* balancer molecules (corresponding to terminal balancer nodes, as given in [1]). In addition, X_c is the (potentially) RPA-capable variable of the CRN. Note that a diverter molecule may also be a terminal balancer molecule.

To achieve RPA, each terminal balancer monomial must be in direct proportion to the diverter variable at steady-state, and each such condition requires a corresponding balancer polynomial (e.g. polynomial (38)) to exist in the ideal associated to the CRN. By an induction argument, starting with *any two* sufficient parallel pathways of the balancer module and adding in collections of chemical reactions corresponding to any additional parallel pathways, one pathway at a time, it is clear from Theorem 3 and the argument presented in the analysis of Figure S8 above, that the connector polynomial (41) (e.g. polynomial (40)), and all requisite balancer polynomials (e.g. polynomial (38)), *always* reside in the rowspan of the CRN's mass-action equations.

2. **Opposer modules.** As we have noted previously, the simple antithetic integral control motif [2] (see Figure S6), together with the network it controls (embedded into a feedback loop), is an example of an Opposer module [1]. As explained in Section S4.2.1, only the *controller reactions*, not the full network, need be considered in determining the RPA capacity and setpoint of the full network. The antithetic integral controller itself is an example of a *deficiency-one* CRN of opposer type, whose mass action equations contain

an RPA polynomial in its rowspan that encodes *opposer kinetics* [1].

Neglecting the 'unnecessary' (and potentially deficiency increasing) reactions in the embedded network (i.e. outside of the controller reactions), the controller CRN of an Opposer module could still increase its deficiency ($\delta > 1$) while still preserving the RPA-capacity of the controller (and ultimately the entire embedded network), if it comprises *multiple independent opposer invariants*, thus possessing an overall topological structure known as an *opposing set* [1].

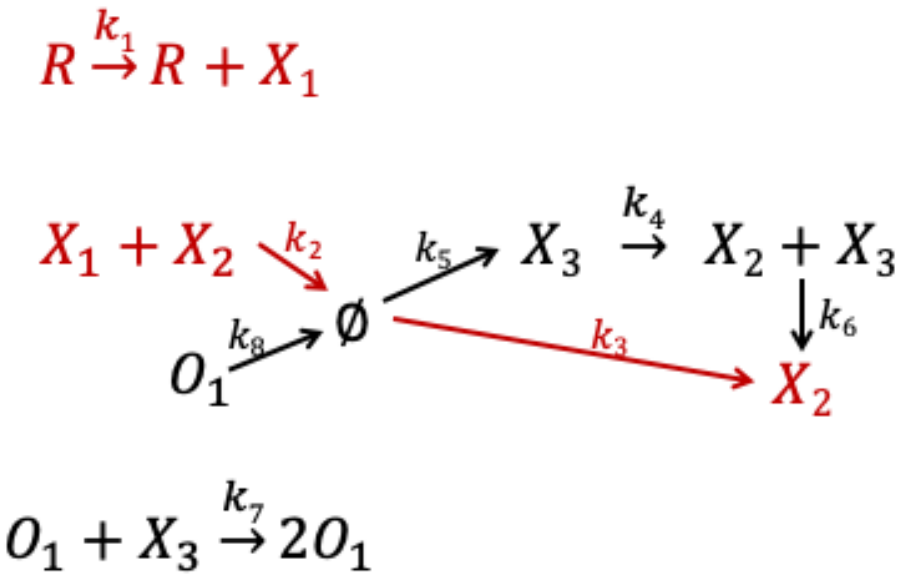

Figure S9: CRN presented in Figure 1b in the main paper, organised into linkage classes, and employing the antithetic integral control motif [2] as one of two opposers in a two-node opposing set.

We illustrate the essential principles of an opposing set at the level of intermolecular interactions through the analysis of the example given in Figure 1b in the main paper. We present this network organised into its correct linkage classes in Figure S9. This CRN has a deficiency of three, since it comprises ten complexes ($m = 10$), three linkage classes ($l = 3$), and has a span of four ($s = 4$).

In Figure S10, we decompose this CRN into two independent subnetworks, each with a span of two ($s = 2$). Each of these subnetworks has a deficiency of two. Since the two subnetworks are independent, they can be analysed separately with respect to the polynomial invariants in their respective rowspans.

