Complex-linear invariants of biochemical networks

Robert L. Karp [a,1], Mercedes Pérez Millán [b,1], Tathagata Dasgupta [a,1], Alicia Dickenstein [b,c], Jeremy Gunawardena [a,*]

[a] Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[b] Departamento de Matemática, FCEN, Universidad de Buenos Aires, Argentina
[c] IMAS-CONICET, C1428EGA Buenos Aires, Argentina

HIGHLIGHTS

► Invariants are polynomial expressions that vanish in any steady state of a network.
► They characterise networks arising in various biological contexts.
► We introduce an efficient linear method for calculating certain types of invariants.
► This exploits Chemical Reaction Network Theory for networks of arbitrary deficiency.
► The new method reveals robustness in complex bifunctional enzymes.

ARTICLE INFO

ABSTRACT

The nonlinearities found in molecular networks usually prevent mathematical analysis of network behaviour, which has largely been studied by numerical simulation. This can lead to difficult problems of parameter determination. However, molecular networks give rise, through mass-action kinetics, to polynomial dynamical systems, whose steady states are zeros of a set of polynomial equations. These equations may be analysed by algebraic methods, in which parameters are treated as symbolic expressions whose numerical values do not have to be known in advance. For instance, an "invariant" of a network is a polynomial expression on selected state variables that vanishes in any steady state. Invariants have been found that encode key network properties and that discriminate between different network structures. Although invariants may be calculated by computational algebraic methods, such as Gröbner bases, these become computationally infeasible for biologically realistic networks. Here, we exploit Chemical Reaction Network Theory (CRNT) to develop an efficient procedure for calculating invariants that are linear combinations of "complexes", or the monomials coming from mass action. We show how this procedure can be used in proving earlier results of Horn and Jackson and of Shinar and Feinberg for networks of deficiency at most one. We then apply our method to enzyme bifunctionality, including the bacterial EnvZ/OmpR osmolarity regulator and the mammalian 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase glycolytic regulator, whose networks have deficiencies up to four. We show that bifunctionality leads to different forms of concentration control that are robust to changes in initial conditions or total amounts. Finally, we outline a systematic procedure for using complex-linear invariants to analyse molecular networks of any deficiency.

1. Introduction

A molecular interaction network within a cell may be decomposed into elementary biochemical reactions, such as

$$2S_1 + 3S_2 \rightarrow 4S_3, \tag{1}$$

where the S_i are distinct chemical species. (This stoichiometry is unlikely but helpful for illustrative purposes.) Under mass-action kinetics, the rate of such a reaction is proportional to the concentrations of the substrates, taking stoichiometry into account. Hence,

$$\frac{dx_3}{dt} = 4\kappa_1 x_1^2 x_2^3, \tag{2}$$

where x_i is the concentration of species S_i, $x_i = [S_i]$, and $\kappa_1 \in \mathbb{R}_{>0}$ is the positive mass-action rate constant. In a biochemical network,

reactions contribute production and consumption terms consisting of monomials like $4\kappa_1 x_1^2 x_2^3$ to the rates of formation of the species in the network. This results in a system of ordinary differential equations (ODEs), $dx/dt = f(x;\kappa)$, in which each component rate function $f_i(x;\kappa)$ is a polynomial in the state variables $x_1, x_2, \ldots, x_n \in \mathbb{R}$ and $\kappa_1, \ldots, \kappa_p \in \mathbb{R}_{>0}$ are positive rate constants.

The nonlinearities in (2) usually preclude mathematical analysis of the dynamical behaviour of such ODE systems, which are customarily studied by numerical simulation. This requires that the rate constants be given numerical values, which in most cases are neither known nor readily measurable. The resulting "parameter problem" remains a major difficulty in exploiting mathematical models (Gunawardena, 2010). However, the steady states of such ODEs are zeros of a set of polynomial equations, $f_1(x,\kappa) = 0, \ldots f_n(x,\kappa) = 0$. Computational algebra and algebraic geometry provide powerful tools for studying these solutions (Cox et al., 1997), and these have recently been used to gain new biological insights (Craciun et al., 2009; Dasgupta et al., submitted for publication; Manrai and Gunawardena, 2008; Pérez Millán et al., 2012; Thomson and Gunawardena, 2009a,b). The rate constants can now be treated as symbolic parameters, whose numerical values do not need to be known in advance. The capability to rise above the parameter problem allows more general results to be obtained than can be expected from numerical simulation (Thomson and Gunawardena, 2009b).

The focus on steady states, rather than transient dynamics, is still of substantial interest. For instance, in time-scale separation, which has been a widespread method of simplification in biochemistry and molecular biology, a fast sub-system is assumed to be at steady state with respect to a slower environment and steady-state analysis is used to eliminate the internal complexity in the sub-system (Gunawardena, 2012). Approximate or quasi-steady states have also been shown to exist under various cellular conditions and can now be engineered *in vivo* (Laurent and Kellershohn, 1999; Kramer et al., 2004). Finally, steady states provide the skeleton around which the transient dynamics unfolds, so knowledge of the former can be helpful for understanding the latter.

The present paper focusses on the algebraic concept of an "invariant": a polynomial expression on selected state variables that is zero in any steady state, with the coefficients of the expression being rational expressions in the symbolic rate constants (Manrai and Gunawardena, 2008). Recall that a rational expression is a quotient of two polynomials; an example of such being the classical Michaelis–Menten constant of an enzyme (Cornish-Bowden, 1995). (A more general definition of an invariant allows the coefficients to include conserved quantities (Xu and Gunawardena, submitted for publication), but this extension is not discussed here.) Since each of the rate functions, $f_i(x;\kappa)$, is zero in any steady state, the force of the definition comes from the restriction to "selected state variables". It is possible that, by performing appropriate algebraic operations on $f_1, \ldots, f_n$, non-selected variables can be eliminated, leaving a polynomial expression on only the selected variables that must be zero in any steady state.

Invariants turn out to be surprisingly useful. They have been shown to characterise the biochemical networks underlying multisite protein phosphorylation (Manrai and Gunawardena, 2008), suggesting that different network architectures can be identified through experimental measurements at steady state. If an invariant has only a single selected variable that appears linearly, this variable has the same value in any steady state since it is determined solely by the rate constants. In particular, its value is unaffected by changes to the initial conditions or to the total amounts of any species. This is "absolute concentration robustness" (ACR), as introduced in Shinar and Feinberg (2010), which accounts for experimental findings in some bacterial bifunctional enzymes (Batchelor and Goulian, 2003; Shinar et al., 2007, 2009).

The mammalian bifunctional enzyme, 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFK-2/FBPase-2), which has a more complex enzymatic network, also yields invariants, with implications for regulation of glycolysis (Dasgupta et al., submitted for publication). The methods developed here provide a systematic way to analyse such bifunctional enzymes, as explained below.

Computational algebra exploits the method of Gröbner bases to provide an Elimination Theorem (Cox et al., 1997), that permits variables to be systematically eliminated among the rate equations, $f_1, \ldots, f_n$ (Manrai and Gunawardena, 2008). Algorithms for calculating Gröbner bases are available in general-purpose tools like Mathematica, Matlab and Maple and in specialised mathematical packages such as Singular and Macaulay2.[2] However, these algorithms are computationally expensive for the task at hand. They have been developed for general sets of polynomials and have not been optimised for those coming from biochemical networks. For instance, Mathematica's Gröbner basis algorithm does not terminate on the network for PFK-2/FBPase-2. If invariants are to be exploited further, alternative approaches are needed.

