SIAM J. Applied Dynamical Systems
Vol. 17, No. 2, pp. 1650–1682

The Structure of MESSI Biological Systems*

Mercedes Pérez Millán† and Alicia Dickenstein†

Abstract. We introduce a general framework for biological systems, called MESSI systems, that describe Modifications of type Enzyme-Substrate or Swap with Intermediates, and we prove general results based on the network structure. Many posttranslational modification networks are MESSI systems. Examples are the motifs in [E. Feliu and C. Wiuf, *J. R. Soc. Interface*, 9 (2012), pp. 1224–1232], sequential distributive and processive multisite phosphorylation networks, most of the examples in [D. Angeli, P. De Leenher, and E. Sontag, *Math. Biosci.*, 210 (2007), pp. 598–618], phosphorylation cascades, two component systems as in [V. B. Kothamachu et al., *J. R. Soc. Interface*, 12 (2015), 20150234], the bacterial EnvZ/OmpR network in [G. Shinar and M. Feinberg, *Science*, 327 (2010), pp. 1389–1391], and all linear networks. We show that, under mass-action kinetics, MESSI systems are conservative. We simplify the study of steady states of these systems by explicit elimination of intermediate complexes, and we give conditions to ensure an explicit rational parametrization of the variety of steady states (inspired by [E. Feliu and C. Wiuf, *J. R. Soc. Interface*, 10 (2013), 20130484, *J. Math. Biol.*, 66 (2013), pp. 281–310; M. Thomson and J. Gunawardena, *J. Theoret. Biol.*, 261 (2009), pp. 626–636]). We define an important subclass of MESSI systems with toric steady states [M. Pérez Millán et al., *Bull. Math. Biol.*, 74 (2012), pp. 1027–1065], and we give for MESSI systems with toric steady states an easy algorithm to determine the capacity for multistationarity. In this case, the algorithm provides rate constants for which multistationarity takes place, based on the theory of oriented matroids.

Key words. biological networks, steady states, MESSI system, multistationarity

AMS subject classifications. 92C42, 13P25

DOI. 10.1137/17M1113722

1. Introduction.

Many processes within cells involve some kind of posttranslational modification of proteins. We introduce a general framework for biological systems that describe Modifications of type Enzyme-Substrate or Swap with Intermediates, which we call MESSI systems, and which allows us to prove general results on their dynamics from the structure of the network, under mass-action kinetics. This subclass of mechanisms has attracted considerable theoretical attention due to its abundance in nature and the special characteristics in the topologies of the networks.

The basic idea in the definition of MESSI systems (see Definitions 2.3 and 2.10) is that the mathematical modeling reflects the different chemical behaviors. The chemical species can be grouped into different subsets according to the way they participate in the reactions,

*Received by the editors January 26, 2017; accepted for publication (in revised form) by M. Chaves March 16, 2018; published electronically June 7, 2018.
http://www.siam.org/journals/siads/17-2/M111372.html
Funding: This work was partially supported by UBACYT 20020100100242, CONICET PIP 11220150100473 and 11220150100483, and ANPCyT PICT 2013-1110, Argentina.
†Dto. de Matemática, FCEN, Universidad de Buenos Aires, Ciudad Universitaria, Pab. I, C1428EGA Buenos Aires, Argentina, and IMAS (UBA-CONICET), Ciudad Universitaria, Pab. I, C1428EGA Buenos Aires, Argentina (mpmillan@dm.uba.ar, alidick@dm.uba.ar).