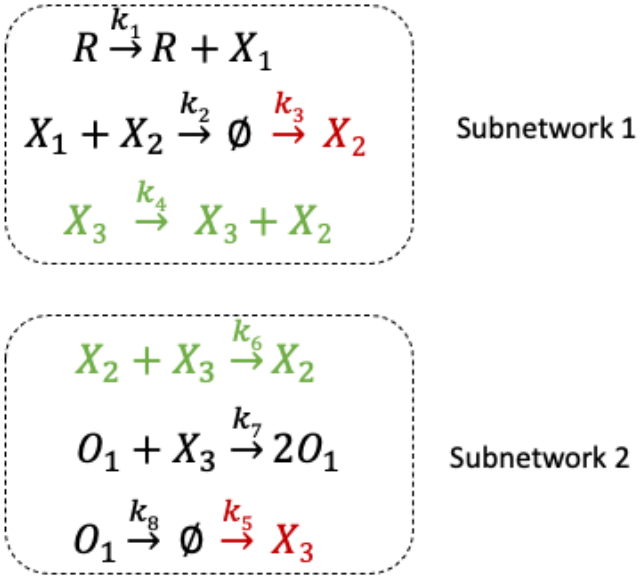

Figure S10: CRN presented in Figure S9, decomposed into two independent subnetworks, each with a span, $s = 2$.

Consider Subnetwork 1. If *either* the reaction noted in red *or* the reaction noted in green is omitted, the deficiency of the remaining subnetwork reduces to one. In particular, neglecting the green reaction, the subnetwork reduces to the antithetical control motif, and contains an RPA polynomial of the form $\alpha_1 R - \alpha_2$ in its rowspan as a consequence of Theorem 3. If, instead, the red reaction is neglected, the remaining deficiency-one network contains a polynomial of the form $\alpha_3 R - \alpha_4 X_3$ in its rowspan. It follows that if *both* the red *and* the green reactions are included, a polynomial of the form $\alpha_5 R - \alpha_6 - \alpha_7 X_3$ is present in the rowspan of the mass-action equations associated to Subnetwork 1. This is no longer an RPA polynomial. Indeed, as we note in the main paper, the inclusion

of the green reaction in this subnetwork prevents this CRN from achieving RPA by itself, and requires an additional controller structure (in this example, Subnetwork 2) in order to preserve the RPA-promoting properties of the antithetic integral control motif.

Consider Subnetwork 2. Here again, if *either* the reaction noted in red *or* the reaction noted in green is omitted, the deficiency of the remaining subnetwork reduces to one. In either case, the deficiency-one subnetwork contains a polynomial of the form $\alpha_8 O_1 X_3 - \alpha_9 O_1$ - an RPA polynomial for X_3 - as a consequence of Theorem 3. Therefore, the full Subnetwork 2 also contains a polynomial $\alpha_8 O_1 X_3 - \alpha_9 O_1$ in its rowspan, and confers the capacity for RPA on X_3. This being the case, this invariant can now be 'passed' to the invariant $\alpha_5 R - \alpha_6 - \alpha_7 X_3$ for Subnetwork 1, which can now be interpreted as an opposer invariant. As indicated in Figure1c of the main paper, combining the opposer invariants from each of the two independent subnetworks requires the *concatenating monomial* O_1 to be applied to the opposer polynomial from Subnetwork 1 (since this contains the monomial X_3) in order to combine with the opposer polynomial from Subnetwork 2 (since this contains the monomial $O_1 X_3$), in order to produce the RPA polynomial given in Figure 1c of the main paper.

The CRN presented in Figure 1c of the main paper, decomposed into two independent subnetworks in Figure S10, is thus a *two-node opposing set*, since it contains two independent *opposer invariants* (or *opposer polynomials*), each contained in the rowspan of the CRN's mass action equations:

$$\alpha_5 R - \alpha_6 - \alpha_7 X_3, \tag{42}$$

$$\alpha_8 O_1 X_3 - \alpha_9 O_1, \tag{43}$$

where it is clear that, for this example, $\alpha_5 = k_1$, $\alpha_6 = k_3$, $\alpha_7 = k_4$, $\alpha_8 = k_7$ and $\alpha_9 = k_8$. The topological structures of *opposing sets* are now well understood, and are fully described in [1]. From this point of view, we consider polynomial (42) to be the *proximal* opposer invariant, since it is directly regulated by the embedded network (which exhibits RPA, and which is not explicitly considered in the analysis of the RPA capacity of the Opposer module as a whole). From a control theory viewpoint, the molecule R could be considered the *sensor molecule*. The polynomial (43), on the other hand, is the most *distal* opposer invariant, being the polynomial furthest away (topologically speaking) from the sensor molecule, and whose non-RPA exhibiting molecule (here O_1) - which must regulate the embedded network in order for the system as a whole to have the capacity for RPA - could be considered an *actuator molecule* from a control theory viewpoint. We summarise the essential features of opposing sets in Figure S11.