The nonlinearity in mass action comes from the pattern of substrate stoichiometry in (1), which gives rise to the monomial nonlinearity in (2). In the language of Chemical Reaction Network Theory (CRNT), the patterns of stoichiometry that appear on either side of a reaction arrow are called "complexes" (Feinberg, 1979; Gunawardena, 2003). Reaction (1) has three species, S_1, S_2 and S_3 and two complexes, $2S_1 + 3S_2$ and $4S_3$. If $C = e_1 S_1 + \cdots + e_n S_n$, where $e_i \in \mathbb{Z}_{\geq 0}$ are nonnegative integer stoichiometries, then the complex C gives the monomial, x^C, where $x^C = x_1^{e_1} \cdots x_n^{e_n}$.

Aside from this nonlinearity, the defining rate equations come from linear processes on complexes. This observation is the starting point of CRNT and reveals that biochemical networks conceal much linearity behind their nonlinearity (Feinberg, 1979; Gunawardena, 2003, and see below). This suggests the possibility of using fast linear methods, in preference to slow polynomial algorithms, to construct a subset of invariants: those that are symbolic linear combinations of the complex monomials, x^C. As before, this definition acquires substance by restricting the complexes that can appear. If $C_1, \ldots, C_k$ are the selected complexes, then a complex-linear invariant is a polynomial expression of the form $a_1 x^{C_1} + \cdots + a_k x^{C_k}$, that is zero in any steady state, where $a_1, \ldots, a_k$ may be rational expressions in the symbolic rate constants.

In this paper, we examine a large class of complex-linear invariants that we call "type 1". We determine the dimension of the space of type 1 invariants (Proposition 1) and provide a linear algorithm for calculating them (Theorem 1). We point out how invariants can be used in proving previous results of Horn and Jackson (1972), and the Shinar–Feinberg Theorem for ACR (Shinar and Feinberg, 2010). We then apply the method to contrast two examples of enzymatic bifunctionality, the bacterial EnvZ/OmpR osmolarity regulator and the mammalian PFK-2/FBPase-2 glycolytic regulator. The method is sufficiently straightforward that the invariants for networks of this kind can be found by manual inspection of an appropriate matrix. Finally, we outline a systematic procedure for analysing any network using type 1 complex-linear invariants.

2. Results

2.1. Background on CRNT

It is assumed that there are n species, $\mathcal{S} = \{S_1, \ldots, S_n\}$, whose concentrations are $x_1, \ldots, x_n \in \mathbb{R}$, respectively, and m complexes,

---

$\mathcal{C} = \{C_1, \ldots, C_m\}$. Complexes are regarded formally as multisets of species, $C_i \in \mathbb{N}^S$, where the value of the multiset, C_i, on the species S_j, denoted $C_i(S_j)$ is the stoichiometry of S_j in C_i. Accordingly, $C_i = C_i(S_1)S_1 + \cdots + C_i(S_n)S_n$. The reactions in the network define a directed graph on the complexes, with an edge $C_i \to C_j$ whenever there is a reaction with substrate stoichiometry given by C_i and product stoichiometry given by C_j. The corresponding mass-action rate constant gives each edge a label, $\kappa_{i,j}$, treated as a positive symbol, $\kappa_{i,j} \in \mathbb{R}_{>0}$.

Any directed graph, G, with labels in $\mathbb{R}_{>0}$ gives rise to an abstract dynamics in which each edge is treated as if it were a first-order chemical reaction with its label as rate constant. Since the rates are all first-order, the dynamics are linear and may therefore be written in matrix terms as $dy/dt = \mathcal{L}(G) \cdot y$, where $y \in \mathbb{R}^m$ is a column vector, consisting of an abstract concentration y_i at each node i of G, and $\mathcal{L}(G)$ is a $m \times m$ matrix called the Laplacian matrix of G. Here, "$\cdot$" signifies matrix multiplication, regarding vectors as matrices of one row or one column. The graph Laplacian has wide application in biology, as described in Gunawardena (2012), where additional information may be found.

In CRNT, the Laplacian, $\mathcal{L}(G): \mathbb{R}^m \to \mathbb{R}^m$, provides a linear analogue for complexes of the nonlinear function, $f: \mathbb{R}^n \to \mathbb{R}^n$, for species, in the following sense. Let $\Psi: \mathbb{R}^n \to \mathbb{R}^m$ be the non-linear function that lists the monomials for each complex, $\Psi(x) = (x^{C_1}, \ldots, x^{C_m})^{\dagger}$. Here, $\dagger$ denotes transpose. Let $Y: \mathbb{R}^m \to \mathbb{R}^n$ be the linear function that associates to each complex, considered as a basis element of $\mathbb{R}^m$, its corresponding stoichiometry pattern; the i-th column of the resulting matrix is then $(C_i(S_1), \ldots, C_i(S_n))^{\dagger}$. With these definitions, it may be checked that $f(x) = Y \cdot \mathcal{L}(G) \cdot \Psi(x)$, for any $x \in \mathbb{R}^n$, as depicted in the commutative diagram in the top left of Fig. 1.

This fundamental decomposition is due to Horn and Jackson (1972), and is the starting point of CRNT. They did not use the Laplacian description, which was introduced in Craciun et al. (2009) and exploited further in Gunawardena (2012) and Thomson and Gunawardena (2009a). To analyse steady states, where $f(x)=0$, it is particularly useful to know the kernel of $\mathcal{L}(G)$: $\ker \mathcal{L}(G) = \{y \in \mathbb{R}^m | \mathcal{L}(G) \cdot y = 0\}$. This was first determined by Feinberg and Horn (1977, Appendix), by a non-constructive method. Here, we briefly describe the constructive method introduced in Gunawardena (2012), which shows how $\ker \mathcal{L}(G)$ can be algorithmically calculated from G.

This can be done in two stages (Gunawardena, 2012). First, if G is "strongly connected", so that any two distinct nodes are linked by a contiguous series of edges in the same direction, then $\dim \ker \mathcal{L}(G) = 1$. The Matrix-Tree Theorem provides an explicit construction of a basis element, $\rho_G \in \mathbb{R}^m$, in terms of the spanning trees of G: $\ker \mathcal{L}(G) = \langle \rho_G \rangle$. The components $(\rho_G)_i$ are polynomials in the symbolic labels but the details of their calculation are not needed here. If G is not strongly connected, it can be partitioned into its maximal strongly connected sub-graphs, or "strongly connected components" (SCCs). These inherit from G a directed graph structure, $\overline{G}$, in which there is an edge in $\overline{G}$ from SCC G_u to SCC G_v whenever there is an edge in G from some node in G_u to some node in G_v. $\overline{G}$ cannot have any directed cycles and so always has terminal SCCs, with no edges leaving them. Let these be $G_1, \ldots, G_T$. For each $1 \le t \le T$, let $\rho^t \in \mathbb{R}^m$ be the vector which, for vertices of G that lie in G_t, agrees with the vector ρ_{G_t}, coming from the Matrix-Tree Theorem applied to G_t as an isolated graph, and, for all other vertices, j, $(\rho^t)_j = 0$. Then, the ρ^t forms a basis for $\ker \mathcal{L}(G)$:

$$\ker \mathcal{L}(G) = \langle \rho^1, \ldots, \rho^T \rangle. \tag{3}$$

This gives algebraic expressions for the components of the basis vectors as polynomials in the symbolic labels. Note that ρ^t may be very sparse, being non-zero only for vertices in the single SCC G_t.
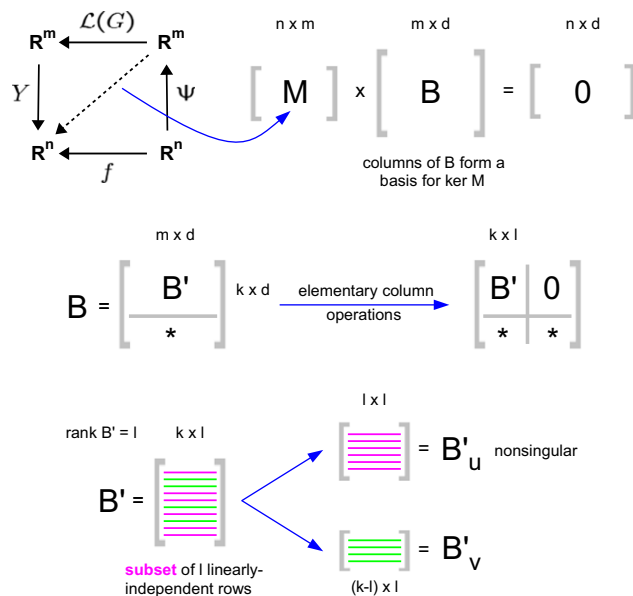
Fig. 1. Schematic illustration of the dual procedure for calculating type 1 complex-linear invariants using (6). The fundamental decomposition of CRNT is shown in the commutative diagram in the top left corner, for which $f(x) = Y \cdot \mathcal{L}(G) \cdot \Psi(x)$. The matrix M corresponds to the linear part of this decomposition, $M = Y \cdot \mathcal{L}(G)$. The matrix B has columns consisting of a basis for $\ker M$. The matrix B' consists of the first k rows of B. Assuming that $rkB' = l$, elementary column operations can be applied to B to bring it into lower-triangular block form. By a mild abuse of notation, the upper-left block continues to be called B'. Finally, a subset, U, of linearly independent rows (magenta) of B' yields the nonsingular matrix B'_U, while the subset V of remaining rows (green) yields the matrix B'_V, from which type 1 complex-linear invariants can be calculated using (6). For interpretation of the references to color in this figures caption, the reader is referred to the web version of this article.

2.2. Generating complex-linear invariants

Depending on the application, invariants may be required that involve only certain complexes, $C_{i_1}, \ldots, C_{i_k}$, for instance, those involving species with more easily measurable concentrations. Since the indices can be permuted so that the complexes of interest appear first in the ordering, it can be assumed that invariants are sought on $C_1, \ldots, C_k$. Let M be the $n \times m$ matrix representing the linear part of the CRNT decomposition, $M = Y \cdot \mathcal{L}(G)$. A simple way to construct a complex-linear invariant on $C_1, \ldots, C_k$ is to find a vector, $a^{\dagger} \in \mathbb{R}^k$, such that, if $(a,0)^{\dagger} \in \mathbb{R}^m$ is a extended with $m-k$ zeros, $(a,0) = (a_1, \ldots, a_k, 0, \ldots, 0)$, then $(a,0)$ is in the rowspan of M. That is, it is a linear combination of the rows of M. If $x \in \mathbb{R}^n$ is any steady state of the system, so that $f(x)=0$, then $\Psi(x) \in \ker M$ because $M \cdot \Psi(x) = Y \cdot \mathcal{L}(G) \cdot \Psi(x) = f(x) = 0$. Since $(a,0)$ is in the rowspan of M, $(a,0) \cdot \Psi(x) = 0$. Hence, by definition of Ψ, $a_1 x^{C_1} + \cdots + a_k x^{C_k} = 0$, giving a complex-linear invariant on $C_1, \ldots, C_k$.

Not all such invariants may arise in this way. For that to happen, it is necessary not just for $(a,0) \cdot \Psi(x) = 0$ whenever x is a steady state but for $(a,0) \cdot v = 0$ for all $v \in \ker M$. The relationship between $\ker M$ and $\{\Psi(x) | x$ is a steady state$\}$ is not straightforward. To sidestep this problem, we focus here only on those invariants, $a_1 x^{C_1} + \cdots + a_k x^{C_k}$, in which $(a,0)$ is in the rowspan of M. We call these type 1 complex-linear invariants. Non-type 1 invariants do exist, as we show in the Supporting Information (SI). The type 1 invariants form a vector space that we abbreviate I_k; note that I_k depends on $C_1, \ldots, C_k$ and not just on k. Two basic problems are, first, to determine the dimension of I_k and, second, to generate its elements.

A simple solution to the second problem is to break the matrix M into the $n \times k$ sub-matrix K consisting of the first k columns of M and the $n \times (m-k)$ sub-matrix N consisting of the remaining $m-k$ columns, so that $M = K|N$. Any vector $b^{\dagger} \in \mathbb{R}^n$ which is in the

left null space of N, $b \in \mathcal{N}_L(N)$, so that $b \cdot N = 0$, gives an $(a,0) = b \cdot M$ that is in the rowspan of M. The assignment $b \to b \cdot M$ thereby defines a surjection, $\mathcal{N}_L(N) \to I_k$. Moreover, $b_1 \cdot M = b_2 \cdot M$, if, and only if, $(b_1 - b_2) \in \mathcal{N}_L(M) \subseteq \mathcal{N}_L(N)$. Hence, there is an isomorphism $I_k \cong \mathcal{N}_L(N)/\mathcal{N}_L(M)$. If X is any $n \times r$ matrix, $\dim \mathcal{N}_L(X) = n - rkX$. We conclude that $\dim I_k = rkM - rkN$, which yields an efficient way to determine the dimension of I_k by Gaussian elimination.

In principle, this provides an automatic procedure for identifying subsets of complexes with non-trivial invariants. First determine rkM and then, for each subset $Z \subseteq \{1, \ldots, m\}$, determine the rank of the submatrix of M formed by those columns not in Z. If the latter is smaller than the former, then there are non-trivial type 1 complex-linear invariants on the complexes in Z. In practice, it is usually more efficient to use biological knowledge of the example being studied and the question being asked to narrow the choice of Z. We outline such a systematic procedure for finding invariants in the last section.

The method above amounts to eliminating the complexes $C_{k+1}, \ldots, C_m$ by taking linear combinations of the defining rate functions, $f_1, \ldots, f_n$. This can be biologically informative because it suggests which rate functions, and, hence, which species at steady state, determine the invariant (Dasgupta et al., submitted for publication).

2.3. Duality and the structure of I_k

In this section, we present an alternative procedure for calculating complex-linear invariants, which is based on duality and exploits the sparsity of (3). The procedure is schematically illustrated in Fig. 1, as an aid to following the details.

We start, as before, with the $n \times m$ matrix, M that represents the linear part of the CRNT decomposition, $M = Y \cdot \mathcal{L}(G)$. Let $d = \dim \ker M$ and let B be any $m \times d$ matrix whose columns form a basis of $\ker M$. Then, $M \cdot B = 0$ and the rowspan of M and the columnspan of B are dual spaces of each other. If $a^{\dagger} \in \mathbb{R}^k$, then $(a,0)$ is in the rowspan of M if, and only if, $(a,0) \cdot B = 0$. If B' is the $k \times d$ sub-matrix of B consisting of the first k rows, then $(a,0) \cdot B = 0$ if, and only if, $a \cdot B' = 0$. Hence, type 1 invariants form the dual space to the columns of B'.