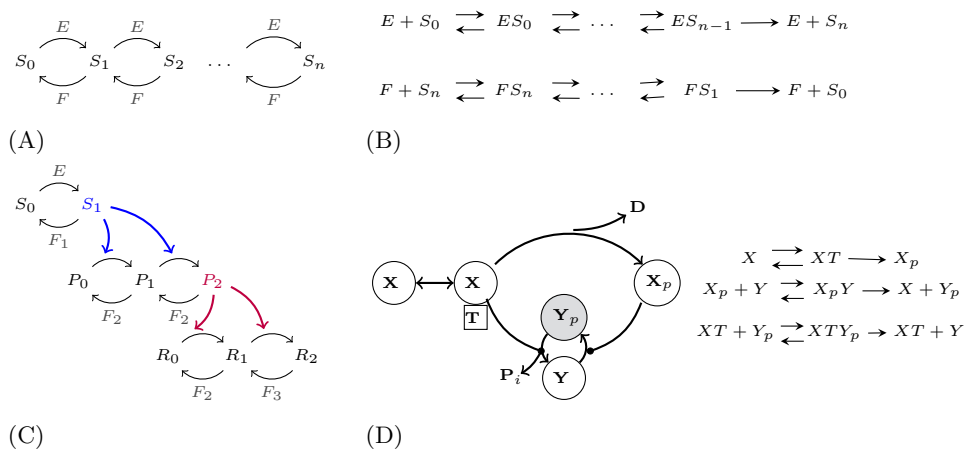


Figure 1. *Examples of MESSI systems: Sequential n-site phosphorylation/dephosphorylation* (A) *distributive case* [34, 49]*;* (B) *processive case* [5, 29]*;* (C) *phosphorylation cascade;* (D) *schematic diagram of an EnvZ-OmpR bacterial model* [42].

very much akin to the intuitive partition of the species according to their function. We show that MESSI systems are conservative (and thus all trajectories are defined for any positive time), and we study the important questions of persistence and multistationarity. Informally, persistence means that no species which is present can tend to be eliminated in the course of the reaction [1]. Multistationarity (see Definition 2.2) is also a crucial property, since its occurrence can be thought of as a mechanism for switching between different response states in cell signaling systems and enables multiple outcomes for cellular decision making, with the same stoichiometric content.

Examples of MESSI systems of major biological importance are phosporylation cascades, such as the mitogen-activated protein kinases (MAPKs) cascades [3, 22, 24]. MAPKs are serine/threonine kinases that play an essential role in signal transduction by modulating gene transcription in the nucleus in response to changes in the cellular environment and participate in a number of disease states including chronic inflammation and cancer [6, 26, 33, 39, 51] as they control key cellular functions [20, 33, 40, 46, 50]. Also, the *multisite phosphorylation system* is a MESSI system. This network describes the phosphorylation of a protein in multiple sites by a kinase/phosphatase pair in a sequential and distributive mechanism [7, 18, 19, 22, 27, 41]. In prokaryotic cells, an example of a MESSI system can be found in [42], representing the *Escherichia coli* EnvZ-OmpR system which consists of the sensor kinase EnvZ, and the response-regulator OmpR (see also [21, 23, 35, 43, 52]). This signaling system is a prototypical two-component signaling system [35, 43]. All linear systems are also MESSI.