The principles considered above for a particular two-node opposing set can thus readily be generalised to any controller CRN of an Opposer module, with arbitrary numbers of independent opposer polynomials, and thus arbitrarily high deficiency, as predicted in the general theory of RPA at the network macroscale given in [1]. Owing to the topological principles governing opposing sets, any controller CRN for an Opposer module can be decomposed into n independent subnetworks ($n \geq 1$), with each such subnetwork contributing one opposer polynomial ρ_i which exists in the *rowspan* of the mass action equations of the respective subnetwork. The distal opposer polynomial, will take the form

$$\rho_n = g(x_{jn})(x_{in} - c_n), \tag{44}$$

where x_{jn} is an 'actuator molecule' (i.e. non RPA-capable) for the distal subnetwork, x_{in} is an RPA-capable molecule in the distal subnetwork, and c_n is the setpoint of x_{in}

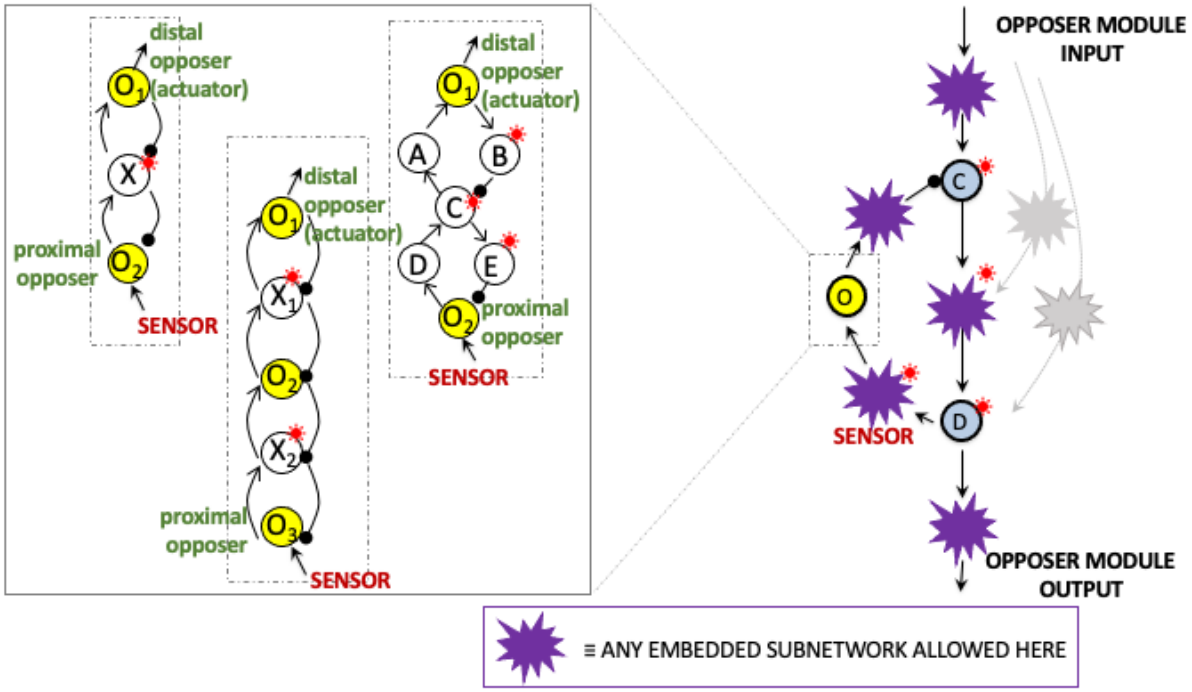

Figure S11: The essential topological features of opposing sets. The schematics above present two examples of two-node opposing sets (left and right), and one example of a three-node opposing set (center). RPA-capable nodes are indicated with a red asterisk. Full details given in [1].

(a rational function of rate constants in the distal subnetwork). For the remaining $n-1$ independent subnetworks, the opposer polynomial will take the form

$$\rho_k = g(x_{jk})(x_{ik} - c_k - \hat{c}_k p_{k+1}), \tag{45}$$

where p_{k+1} is a polynomial involving RPA-capable molecule(s) from the $(k+1)$-th (i.e. next most distal) subnetwork that also participate in subnetwork k, and where c_k and $\hat{c}_k$ are rational functions of rate constants in subnetwork k. The setpoint for the RPA-capable molecule x_{ik} is thus $c_k + \hat{c}_k c_{k+1}$.