Proposition 1. *The space I_k of type 1 complex-linear invariants on the complexes $C_1, \ldots, C_k$ satisfies $\dim I_k = rkM - rkN = k - rkB'$.*

Let $l = rkB'$. Note that $l \leq \min(k,d)$. If $l = k$, then $\dim I_k = 0$ and there are no type 1 invariants on $C_1, \ldots, C_k$. If, however, $l < k$, then the original matrix B can be simplified in two steps. First the columns. Since the column rank of B' is l, elementary column operations – interchange of two columns, multiplication of a column by a scalar, addition of one column to another – can be applied to the columns of B', to bring the last $d - l$ columns to zero. If exactly the same elementary column operations are applied to the full matrix B, a new matrix is obtained, which we still call B, whose columns still form a basis for $\ker M$. B is now in lower-triangular block form

$$B = \begin{pmatrix} B' & 0 \\ * & * \end{pmatrix}, \tag{4}$$

where, as before, B' is the $k \times l$ sub-matrix consisting of the first k rows and l columns.

For the rows, since the row rank of B' is still l, there are l rows of B' that are linearly independent. Let $U \subseteq \{1, \ldots, k\}$ be the corresponding subset of l indices and let $V \subseteq \{1, \ldots, k\}$ be the subset of $k - l$ remaining indices. This defines a partition of the row indices of $B': U \cap V = \emptyset$ and $U \cup V = \{1, \ldots, k\}$. Let B'_U be the $l \times l$ sub-matrix of B' consisting of the rows with indices in U and B'_V be the $(k - l) \times l$ sub-matrix consisting of the remaining rows of B'.

Using the same notation for $a^{\dagger} \in \mathbb{R}^k$, $a \cdot B' = 0$ if, and only if, $a_U \cdot (B'_U) + a_V \cdot (B'_V) = 0$. Since, by construction, B'_U has full rank and is hence invertible, this may be rewritten as

$$a_U = -a_V \cdot (B'_V) \cdot (B'_U)^{-1}. \tag{5}$$

This gives a non-redundant procedure for generating all elements of I_k by choosing $a_V^{\dagger} \in \mathbb{R}^{k-l}$ arbitrarily and $a_U^{\dagger} \in \mathbb{R}^l$ to satisfy (5). The resulting $a^{\dagger} \in \mathbb{R}^k$ satisfy $a \cdot B' = 0$ and give exactly the type 1 complex-linear invariants on $C_1, \ldots, C_k$.

Using the same notation for $\Psi(x) \in \mathbb{R}^m$, the invariants themselves are given by $a_U \cdot \Psi(x)_U + a_V \cdot \Psi(x)_V = 0$, for any steady state $x \in \mathbb{R}^n$. Substituting (5) and rearranging gives $a_V \cdot (\Psi(x)_V - (B'_V) \cdot (B'_U)^{-1} \cdot \Psi(x)_U) = 0$. Since a_V can be chosen arbitrarily in the dual space, we conclude that

$$\Psi(x)_V = (B'_V) \cdot (B'_U)^{-1} \cdot \Psi(x)_U, \tag{6}$$

which we summarise as follows.

Theorem 1. *Each of the $k - l$ rows of the matrix equation in (6) gives an independent type 1 complex-linear invariant on $C_1, \ldots, C_k$.*

This procedure relies on the choice of basis elements for $\ker M$ that make up the columns of B and on the choice of the subset, U, of linearly independent rows of B'. These choices are not critical; different ones yield different bases for I_k. Up to linear combinations, the same invariants are found irrespective of the choices. All the calculations required are linear and can be readily undertaken in any computer algebra system with the rate constants treated as symbols. (Mathematica was used for the calculations in the SI.) The coefficients are then rational expressions in the symbolic rate constants.

2.4. Haldane relationships and the Shinar–Feinberg theorem

Since $M = Y \cdot \mathcal{L}(G)$, $\ker M$ contains the subspace $\ker \mathcal{L}(G)$. The structure of the latter is known from (3). This should assist in the calculation of invariants, especially when $\ker \mathcal{L}(G)$ is close to $\ker M$. We discuss two instances of this, which illustrate how complex-linear invariants are related to previous studies.

Define the "dynamic deficiency" of a biochemical network, $\delta_D \in \mathbb{N}_{\geq 0}$, to be the difference in dimension between the two subspaces: $\delta_D = \dim \ker M - \dim \ker \mathcal{L}(G)$, or, equivalently, $\delta_D = \dim(\ker Y \cap \text{Image } \mathcal{L}(G))$. This is different from the "deficiency" as usually defined in CRNT (Feinberg, 1979; Gunawardena, 2003), which we call the "structural deficiency", $\delta_S \in \mathbb{N}_{\geq 0}$. While δ_D may depend on the values of rate constants, δ_S is independent of them. However, the former is more convenient for our purposes.

It is known that $\delta_D \leq \delta_S$. Furthermore, if there is only a single terminal SCC in each connected component of G, which holds for the graph in Fig. 2C but not for that in Fig. 4A, then $\delta_D = \delta_S$ (Feinberg, 1979; Gunawardena, 2003). Recall that a graph is connected if any two distinct nodes are linked by a path of contiguous edges, ignoring directions. A connected component of G is then a maximal connected sub-graph. Distinct connected components are totally disconnected, with no edges between them.

Suppose first that $\delta_D = 0$ and that there is a positive steady state $x \in (\mathbb{R}_{>0})^n$. Since $\mathcal{L}(G) \cdot \Psi(x) = 0$, x is a "complex-balanced" steady state, in the terminology of (Horn and Jackson, 1972). According to (3), the vectors ρ^t provide a basis for $\ker M = \ker \mathcal{L}(G)$ and, furthermore, $(\rho^t)_j \neq 0$ if, and only if, $C_j \in G_t$. Choose any terminal SCC of G, which we may suppose to be G_1, and suppose that $C_1, \ldots, C_k$ are the complexes in G_1. Choose the matrix B so that ρ^1 is its first column and the other ρ^t for $t > 1$ are assigned to columns arbitrarily. By construction, B is already in lower-triangular block form and $l = 1$. Setting $U = \{1\}$ and $V = \{2, \ldots, k\}$ the $k - 1$ type 1 invariants coming from (6) are
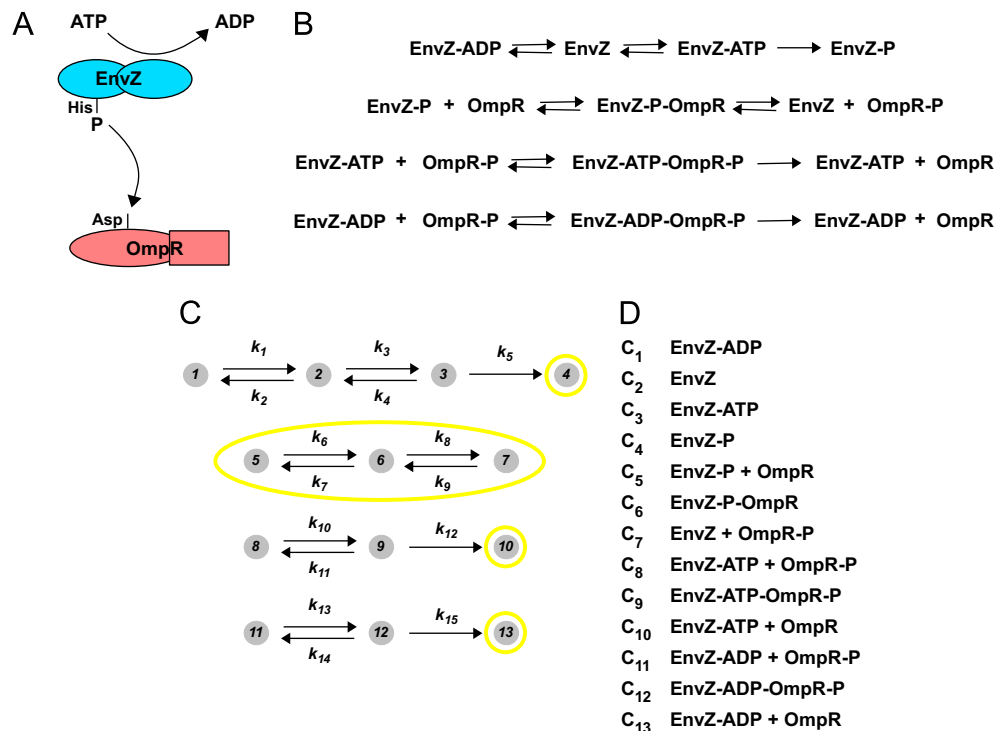
Fig. 2. Two component signalling and the *E. coli* osmolarity network. (A) Schematic of two-component phospho-transfer between a histidine residue on the autophosphorylating sensor kinase (light blue) and an aspartate on the response regulator (red). (B) Extended reaction network for the EnvZ/OmpR two-component osmoregulator in *E. coli*, following Shinar et al. (2007, Supplementary Information). Hyphens, as in EnvZ-ATP, indicate the formation of a biochemical complex between the components. (C) Corresponding labelled, directed graph on the complexes, with the terminal strongly connected components outlined in yellow. Each connected component has only a single terminal SCC. (D) Numbering scheme for the complexes. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