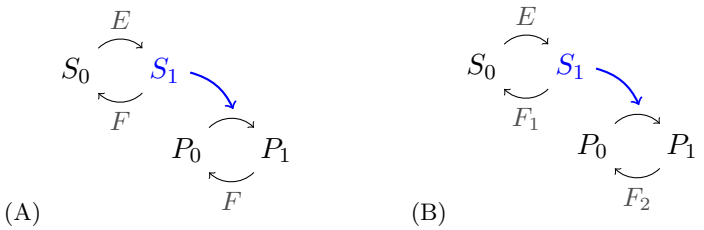


Figure 2. *Same and different phosphatases in two different MESSI cascades.*

We depict in Figures 1 and 2 some examples of important biochemical networks which are MESSI networks.[1] Figure 1(A) features the n-site phosphorylation-dephosphorylation of a protein by a kinase-phosphatase pair in a sequential and distributive mechanism. The total of n phosphate groups is allowed to be added to the unphosphorylated substrate S_0 by an enzyme E. The substrate S_i is the phosphoform obtained from S_0 by attaching i phosphate groups to it. Each phosphoform can accept (via an enzymatic reaction involving E) or lose (via a reaction involving the phosphatase F) at most one phosphate; this means that the mechanism is "distributive." In addition, the phosphorylation is said to be "sequential" because multiple phosphate groups must be added in a specific order and removed in a specific order as well. The sequential and processive phosphorylation/dephosphorylation of a substrate at n sites [29, 5] is depicted in Figure 1(B). The substrate undergoes $n \geq 1$ phosphorylations after binding to the kinase and forming the enzyme-substrate complex; only the fully phosphorylated substrate is released, and hence only two phosphoforms have to be considered: the unphosphorylated substrate S_0 and the fully phosphorylated substrate S_n. Processive dephosphorylation proceeds similarly. All the motifs in [13] are MESSI networks, as are the phosphorylation cascades shown in Figure 2. The cascade in Figure 1(C) features the sequential activation of a specific MAPK kinase kinase (MAPKKK, denoted S) and a MAPK kinase (MAPKK, denoted P), which in turn phosphorylates and activates the downstream MAPK (denoted R). The activated forms are S_1, P_2, and R_2, respectively. Figure 2 features two cascade motifs with two layers, which are a combination of two one-site modification cycles with either a specific or the same phosphatase acting in each layer. It is already known [13] that the cascade in (A) exhibits multistationarity while the cascade in (B) is monostationary. We will recover these results under the framework of MESSI systems (they will both prove to be s-toric MESSI systems; see Definition 4.3). We will moreover consider the cascade in Figure 2(A) as one of our running examples in this article, and sometimes we will also include a drug D acting by a sequestration mechanism such as $P_1 + D \rightleftarrows P_1 D$. Figure 1(D) depicts a schematic diagram of an EnvZ-OmpR bacterial model [42], which is a MESSI network. The sensor EnvZ (X) phosphorylates itself by binding and breaking down ATP (T). The phosphorylated form X_p catalyzes the transfer of a phosphoryl group to the response-regulator OmpR

[1] As usual, in the figures we summarize with the scheme $S_0 \overset{E}{\frown} S_1$ a sequence of reactions with intermediates such as $S_0 + E \underset{\kappa_2}{\overset{\kappa_1}{\rightleftarrows}} ES_0 \overset{\kappa_3}{\to} S_1 + E$.

(Y). X, together with ATP, dephosphorylates Y_p, a transcription factor that regulates the expression of various protein pores.

Our work continues the ideas in chemical reaction network theory (CNRT), which connects qualitative properties of ordinary differential equations corresponding to a reaction network to the network structure. CNRT has been developed over the last 40 years, initially through the work of Horn and Jackson and subsequently by Feinberg and his students and collaborators (for example, see [9, 10]) and Vol'pert [48]. Biochemical reaction networks, that is, chemical reaction networks in biochemistry, are the principal current application of these developments. In particular, our work is inspired by previous articles by Thomson and Gunawardena [45], who set the posttranslational modification (PTM) framework; Mirzaev and Gunawardena [30], who detailed the Laplacian dynamics; Feliu and Wiuf [14, 15], who clarified the elimination of intermediate complexes; and Müller et al. [32], who collected and clarified the role of signs in the determination of multistationarity. Also related to our work are the papers by Gnacadja on constructive chemical reaction networks [16, 17], who gave an alternative approach to the PTM setting. The MESSI structure we propose simplifies and unifies most of these approaches.

The precise conditions are given in Definitions 2.3 and 2.10. In particular, complexes in a MESSI network are mono or bimolecular. As remarked in [45], one main assumption for this modeling is that donor molecules that provide modifiers are kept at constant concentration on the time scaling of the reactions we are modeling, and their effects can be absorbed into the rate constants. The main difference between our approach and theirs is that they do not allow a species to act as a substrate in one reaction and then as an enzyme in another (neither does [30]), which in particular excludes all enzymatic cascades. This is considered in [16, 17]. However, none of these previous settings allow swaps and monomolecular reactions between core species that our framework incorporates. Regarding [14, 15, 32], we pay special attention to networks with toric steady states [34].