S4.3 Summary of general principles for RPA in CRNs

The following represents a summary of the general principles governing the RPA capacity of all CRNs, as outlined in the preceding sections:

1. For each RPA-promoting module, the key subsidiary polynomial invariants reside in the *rowspan* of the CRN. In particular,

   (a) For a Balancer module, the connector polynomial, along with all requisite balancer polynomials, reside in the rowspan of the CRN.

   (b) For an Opposer module, comprising an n-node opposing set ($n \geq 1$), each of the n opposer polynomials resides in the rowspan of the CRN.

2. Large and complex RPA-capable CRNs could be composed of multiple such modules, connected together. The full set of topological possibilities for how this may occur is now understood in complete generality (see [1]). It is clear from the analysis in the preceding sections that a CRN can be decomposed into independent subnetworks corresponding to the individual modules of a multi-modular network, which can be analysed separately.

3. For Balancer modules, invariants are 'passed' downstream from the diverter to each terminal balancer, culminating in the construction of an RPA polynomial from the connector polynomial. By contrast, for Opposer modules, invariants are passed from the distal opposer to the proximal opposer.

4. If the subsidiary polynomials (balancers, connectors, opposers) are *stoichiometrically independent* (see Section S3), they require a concatenating monomial to combine invariants to ultimately produce the RPA polynomial. Otherwise, they can combine without leaving the rowspan of the CRN's mass-action equations. Thus, in special cases, the RPA poly-

nomial itself may reside in the rowspan. (CRNs satisfying the Shinar-Feinberg theorem are included among these special cases.)

5. In all cases, the fact that *all* the key invariants of all RPA-capable CRNs exist in the rowspan of the mass-action equation, reveals a well-defined connection to integral control, as outlined in the main paper.

The analysis in this Section (S4) makes clear that sets of polynomials $\{h_1, \ldots, h_n\}$ that are able to eliminate all but two variables from the rate equations to 'reach' the RPA polynomial of the CRN (Theorem 1) possess a very special structure. In particular,

$$[h_1, \ldots, h_n]^T \in Z \oplus \hat{h}, \tag{46}$$

where Z is the module of syzygies of the CRN rate equations, and $\hat{h}$ is the 'canonical' basis set of polynomials $h_1, \ldots, h_n$ that 'reaches' the RPA polynomial of the CRN.

Some CRNs will contain mass-conservation equations. For these, the module of syzygies may be further decomposed into the form

$$Z \oplus \hat{h} = Z_x \oplus Z_\phi \oplus \hat{h}, \tag{47}$$

where Z_ϕ is the co-kernel (left nullspace) of $Y.\mathcal{L}(G)$. CRNs without mass-conservation relations will have a trivial co-kernel, with $Z = Z_x$. In any case, it follows from the analysis earlier in this Section that the only monomials contained in any polynomial in $Z_\phi \oplus \hat{h}$, for judicious choice of two projection variables (see Section S2), are the requisite concatenating monomials.

Given this special 'almost linear' structure of the ideal associated to an RPA polynomial, RPA capacity may be tested systematically via Buchberger's algorithm, which computes the Gröbner basis of an ideal, with an elimination ordering on monomials. In addition, using this algorithm, a set of elimination polynomials $\{h_1, \ldots, h_n\} \in Z_\phi \oplus \hat{h}$ can be computed auto-

matically. In the next Section, we provide code to make these systematic computations in the open-source software *Singular* (www.singular.uni-kl.de), which can be adapted to the study of any CRN.

S5 Singular code for analysing the RPA capacity of CRNs

S5.1 General structure of Singular code

The general structure of a *Singular* script that systematically tests the RPA capacity of a CRN and automatically determines the subsidiary polynomial invariants along with a set of elimination polynomials, $\hat{h}$, involves the following commands:

1. `> ring R = (0, {parameters listed in any order}),`
   `(variables listed with two projection variables LAST), lp;`

   Note that the '0' noted before the parameter list denotes the characteristic of the field. For the analysis of CRNs, the relevant underlying field is the real numbers, which has characteristic zero. The command 'lp' denotes *lexicographic order*, which is an elimination order on the monomials of the system. The command 'lp' can be always be replaced with the product order '(dp($n-2$), dp(2))', which imposes the much more computationally efficient *degree reverse lexicographic order* on both the block of $(n-2)$ variables to be eliminated and the block of 2 variables that remain. But as we note, the special structure of RPA-capable CRNs allows the Gröbner basis to be computed in polynomial time for these networks, such that the conceptually simpler (albeit computationally more expensive) lexicographic ordering will generally perform well if RPA does indeed obtain.