$x^{C_i} = ((\rho^1)_i/(\rho^1)_1)x^{C_1}$ for $2 \leq i \leq k$. It is not difficult to see from the structure of B that these are the only type 1 invariants.

These invariants may be rewritten $x^{C_i}/x^{C_1} = (\rho^1)_i/(\rho^1)_1$ to resemble the Haldane relationships that hold between substrates and products of a reaction at equilibrium (Cornish-Bowden, 1995). It follows from the construction of ρ^t by the Matrix-Tree Theorem that the right-hand side of this relationship is determined by the rate constants, as expected for a Haldane relationship (Cornish-Bowden, 1995). Horn and Jackson introduced the concept of a complex-balanced steady state, in part, to recover such generalised Haldane relationships for networks of reactions that might be in steady state but not at thermodynamic equilibrium (Horn and Jackson, 1972; Gunawardena, 2012).

Now suppose that $\delta_D = 1$. Then, ker $M = \langle \chi, \rho^1, \ldots, \rho^T \rangle$, where $\chi \in \mathbb{R}^m$ is any vector in ker M that is not in ker $\mathcal{L}(G)$. Choose B to have columns in the same order. Suppose that there are k complexes that are not in any terminal SCC and that indices are chosen so that these are $C_1, \ldots, C_k$. Then, $(\rho^t)_i = 0$ for $1 \leq i \leq k$ and $1 \leq t \leq T$, so that B is already in lower-triangular block form with $l = 1$. If $x \in (\mathbb{R}_{>0})^n$ is a positive steady state, then $\Psi(x) \in$ ker M and $\Psi(x)_i \neq 0$ for $1 \leq i \leq m$. It follows that $\chi_i \neq 0$ for $1 \leq i \leq k$. We may therefore choose $U = \{1\}$ and $V = \{2, \ldots, k\}$ and deduce from (6) that $x^{C_i} = (\chi_i/\chi_1)x^{C_1}$ for $2 \leq i \leq k$.

These type 1 complex-linear invariants lead to the theorem of Shinar and Feinberg on ACR (Shinar and Feinberg, 2010). Suppose that the structural deficiency of a network satisfies $\delta_S = 1$. Suppose further that C_1 and C_2 are two complexes that are not in any terminal SCC, whose stoichiometry differs only in species S_q. Since $\delta_D \leq \delta_S$ it must be that either $\delta_D = 0$ or $\delta_D = 1$. Suppose the former. The ρ^t then form a basis for ker M. Because C_1 is not in any non-terminal SCC, $\nu_1 = 0$ for any $\nu \in$ ker M. However, $\Psi(x) \in$ ker M and, since $x \in (\mathbb{R}_{>0})^n$, $\Psi(x)_1 \neq 0$. This contradiction

shows that $\delta_D = 1$. It then follows from the invariant above that $(x_q)^{C_2(S_q)-C_1(S_q)} = \chi_2/\chi_1$. Hence, the steady-state concentration of S_q depends only on the rate constants and not on the initial conditions or the total amounts and thereby exhibits ACR (Shinar and Feinberg, 2010).

2.5. Bifunctional enzymes

The previous calculations only exploited Theorem 1 when $l = 1$. We now consider examples with $l > 1$. Details of the calculations are given in the SI. The examples concern enzyme bifunctionality. Enzymes are known for being highly specific but some exhibit multiple activities. One form of this arises when a protein catalyses both a forward phosphorylation – covalent addition of phosphate, with ATP as the donor – and its reverse dephosphorylation—hydrolysis of the phosphate group. What advantage does such bifunctionality bring over having two separate enzymes?

We discuss one bacterial and one mammalian example. In *Escherichia coli*, osmolarity regulation is implemented in part by the EnvZ/OmpR two-component system (Fig. 2A); for references, see Shinar et al. (2007). Here, the sensor kinase, EnvZ, autophosphorylates on a histidine residue and catalyses the transfer of the phosphate group to the aspartate residue of the response regulator, OmpR, which then acts as an effector. Bifunctionality arises because EnvZ, when ATP is bound, also catalyses hydrolysis of phosphorylated OmpR-P.

It was suggested early on that the unusual design of the EnvZ/OmpR system might keep the absolute concentration of OmpR-P stable (Russo and Silhavy, 1993). This was later supported by experimental and theoretical analyses (Batchelor and Goulian, 2003), and the theoretical analysis was extended to other bifunctional two-component systems (Shinar et al., 2007). These *ad hoc*

calculations were clarified when a core network for EnvZ/OmpR was found to have $\delta_S = 1$ and the Shinar–Feinberg Theorem could be applied to confirm ACR for OmpR-P (Shinar and Feinberg, 2010). Attempts were made to broaden the analysis by extending the core network to include additional reactions thought to be present. For instance, EnvZ bound to ADP may also dephosphorylate OmpR-P. Adding these reactions to the core gives a network (Fig. 2B) with $\delta_S = 2$, so that Shinar–Feinberg can no longer be applied. However, it was shown by direct calculation in Shinar et al. (2007, Supplementary Information) that this network also satisfies ACR for OmpR-P.

Here, we use complex-linear invariants to confirm ACR and to find a formula for the absolute concentration value of OmpR-P in terms of the rate constants. The labelled, directed graph on the complexes has 13 nodes and 15 edges (Fig. 2C). Each connected component has only a single terminal SCC and $\delta_D = \delta_S = 2$. We can apply Theorem 1 to systematically find two new invariants.

Corollary 1. *If the complexes in the reaction network in Fig. 2B are ordered as shown in Fig. 2D, then the space of type 1 complex-linear invariants on the complexes C_1, C_3, C_8, C_{11} has dimension 2 and the following are independent invariants,*

$$\left(\frac{k_1 k_3}{k_2}\right) x^{C_1} - (k_4 + k_5) x^{C_3} = 0,$$

$$k_5 x^{C_3} - \left(\frac{k_{12} k_{10}}{k_{11} + k_{12}}\right) x^{C_8} - \left(\frac{k_{15} k_{13}}{k_{14} + k_{15}}\right) x^{C_{11}} = 0.$$

Using the expressions for the complexes in Fig. 2D, it can be seen that

$$x^{C_8} = x^{C_3} [OmpR-P], \quad x^{C_{11}} = x^{C_1} [OmpR-P].$$

Provided that $[EnvZ - ATP] = x^{C_3} \neq 0$, the invariants can be combined and simplified to yield the following expression:

$$[OmpR-P] = \frac{k_1 k_3 k_5 (k_{11} + k_{12})(k_{14} + k_{15})}{k_1 k_3 k_{10} k_{12}(k_{14} + k_{15}) + k_2 k_{13} k_{15}(k_4 + k_5)(k_{11} + k_{12})}.$$

$$(7)$$

This confirms that, as long as there is a positive steady state, the steady-state concentration of OmpR-P is not affected by changes in either the amount of OmpR or of EnvZ. The network exhibits ACR for OmpR-P, with the absolute value being given in terms of the rate constants by (7).