Theorem 3.2 explicitly describes conservation relations that imply that any MESSI system is conservative. Theorem 3.15 gives conditions that ensure that a MESSI system is persistent. We give necessary conditions for the existence of a rational parametrization of the variety of positive steady states in Theorem 4.1, which is the generalization of the main theorem in [45] to our setting. Proposition 4.7 expresses the role of intermediates in the steady states of the system. Theorem 4.8 shows a frequent class of MESSI systems with special steady states, cut out by binomial equations and termed as toric steady states [34], that allow for an easier determination of multistationarity.

We give for MESSI systems with toric steady states an algorithm to determine the capacity for multistationarity based on Theorems 5.4 and 5.8. If this is the case, the algorithm provides rate constants for which multistationarity takes place, based on the theory of oriented matroids [2]. This is a specialized procedure, easy to tune to produce different choices of rate constants, besides the general algorithms for injectivity implemented, for instance, by Feinberg and his group in the Chemical Reaction Network Toolbox [11]. Links to other algorithms can be found at https://reaction-networks.net/wiki/Mathematics_of_Reaction_Networks. The proofs of our statements are concentrated in Appendix A.

2. MESSI systems.

In this section we review the notion of a chemical reaction network in order to introduce the definition of MESSI networks and MESSI systems (when these networks

are endowed with mass-action kinetics). The conditions in the definition might seem to be very restrictive (mathematically), but indeed we show many examples of *popular* networks in systems biology that lie in this framework.

Chemical reaction systems. We briefly recall the basic setup of chemical reaction networks and how they give rise to autonomous dynamical systems under mass-action kinetics (see Example 2.1). Given a set of s chemical species, a *chemical reaction network* on this set of species is a finite directed graph whose vertices are indicated by complexes and whose edges are labeled by parameters (reaction rate constants). The labeled digraph is denoted $G = (\mathcal{V}, \mathcal{E}, \boldsymbol{\kappa})$, with vertex set $\mathcal{V}$, edge set $\mathcal{E}$, and edge labels $\boldsymbol{\kappa} \in \mathbb{R}^{\#\mathcal{E}}_{>0}$. If $(y, y') \in \mathcal{E}$, we denote $y \to y'$. Complexes determine vectors in $\mathbb{Z}^s_{\geq 0}$ according to the stoichiometry of the species they consist of. We identify a complex with its corresponding vector and also with the formal linear combination of species specified by its coordinates.

Example 2.1 (basic example of an enzymatic network). We present a basic example that illustrates how a chemical reaction network gives rise to a dynamical system. This example represents a classical mechanism of enzymatic reactions, usually known as the futile cycle [22, 24, 49]:

$$S_0 + E \underset{\kappa_2}{\overset{\kappa_1}{\rightleftarrows}} U_1 \overset{\kappa_3}{\to} S_1 + E, \qquad S_1 + F \underset{\kappa_5}{\overset{\kappa_4}{\rightleftarrows}} U_2 \overset{\kappa_6}{\to} S_0 + F, \tag{1}$$

where U_1 and U_2 are *intermediate species*, S_0 and S_1 are *substrates*, and E and F are *enzymes*. The source and the product of each reaction are called *complexes*. The concentrations of the six species change in time as the reactions occur. We order the $s = 6$ species as follows: U_1, U_2, S_0, S_1, E, F, and we denote the concentrations by $[U_1] = u_1$, $[U_2] = u_2$, $[S_0] = x_1$, $[S_1] = x_2$, $[E] = x_3$, $[F] = x_4$. The first three complexes in the network (1) give rise to the vectors $(0,0,1,0,1,0)$, $(1,0,0,0,0,0)$, and $(0,0,0,1,1,0)$. Under the assumption of *mass-action kinetics*, we obtain then the following polynomial dynamical system:

$$\begin{aligned}
\tfrac{du_1}{dt} &= \kappa_1 x_1.x_3 - (\kappa_2 + \kappa_3)u_1, & \tfrac{du_2}{dt} &= \kappa_4 x_2.x_4 - (\kappa_5 + \kappa_6)u_2,\\
\tfrac{dx_1}{dt} &= -\kappa_1 x_1.x_3 + \kappa_2 u_1 + \kappa_6 u_2, & \tfrac{dx_2}{dt} &= -\kappa_4 x_2.x_4 + \kappa_5 u_2 + \kappa_3 u_1,\\
\tfrac{dx_3}{dt} &= -\kappa_1 x_1.x_3 + (\kappa_2 + \kappa_3)u_1, & \tfrac{dx_4}{dt} &= -\kappa_4 x_2.x_4 + (\kappa_4 + \kappa_5)u_2.
\end{aligned}$$