2. Next, the n mass-action rate equations are listed in any order.

   `> poly f1 = ... ;`

⋮

```
> poly fn = ... ;
```

3. We can now compute the ideal generated by the rate equations $f_1, \ldots, f_n$:

```
> ideal I = f1, ... , fn;
```

4. Finally, we can compute the Gröbner basis for the ideal with the chosen monomial ordering:

```
> ideal GI = groebner(I);
> GI;
```

   For the CRN to be RPA-capable, only the first polynomial in the Gröbner basis (listed in Singular as $GI[1]$) can comprise only the two projection variables, and it must be factorizable into an RPA polynomial.

5. We can now automatically calculate the subsidiary polynomial invariants, as well as a set of polynomials $h_1, \ldots, h_n$ for which $h_1 f_1 + \ldots + h_n f_n = \rho = GI[1]$ via Singular's *lift* command:

```
> lift(I, GI[1]);
```

6. The presence of any conservation laws in the system can also be quickly checked by calculating the syzygies of the system via the *syz* command:

```
syz(I);
```

If conservation laws do exist for the system in question, the module of syzygies will contain generators (one corresponding to each conservation law) containing only integer elements.

We demonstrate the application of this Singular code through a listing of scripts used to analyse the illustrative examples featured in this work.

S5.2 Singular code for Example in Section S1.5 (Figure 1, main paper)

```
> ring R1 = (0, k1, k2, k3, k4, k5, k6, k7, k8), (x3, x1, o1, x2, r), lp;
> poly f1 = k1*r - k2*x1*x2;
> poly f2 = k3 - k2*x1*x2 + k4*x3;
> poly f3 = k5 - k7*o1*x3 - k6*x2*x3;
> poly f4 = k7*o1*x3 - k8*o1;
> ideal I = f1, f2, f3, f4;
> ideal GI = groebner(I);
> GI;
```

This code projects the system onto the two variables $x2$ and r, and generates the following

output:

$$GI[1] = (k1^2 * k6 * k7) * x2 * r^2 + (-2 * k1 * k3 * k6 * k7 - k1 * k4 * k6 * k8) * x2 * r$$
$$+ (k3^2 * k6 * k7 + k3 * k4 * k6 * k8) * x2 + (-k1 * k4 * k5 * k7) * r$$
$$+ (k3 * k4 * k5 * k7 + k4^2 * k5 * k8)$$
$$GI[2] = (k4 * k8) * o1 + (k1 * k6) * x2 * r + (-k3 * k6) * x2 + (-k4 * k5)$$
$$GI[3] = (k1 * k2 * k4 * k5 * k7) * x1 * r + (-k2 * k3 * k4 * k5 * k7 - k2 * k4^2 * k5 * k8) * x1$$
$$+ (-k1^3 * k6 * k7) * r^3 + (2 * k1^2 * k3 * k6 * k7 + k1^2 * k4 * k6 * k8) * r^2$$
$$+ (-k1 * k3^2 * k6 * k7 - k1 * k3 * k4 * k6 * k8) * r$$
$$GI[4] = (k2) * x1 * x2 + (-k1) * r$$
$$GI[5] = (k4) * x3 + (-k2) * x1 * x2 + (k3)$$

As shown in-text, G[1] has the form of an RPA polynomial. We can also repeat this process for the range of alternative projections noted in Section S1.5. For instance, projecting onto $o1$ and r requires the code:

```
> ring R2 = (0, k1, k2, k3, k4, k5, k6, k7, k8), (x3, x1, x2, o1, r), lp;
> poly f1 = k1*r - k2*x1*x2;
> poly f2 = k3 - k2*x1*x2 + k4*x3;
> poly f3 = k5 - k7*o1*x3 - k6*x2*x3;
> poly f4 = k7*o1*x3 - k8*o1;
> ideal I = f1, f2, f3, f4;
> ideal GI = groebner(I);
> GI;
```

which generates the following output in Singular:

$$GI[1] = (k1 * k7) * o1 * r + (-k3 * k7 - k4 * k8) * o1$$

$$GI[2] = (k1 * k6) * x2 * r + (-k3 * k6) * x2 + (k4 * k8) * o1 + (-k4 * k5)$$

$$GI[3] = (k6 * k8) * x2 * o1 + (k7 * k8) * o1^2 + (-k5 * k7) * o1$$

$$\begin{aligned} GI[4] = {} & (k1 * k2 * k4 * k5 * k7) * x1 * r + (-k2 * k3 * k4 * k5 * k7 - k2 * k4^2 * k5 * k8) * x1 \\ & + (-k1^3 * k6 * k7) * r^3 + (2 * k1^2 * k3 * k6 * k7 + k1^2 * k4 * k6 * k8) * r^2 \\ & + (-k1 * k3^2 * k6 * k7 - k1 * k3 * k4 * k6 * k8) * r \end{aligned}$$

$$GI[5] = (k2 * k4 * k8) * x1 * o1 + (-k2 * k4 * k5) * x1 + (k1^2 * k6) * r^2 + (-k1 * k3 * k6) * r$$

$$GI[6] = (k2) * x1 * x2 + (-k1) * r$$

$$GI[7] = (k4) * x3 + (-k2) * x1 * x2 + (k3)$$

From this it is clear by inspection that $GI[1]$ is an RPA polynomial. The subsidiary polynomial invariants and elimination polynomials may now be calculated using the *lift* command:

```
> lift(I, G[1]);
```

which returns the following output:

$$[1, 1] = (k7) * o1$$

$$[2, 1] = (-k7) * o1$$

$$[3, 1] = 0$$

$$[4, 1] = (k4)$$

This reveals that

$$G[1] = k7.o1.f1 - k7.o1.f2 + k4.f4,$$
$$= k7.o1(f1 - f2) + k4(f4),$$

which demonstrates that one concatenating monomial, $o1$, is required, and that the two subsidiary polynomial invariants (opposer polynomials) are $f_1 - f_2 = k_1.r - k_3 - k_4.x_3$ and $f_4 = k_7.o_1.x_3 - k_8.o_1$. The command *syz* reveals that this system has no non-trivial left nullspace (co-kernel). These two linear coordinate changes are thus the canonical transformations for identifying the two subsidiary polynomial invariants.

S5.3 Singular code for Example in Section S3.1 (Cappelletti et al. [3] toy model)

```
> ring R = (0, k1,k2,k3,k4,k5,k6,k7,k8), (E,D,C,B,A), lp;
> poly f1 = -k1*A*B + k7*C;
> poly f2 = k2*B*C - k3*B^2 + k5*E^2 -k4*B;
> poly f3 = k1*A*B -k7*C - k2*B*C + k3*B^2;
> poly f4 = 2*k6*E^2 - k8*D;
> poly f5 = k8*D + 2*k4*B - 2*k5*E^2 - 2*k6*E^2;
> ideal I = f1, f2, f3, f4, f5;
> ideal GI = groebner(I);
> GI;
```

Running this code in Singular generates the following output:

$$GI[1] = (k1 * k2) * B^2 * A + (-k3 * k7) * B^2$$

$$GI[2] = (k7) * C + (-k1) * B * A$$

$$GI[3] = (k5 * k8) * D + (-2 * k4 * k6) * B$$

$$GI[4] = (2 * k6) * E^2 + (-k8) * D$$

from which it is clear that $GI[1]$ is an RPA polynomial. Using the *lift* command now calculates the subsidiary polynomial invariants and elimination polynomials:

```
> lift(I, G[1]);
```

which returns the following output:

$$[1, 1] = (-k2) * B$$

$$[2, 1] = (k7)$$

$$[3, 1] = 0$$

$$[4, 1] = (k7)/2$$

$$[5, 1] = (k7)/2$$

This demonstrates that one concatenating monomial, B, is required, and that the two subsidiary polynomial invariants (opposer polynomials) are $f_1 = -k_1.A.B + k_7.C$ (the connector invariant) and $k_7(f_2 + (1/2)f_4 + (1/2)f_5) = k_2.B.C - k_3.B^2$ (the balancer invariant).

The command *syz* reveals that this system has one conservation law, with

```
> syz{I};
```

producing an output in which only one generator contains integer elements:

$$[1] = gen(5) + gen(4) + 2 * gen(3) + 2 * gen(2) + 2 * gen(1)$$

This demonstrates that, for this CRN, $E + D + 2C + 2B + 2A$ is a constant for all time. Thus, the linear coordinate changes noted above to produce the balancer/connector invariants are not unique. But with the connector and balancer invariants automatically calculated by this algorithm, the canonical linear transformation can quickly be determined by routine calculation by the methods developed by [9, 3].