We now turn to our second example. 6-Phosphofructo-1-kinase (PFK-1) is one of the key regulatory enzymes in glycolysis, converting the small molecule fructose-6-phosphate to fructose-1,6-bisphosphate (Fig. 3A); for references, see Dasgupta et al. (submitted for publication). In mammalian cells, the bifunctional PFK-2/FBPase-2 has two domains. PFK-2 has the same substrate as PFK-1 but produces fructose-2,6-bisphosphate. This is a terminal metabolite that is not consumed by other metabolic processes. Instead, it acts as an allosteric effector, activating PFK-1 and inhibiting fructose-1,6-bisphosphatase, the reverse enzyme present in gluconeogenic cells, such as hepatocytes. The other domain, FBPase-2, catalyses the dephosphorylation of F2,6BP and produces F6P.

Biochemical studies lead to the reaction network in Fig. 3B. The kinase domain has an ordered, sequential mechanism and the kinase and phosphatase domains operate simultaneously; for more details, see Dasgupta et al. (submitted for publication). The corresponding labelled, directed graph on the complexes has 14 nodes and 19 edges (Fig. 4A). One of the connected components has two terminal SCCs, $\delta_D = 4$ and $\delta_S = 5$.
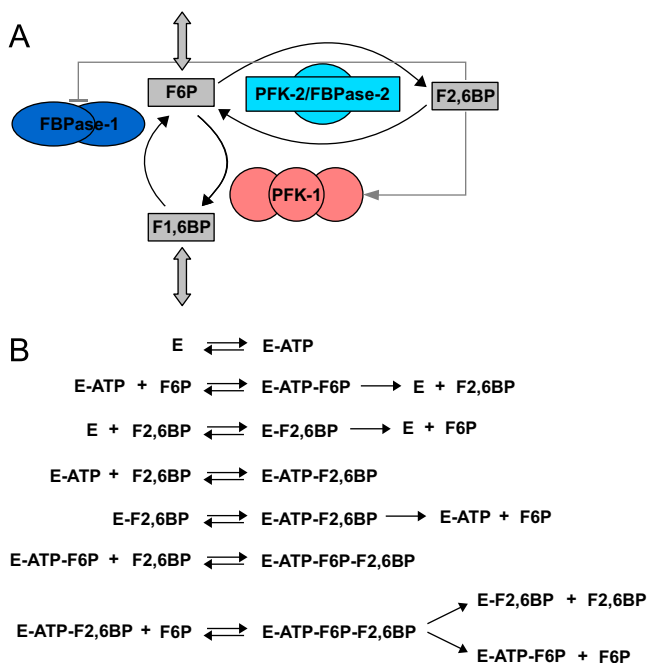
Fig. 3. The bifunctional enzyme 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFK-2/FBPase-2). (A) Schematic of the glycolysis/gluconeogenic pathway (broad gray arrows) at the step involving 6-phosphofructo-1-kinase (PFK-1, in red), that converts fructose-6-phosphate (F6P) into fructose-1,6-bisphosphate (F1,6BP), and fructose-1,6-bisphosphatase (FBPase-1, in dark blue), that catalyses the opposing reaction in gluconeogenic tissues. PFK-2/FBPase-2 (light blue) operates bifunctionally to produce and consume fructose-2,6-bisphosphate (F2,6BP), which allosterically regulates PFK-1 and FBPase-1, as shown. (B) The corresponding reaction network. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

Corollary 2. *If the complexes in the reaction network in Fig. 3B are ordered as shown in Fig. 4B, then the space of type 1 complex-linear invariants on the complexes $C_1, C_2, C_4, C_6, C_8, C_{11}$ has dimension 2 and the following are the independent invariants*

$$k_1 x^{C_1} - k_2 x^{C_2} + (k_{10} - k_8) x^{C_6} - (k_9 + k_{11}) x^{C_8} - k_{19} x^{C_{11}} = 0,$$

$$k_5 x^{C_4} - k_8 x^{C_6} - k_{11} x^{C_8} + (k_{18} - k_{19}) x^{C_{11}} = 0.$$

The second invariant in Corollary 2 was originally discovered by *ad hoc* algebraic calculation. It is used in Dasgupta et al. (submitted for publication) to show that, if the kinase dominates the phosphatase, in the sense that $k_{18} > k_{19}$, then the steady state concentration of F6P is held below a level that depends only on the rate constants and not on the amounts of the enzymes or the substrate. Conversely, if the phosphatase dominates the kinase, so that $k_{18} < k_{19}$, then the steady state concentration of F2,6BP is similarly constrained below a level that depends only on the rate constants and not on the amounts. Interestingly, regulation of PFK-2/FBPase-2 by phosphorylation, under the influence of the insulin and glucagon, causes the kinase and phosphatase activities to be shifted between the regimes $k_{18} > k_{19}$ and $k_{18} < k_{19}$. The implications of this for control of glycolysis are discussed in Dasgupta et al. (submitted for publication).

2.6. A systematic procedure

The two examples discussed above had already been analysed by other methods, so we had an idea of which invariants to expect and which subset of complexes to consider. For a new network,
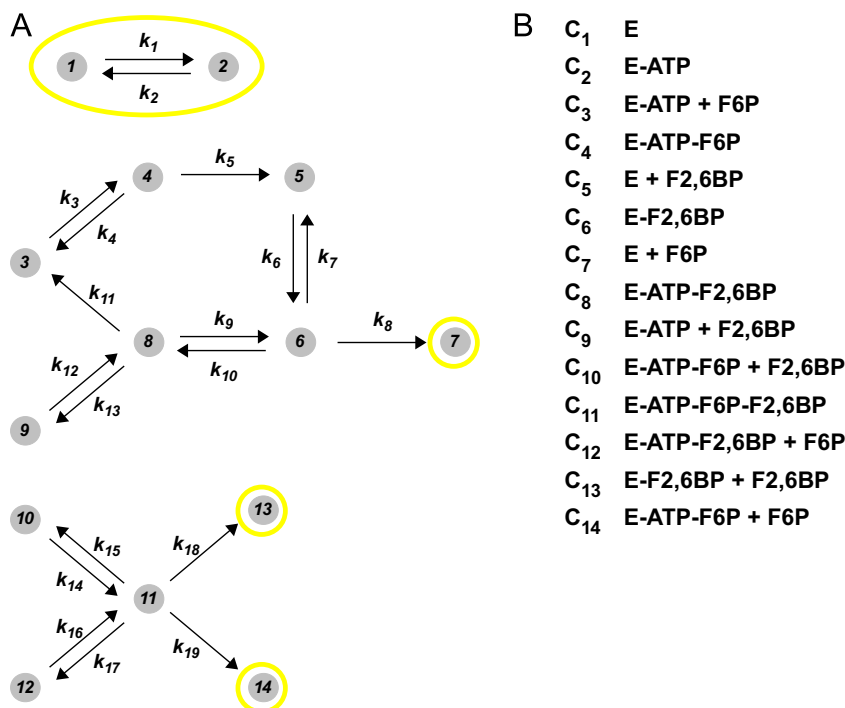
Fig. 4. PFK-2/FBPase-2, as in Fig. 3, in terms of complexes. (A) Labelled, directed graph on the complexes, with the terminal SCCs outlined in yellow. The last connected component has two terminal SCCs. (B) Numbering scheme for the complexes. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

such information may not be available, so how can non-trivial type 1 complex-linear invariants (simply, "invariants") be found? The automatic procedure outlined in Section 2.2 can be used in principle but this becomes computationally infeasible when there are many complexes. We have found the following systematic procedure to be helpful on several examples.