The unknowns $x_1, x_2, \ldots, x_s$ represent the concentrations of the species in the network, and we regard them as functions of time t. Under mass-action kinetics, the chemical reaction network G defines the following chemical reaction dynamical system:

$$\dot{x} = \left(\frac{dx_1}{dt}, \frac{dx_2}{dt}, \ldots, \frac{dx_s}{dt}\right) = \sum_{y \to y'} \kappa_{yy'}\, x^y\, (y' - y), \tag{2}$$

where $x = (x_1, \ldots, x_s)$ and $x^y = x_1^{y_1} \cdots x_s^{y_s}$. The right-hand side of each differential equation dx_ℓ/dt is a polynomial $f_\ell(x, \boldsymbol{\kappa})$, in the variables $x_1, \ldots, x_s$ with real coefficients $\boldsymbol{\kappa}$. The associated *steady state variety* V_f is defined as the common nonnegative zeros of the polynomials

f_ℓ, that is,

$$V_f := \{x \in \mathbb{R}^s_{\geq 0} \,:\, f_\ell(x, \boldsymbol{\kappa}) = 0, \quad \ell = 1, \ldots, s\}. \tag{3}$$

The linear subspace spanned by the reaction vectors $S = \{y' - y \,:\, y \to y'\}$ is called the *stoichiometric subspace*. Notice from (2) that the vector $\dot{x}(t)$ lies in S for all time t. In fact, a trajectory $x(t)$ beginning at a vector $x(0) = x^0 \in \mathbb{R}^s_{\geq 0}$ remains in the *stoichiometric compatibility class* $(x^0 + S) \cap \mathbb{R}^s_{\geq 0}$ for all positive time. The equations of $x^0 + S$ give rise to linear *conservation relations* of the system.

Definition 2.2. *We say that the system* has the capacity for multistationarity *if there exists a choice of rate constants* $\boldsymbol{\kappa}$ *such that there are two or more steady states in one stoichiometric compatibility class. On the other hand, if for any choice of rate constants there is at most one steady state in each stoichiometric compatibility class, the system is said to be* monostationary.

It may happen that the vectors $\dot{x}(t)$ lie in a smaller subspace $S' \subseteq S$, called the *kinetic subspace* [12]. In this case, the trajectories live in $(x^0 + S') \cap \mathbb{R}^s_{\geq 0}$ for some initial state $x^0 \in \mathbb{R}^s_{\geq 0}$, and the concepts of mono- and multistationarity might be defined with respect to this smaller affine subspace. In this article, we focus on the classical Definition 2.2.

Definition of MESSI systems. A MESSI network is a particular type of chemical reaction network, which includes all monomolecular (linear) ones. As we mentioned in the introduction, the main ingredient in the definition is the existence of a partition of the set of species, that is, a decomposition into disjoint subsets, with the following properties.

Definition 2.3. *A chemical reaction network is called a MESSI network if there is a* partition *of the set of species* $\mathscr{S}$

$$\mathscr{S} = \mathscr{S}^{(0)} \bigsqcup \mathscr{S}^{(1)} \bigsqcup \mathscr{S}^{(2)} \bigsqcup \cdots \bigsqcup \mathscr{S}^{(m)}, \tag{4}$$

where $m \geq 1$ *and* $\bigsqcup$ *denotes disjoint union, such that the complexes and reactions satisfy the conditions below.*