S5.4 Singular code for Example in Section S3.2 (EnvZ-OmpR Osmoregulation, Figure 2, main paper)

```
> ring R = (0, a1, d1, k1, a2, d2, k2, a3, d3, k3, a4, d4, k4, a5, d5),
                  (x9, x8, x7, x6, x5, x4, x3, x2, x1), lp;
> poly f1 = a1*x4 - (d1+k1)*x1 - a3*x1*x2 + (d3+k3)*x8;
> poly f2 = k2*x7 - a3*x1*x2 + d3*x8 - a4*x3*x2 + d4*x9;
> poly f3 = d5*x4 - a5*x3 - a4*x3*x2 + (k4+d4)*x9;
> poly f4 = a5*x3 + d1*x1 - (a1+d5)*x4 + k2*x7;
> poly f5 = k1*x1 - a2*x5*x6 + d2*x7;
> poly f6 = d2*x7 - a2*x5*x6 + k3*x8 + k4*x9;
> poly f7 = a2*x5*x6 - (d2+k2)*x7;
> poly f8 = a3*x1*x2 - (d3+k3)*x8;
> poly f9 = a4*x2*x3 - (d4+k4)*x9;
> ideal I = f1, f2, f3, f4, f5, f6, f7, f8, f9;
> ideal GI = groebner(I);
> GI;
```

Running this code in Singular generates the following output:

$$
\begin{aligned}
GI[1] &= (a1*a3*k3*d4*a5 + a1*a3*k3*k4*a5 + d1*d3*a4*k4*d5 \\
&\quad + d1*k3*a4*k4*d5 + k1*d3*a4*k4*d5 + k1*k3*a4*k4*d5)*x2*x1 \\
&\quad + (-a1*k1*d3*d4*a5 - a1*k1*d3*k4*a5 - a1*k1*k3*d4*a5 \\
&\quad - a1*k1*k3*k4*a5)*x1 \\
GI[2] &= (a1*a5)*x3 + (-d1*d5 - k1*d5)*x1 \\
GI[3] &= (d5)*x4 + (-a5)*x3 \\
GI[4] &= (a2*k2)*x6*x5 + (-a1*d2 - a1*k2 - d2*d5 - k2*d5)*x4 \\
&\quad + (d2*a5 + k2*a5)*x3 + (d1*d2 + d1*k2)*x1 \\
GI[5] &= (k2)*x7 + (-a1 - d5)*x4 + (a5)*x3 + (d1)*x1 \\
GI[6] &= (d3 + k3)*x8 + (-a3)*x2*x1 \\
GI[7] &= (k4)*x9 + (k3)*x8 + (-k1)*x1
\end{aligned}
$$

It is clear that $G[1]$ is an RPA polynomial. In addition running the *lift* command,

```
> lift(I, GI[1]);
```

produces the following output:

$$[1,1] = (-a1 * d3 * a4 * k4 - a1 * k3 * a4 * k4 - d3 * a4 * k4 * d5 - k3 * a4 * k4 * d5) * x2$$

$$[2,1] = 0$$

$$[3,1] = 0$$

$$[4,1] = (-a1 * d3 * a4 * k4 - a1 * k3 * a4 * k4) * x2$$

$$[5,1] = (-a1 * d3 * a4 * k4 - a1 * k3 * a4 * k4) * x2$$
$$+ (-a1 * d3 * d4 * a5 - a1 * d3 * k4 * a5 - a1 * k3 * d4 * a5 - a1 * k3 * k4 * a5)$$

$$[6,1] = (a1 * d3 * d4 * a5 + a1 * d3 * k4 * a5 + a1 * k3 * d4 * a5 + a1 * k3 * k4 * a5)$$

$$[7,1] = (-a1 * d3 * a4 * k4 - a1 * k3 * a4 * k4) * x2$$

$$[8,1] = (-a1 * d3 * a4 * k4 - a1 * k3 * a4 * k4 - d3 * a4 * k4 * d5 - k3 * a4 * k4 * d5) * x2$$
$$+ (a1 * k3 * d4 * a5 + a1 * k3 * k4 * a5)$$

$$[9,1] = (a1 * d3 * k4 * a5 + a1 * k3 * k4 * a5)$$

Thus, only one concatenating monomial $x2$ is required, and there are two subsidiary polynomial invariants, namely