First determine the matrix M and from it the dual matrix B using Section 2.3 and Fig. 2. The biological context and the question being asked typically suggest one or more species of interest. For the initial subset of complexes, Z, choose all those complexes in which the species of interest have positive stoichiometry. Check if there are any invariants on Z using Proposition 1. If not, then consider any additional complexes that have at least one species in common with the complexes in Z. Add each of these complexes to Z in turn, starting with those that introduce the fewest new species and allowing the number of new species to increase as slowly as possible. With each addition, test for invariants as before. If this fails, consider adding the new complexes in groups, trying, as before, to minimise the number of new species that are introduced.

To demonstrate this procedure, we use a modification of the EnvZ/OmpR network in Fig. 3B. We add a single new reaction

$$OmpR-P \longrightarrow OmpR$$
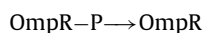
for spontaneous (non-catalysed) dephosphorylation of OmpR-P. The phospho-aspartate bond in response regulators is labile and may be spontaneously hydrolysed, so this new reaction is biochemically plausible. The labelled, directed graph for the modified network in Fig. 5A has 15 nodes and 16 edges. Each connected component still has only a single terminal SCC, like the graph in Fig. 2C, but now $\delta_D = \delta_S = 3$. The matrices M and B are provided in the SI, along with other details of the calculation.

The biological context suggests that the active state of the response regulator, OmpR-P, is of most interest. Following the procedure above and using the table in Fig. 5B leads to

$Z = \{C_7, C_8, C_{11}, C_{14}\}$ as an initial subset of complexes. It can be readily checked by inspection of B that the corresponding rows yield a submatrix of full rank 4, so that Proposition 1 tell us that there are no non-trivial invariants on Z. Among the remaining complexes, C_1, C_2 and C_3 each involve only species that are already present among the complexes in Z. Adding each to Z in turn, it can be checked that each of the subsets $Z \cup \{C_1\}$, $Z \cup \{C_2\}$ and $Z \cup \{C_3\}$ have a space of invariants of dimension 1, which respectively yield the following non-trivial invariants:

$$k_{16}x^{C_{14}} - \left(\frac{k_1 k_3 k_5}{k_2(k_4+k_5)}\right)x^{C_1} + \left(\frac{k_{10}k_{12}}{k_{11}+k_{12}}\right)x^{C_8} + \left(\frac{k_{13}k_{15}}{k_{14}+k_{15}}\right)x^{C_{11}} = 0,$$

$$k_{16}x^{C_{14}} - \left(\frac{k_3 k_5}{k_4+k_5}\right)x^{C_2} + \left(\frac{k_{10}k_{12}}{k_{11}+k_{12}}\right)x^{C_8} + \left(\frac{k_{13}k_{15}}{k_{14}+k_{15}}\right)x^{C_{11}} = 0,$$

$$k_{16}x^{C_{14}} - k_5 x^{C_3} + \left(\frac{k_{10}k_{12}}{k_{11}+k_{12}}\right)x^{C_8} + \left(\frac{k_{13}k_{15}}{k_{14}+k_{15}}\right)x^{C_{11}} = 0. \tag{8}$$

These all have a similar form, due to the common subset Z. The absence of C_7 in these invariants could have been inferred directly from the pattern of entries in B (SI).

A non-trivial invariant does not necessarily provide helpful biological insights. This depends crucially on the context and the question being studied. For instance, assuming that we have a positive steady-state, the second invariant in (8) may be rewritten in terms of steady-state species concentrations as

$$[OmpR-P] = \frac{\left(\frac{k_3 k_5}{k_4+k_5}\right)[EnvZ]}{k_{16} + \left(\frac{k_{10}k_{12}}{k_{11}+k_{12}}\right)[EnvZ-ATP] + \left(\frac{k_{13}k_{15}}{k_{14}+k_{15}}\right)[EnvZ-ADP]}. \tag{9}$$

Eq. (9) establishes a steady-state relationship between the activated response regulator, OmpR-P, and the sensor kinase, EnvZ, under nucleotide loading. This may be useful depending on the available experimental data. It does suggest that OmpR-P no longer exhibits ACR, as it did for the network in Fig. 2B, since its
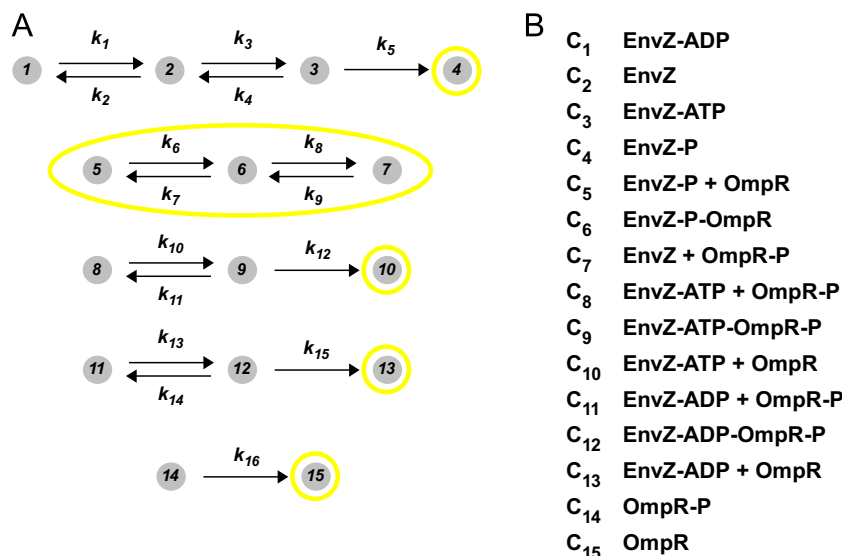
Fig. 5. Complexes for the modified EnvZ/OmpR network discussed in Section 2.6, in which a single reaction is added to the network in Fig. 2B. (A) Labelled, directed graph on the complexes, with the terminal SCCs outlined in yellow. Only the connected component at the bottom differs from the graph in Fig. 2C. (B) Numbering scheme, with two additional complexes, C_{14} and C_{15}, beyond those in Fig. 2D.

steady-state level appears to depend on the amount of EnvZ present. However, not much more can be said just from (9).

The situation is different for the first and third invariants in (8). They lead to a similar equation for [OmpR-P] as in (9) but with the numerators on the right hand side given by, respectively

$$\left(\frac{k_1 k_3 k_5}{k_2(k_4+k_5)}\right)[\text{EnvZ–ADP}] \quad \text{and} \quad k_5[\text{EnvZ–ATP}].$$

Because the same species now appears in both the numerator and the denominator, a simple comparison and cancellation yields the inequalities

$$[\text{OmpR–P}] < \begin{cases} \dfrac{k_1}{k_2}\dfrac{k_3 k_5}{(k_4+k_5)}\dfrac{(k_{14}+k_{15})}{k_{13}k_{15}}, \\[2mm] k_5\dfrac{(k_{11}+k_{12})}{k_{10}k_{12}}. \end{cases} \quad (10)$$

The strictness of the inequality comes from the assumption that the steady state is positive. We see that the activated response regulator has two upper bounds that are robust: they depend only on the rate constants and not on the initial conditions or the total amounts of either EnvZ or OmpR. An interesting aspect of (10) is the absence of parameters $k_6,\ldots,k_9$, that relate to the phosphorylation of OmpR. The other reactions contribute factors to the bounds that can be biochemically interpreted. Recall that the catalytic efficiency of an enzyme is the ratio of its catalytic rate to its Michaelis–Menten constant, k_{cat}/K_M (Cornish-Bowden, 1995). The factor $k_{13}k_{15}/(k_{14}+k_{15})$ in the first bound is the catalytic efficiency of OmpR-P dephosphorylation by EnvZ-ADP, while the factor $k_{10}k_{12}/(k_{11}+k_{12})$ in the second bound is the catalytic efficiency of OmpR-P dephosphorylation by EnvZ-ATP. The balance between these redundant dephosphorylation routes will influence which of the two bounds is the tighter and this balance is further modulated by the efficiency of nucleotide binding to EnvZ. The first bound is modulated by the factor $k_3 k_5/(k_4+k_5)$, which may be treated as an effective catalytic efficiency for EnvZ phosphorylation by ATP (it has units of s^{-1} rather than $\text{M}^{-1}\,\text{s}^{-1}$), and the factor k_1/k_2, which is the equilibrium constant for ADP binding. The second bound is only modulated by k_5, which is the catalytic rate for EnvZ phosphorylation.

The existence of these robust bounds also strongly suggests that OmpR-P does not satisfy ACR. Indeed, it can be shown by

algebraic calculation that [OmpR-P] can take different values in different steady states (SI). In contrast to the second invariant in (8), the first and third invariants yield interesting and unexpected biological insights. Further non-trivial invariants can be sought using the procedure above and we leave this for the reader.

We note two further points of interest. First, the addition of a single new reaction to a network can markedly change its behaviour from exhibiting ACR to only having robust bounds. This is a feature of biochemical networks under mass-action kinetics. It raises difficult problems of interpretation because there is always the possibility that the actual cellular network may include reactions that have been missed in a model. Very little work has been done on this difficult question. Invariants may provide a way to study it: perhaps certain invariants can be shown to remain "invariant" when a network is enlarged in a particular way. Second, a theme is emerging from the examples considered here. Despite the differences in network structures between Figs. 2, 3 and 5, the bifunctionality in each case serves to limit the steady state concentration of a substrate form, either absolutely, as in Fig. 2, or relative to some robust upper bound, as in Figs. 3 and 5. We speculate that this may be a design principle of those bifunctional enzymes that catalyse forward and reverse modifications. There are other forms of bifunctionality, such as enzymes that catalyse successive steps in a metabolic pathway, and preliminary studies suggest that these behave very differently. If modification bifunctionality did evolve to implement concentration control, markedly different network structures seem to have converged upon it.

3. Discussion

The nonlinearity of molecular networks makes it impossible to solve their dynamical behaviour in closed form. Their analysis has therefore relied on numerical integration and simulation, for which the biochemical details and the numerical values of all parameters must be specified in advance. This has made it difficult, if not impossible, to "see the wood for the trees" and to discern general principles within the overwhelming molecular complexity of cellular processes. Invariants are part of a new approach in which network behaviour at steady state can be

analysed with the parameters treated symbolically. There are now several examples, drawn from different biological contexts, in which the invariants summarise the essential steady-state properties of the network. The key to wide exploitation of this method is that it should be readily applicable to realistic networks. In principle, Gröbner bases allow any invariant to be calculated but this is computationally infeasible in practice. Complex-linear invariants form a limited subset of all invariants but, as shown here, they have biological significance and can be efficiently calculated for realistic networks. Our work clarifies previous results and provides a new tool for symbolic, steady-state analysis of molecular networks.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at http://dx.doi.org.10.1016/j.jtbi.2012.07.004.

References

Batchelor, E., Goulian, M., 2003. Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. Proc. Natl. Acad. Sci. U.S.A. 100, 691–696.

Cornish-Bowden, A., 1995. Fundamentals of Enzyme Kinetics, 2nd edition Portland Press, London, UK.

Cox, D., Little, J., O'Shea, D., 1997. Ideals, Varieties and Algorithms, 2nd edition Springer.

Craciun, G., Dickenstein, A., Shiu, A., Sturmfels, B., 2009. Toric dynamical systems. J. Symb. Comput. 44, 1551–1565.

Dasgupta, T., Croll, D.H., Owen, J. A., Vander Heiden, M.H., Locasale, J.W., Alon, U., Cantley, L.C., Gunawardena, J. A fundamental trade off in covalent switching and its circumvention in glucose homeostasis. submitted for publication.

Feinberg, M., 1979. Lectures on Chemical Reaction Networks. Lecture Notes. Mathematics Research Center, University of Wisconsin.

Feinberg, M., Horn, F., 1977. Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspace. Arch. Rational Mech. Anal. 66, 83–97.

Gunawardena, J., 2003. Chemical Reaction Network Theory for *In-Silico* Biologists. Lecture Notes. Harvard University. vcp.med.harvard.edu/papers/crnt.pdf.

Gunawardena, J., 2010. Models in systems biology: the parameter problem and the meanings of robustness. In: Lodhi, H., Muggleton, S. (Eds.), Elements of Computational Systems Biology. Wiley Book Series on Bioinformatics. John Wiley and Sons, Inc.

Gunawardena, J., 2012. A linear framework for time-scale separation in nonlinear biochemical systems. PLoS ONE 7, e36321.

Horn, F., Jackson, R., 1972. General mass action kinetics. Arch. Rational Mech. Anal. 47, 81–116.

Kramer, B.P., Viretta, A.U., Daoud-El-Baba, M., Aubel, D., Weber, W., Fussenegger, M., 2004. An engineered epigenetic transgene switch in mammalian cells. Science 22, 867–870.

Laurent, M., Kellershohn, N., 1999. Multistability: a major means of differentiation and evolution in biological systems. Trends Biochem. Sci. 24, 418–422.

Manrai, A., Gunawardena, J., 2008. The geometry of multisite phosphorylation. Biophys. J. 95, 5533–5543.

Pérez Millán, M., Dickenstein, A., Shiu, A., Conradi, C., 2012. Chemical reaction systems with toric steady states. Bull. Math. Biol. 74, 1027–1065.

Russo, F.D., Silhavy, T.J., 1993. The essential tension: opposed reactions in bacterial two-component regulatory systems. Trends Microbiol. 1, 306–310.

Shinar, G., Feinberg, M., 2010. Structural sources of robustness in biochemical networks. Science 327, 1389–1391.

Shinar, G., Milo, R., Martínez, M.R., Alon, U., 2007. Input–output robustness in simple bacterial signaling systems. Proc. Natl. Acad. Sci. U.S.A. 104, 19931–19935.

Shinar, G., Rabinowitz, J.D., Alon, U., 2009. Robustness in glyoxylate bypass regulation. PLoS Comp. Biol. 5, e1000297.

Thomson, M., Gunawardena, J., 2009a. The rational parameterisation theorem for multisite post-translational modification systems. J. Theor. Biol. 261, 626–636.

Thomson, M., Gunawardena, J., 2009b. Unlimited multistability in multisite phosphorylation systems. Nature 460, 274–277.

Xu, Y., Gunawardena, J. Realistic enzymology for post-translational modification: zero-order ultrasensitivity revisited. J. Theor. Biol., http://dx.doi.org/1016/j.jtbi.2012.07.012, 2012, in press.