$$-\alpha f_5 + \alpha f_6 + \beta f_8 + \gamma f_9 = (a1 * d3 * a4 * k4 * a5 + a1 * k3 * a4 * k4 * a5) * x3 * x2$$
$$+ (a1 * a3 * k3 * d4 * a5 + a1 * a3 * k3 * k4 * a5) * x2 * x1$$
$$+ (-a1 * k1 * d3 * d4 * a5 - a1 * k1 * d3 * k4 * a5 - a1 * k1 * k3 * d4 * a5 - a1 * k1 * k3 * k4 * a5) * x1,$$

being the connector invariant, with $\alpha = (a1*d3*d4*a5+a1*d3*k4*a5+a1*k3*d4*a5+a1*k3*k4*a5)$, $\beta = (a1*k3*d4*a5+a1*k3*k4*a5)$, and $\gamma = (a1*d3*k4*a5+a1*k3*k4*a5)$;

and

$$(\eta+\zeta)(f_1+f_8)+\eta(f_4+f_5+f_7)=(-a1*d3*a4*k4*a5-a1*k3*a4*k4*a5)*x3$$
$$+(d1*d3*a4*k4*d5+d1*k3*a4*k4*d5+k1*d3*a4*k4*d5+k1*k3*a4*k4*d5)*x1,$$

being the balancer invariant, with $\eta = -a1*d3*a4*k4-a1*k3*a4*k4$, and $\zeta = -d3*a4*k4*d5-k3*a4*k4*d5$.

The command,

```
> syz(I);
```

produces an output for which only two generators contain integer elements:

$$[1] = gen(6)-gen(5)-gen(4)-gen(3)+gen(2)-gen(1)$$
$$[2] = gen(9)+gen(8)+gen(7)+gen(5)+gen(4)+gen(3)+gen(1)$$

This confirms that there are two conservation laws, with $x_9+x_8+x_7+x_5+x_4+x_3+x_1$ equal to a constant for all time (conserving the total abundance of X), and $x_6+x_2-(x_5+x_4+x_3+x_1) = x_6+x_2+x_9+x_8+x_7$ equal to a constant for all time (conserving the total abundance of Y). Thus, once again, the linear coordinate changes by which the connector and balancer invariants above are obtained are not unique. But having systematically calculated these two subsidiary invariants as shown above, the canonical linear transformation can quickly be determined by routine calculation by the methods developed by [9, 3]. It is this canonical transformation that we depict in Figure 2c of the main paper.

References

[1] Robyn P Araujo and Lance A Liotta. The topological requirements for robust perfect adaptation in networks of any size. *Nature Communications*, 9(1):1–12, 2018.

[2] Corentin Briat, Ankit Gupta, and Mustafa Khammash. Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks. *Cell systems*, 2(1):15–26, 2016.

[3] Daniele Cappelletti, Ankit Gupta, and Mustafa Khammash. A hidden integral structure endows absolute concentration robust systems with resilience to dynamical concentration disturbances. *Journal of the Royal Society Interface*, 17(171):20200437, 2020.

[4] Martin Feinberg. *Foundations of chemical reaction network theory*. Springer, 2019.

[5] Martin Feinberg and Friedrich JM Horn. Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspaces. *Archive for Rational Mechanics and Analysis*, 66(1):83–97, 1977.

[6] Bruce A Francis and Walter Murray Wonham. The internal model principle of control theory. *Automatica*, 12(5):457–465, 1976.

[7] Bruce A Francis and William M Wonham. The internal model principle for linear multivariable regulators. *Applied mathematics and optimization*, 2(2):170–194, 1975.

[8] Fritz Horn and Roy Jackson. General mass action kinetics. *Archive for rational mechanics and analysis*, 47(2):81–116, 1972.

[9] Robert L Karp, Mercedes Pérez Millán, Tathagata Dasgupta, Alicia Dickenstein, and Jeremy Gunawardena. Complex-linear invariants of biochemical networks. *Journal of theoretical biology*, 311:130–138, 2012.

[10] Guy Shinar and Martin Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327(5971):1389–1391, 2010.

[11] Oren Shoval, Uri Alon, and Eduardo Sontag. Symmetry invariance for adapting biological systems. *SIAM Journal on Applied Dynamical Systems*, 10(3):857–886, 2011.

[12] Eduardo D Sontag. Adaptation and regulation with signal detection implies internal model. *Systems & control letters*, 50(2):119–126, 2003.

[13] Fangzhou Xiao and John C Doyle. Robust perfect adaptation in biomolecular reaction networks. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 4345–4352. IEEE, 2018.