We call the cardinalities $\#\mathscr{S}^{(0)} = p$, $\#\mathscr{S}^{(\alpha)} = n_\alpha$ *for any* $\alpha > 0$ *and* $\sum_{\alpha>0} n_\alpha = n$. *We allow* p *to be* 0, *but we assume that all* n_α *are positive. Species in* $\mathscr{S}^{(0)}$ *are called* intermediate, *and species in* $\mathscr{S}_1 := \mathscr{S} \setminus \mathscr{S}^{(0)}$ *are termed* core. *When convenient, we will distinguish intermediate and core species in the notation in the following way:* $\mathscr{S}^{(0)} = \{U_1, \ldots, U_p\}$, $\mathscr{S}_1 = \{X_1, \ldots, X_n\}$. *Thus, the vectors determined by the complexes* $(\lambda_1, \ldots, \lambda_p, \nu_1, \ldots, \nu_n)$ *live in* $\mathbb{Z}^{p+n}_{\geq 0}$ *and define the formal linear combination of species* $\sum_{i=1}^p \lambda_i U_i + \sum_{j=1}^n \nu_j X_j$.

Complexes are also partitioned into two disjoint sets, and the following conditions hold:

($\mathcal{N}_1$) Intermediate complexes *are complexes that consist of a unique intermediate species that only appears in that complex. The vector corresponding to the unimolecular complex* U_i *is denoted by* y_i.

($\mathcal{N}_2$) Core complexes [14] *are mono or bimolecular and consist of either one or two core species. If the core complex consists of only the species* X_i, *the corresponding vector will be denoted by* y_{i0}.

($\mathcal{N}_3$) *When a core complex consists of two species X_i, X_j, they* must *belong to* different *sets $\mathscr{S}^{(\alpha)}, \mathscr{S}^{(\beta)}$ with $\alpha \neq \beta, \alpha, \beta \geq 1$. We also denote the complex $X_i + X_j = X_j + X_i$ by $y_{ij} = y_{ji}$.*

We say that complex y reacts to complex y' via intermediates *if either $y \to y'$ or there exists a path of reactions from y to y'* only through intermediate complexes. *This is denoted by $y \to_\circ y'$. The intermediate complexes of a MESSI network satisfy, moreover, the following condition:*

($\mathcal{C}$) *For every intermediate complex y_k, there exist core complexes y_{ij} and $y_{\ell m}$ such that $y_{ij} \to_\circ y_k$ and $y_k \to_\circ y_{\ell m}$.*

Finally, reactions are constrained by the following rules:

($\mathcal{R}_1$) *If three species are related by $X_i + X_j \to_\circ X_k$ or $X_k \to_\circ X_i + X_j$, then X_k is an intermediate species.*

($\mathcal{R}_2$) *If two core species X_i, X_j are related by $X_i \to_\circ X_j$, then there exists $\alpha \geq 1$ such that both belong to $\mathscr{S}^{(\alpha)}$.*

($\mathcal{R}_3$) *If $X_i + X_j \to_\circ X_k + X_\ell$, then there exist $\alpha \neq \beta$ such that $X_i, X_k \in \mathscr{S}^{(\alpha)}$, $X_j, X_\ell \in \mathscr{S}^{(\beta)}$ or $X_i, X_\ell \in \mathscr{S}^{(\alpha)}$, $X_j, X_k \in \mathscr{S}^{(\beta)}$.*

We will say that the partition (4) defines a MESSI structure *on the network.*

Example 2.4. We present a toy example that shows which kinds of reactions are allowed and which are not. Consider the following digraph, where we assume Y_1 and Y_2 to be monomolecular complexes:

$$X_1 + X_2 \to Y_1 \rightleftarrows Y_2 \to Y_3.$$

Then, Y_1 and Y_2 must consist of an intermediate species by rule ($\mathcal{R}_1$). For Condition ($\mathcal{C}$) to hold, necessarily Y_3 must be a core complex since there are no arrows leaving from Y_3. Moreover, rule ($\mathcal{R}_1$) imposes that Y_3 is of the form $X_\ell + X_m$, and by rule ($\mathcal{R}_3$), if $X_1 \in \mathscr{S}^{(\alpha)}$ and $X_2 \in \mathscr{S}^{(\beta)}$, then $\alpha \neq \beta$ and either $X_\ell \in \mathscr{S}^{(\alpha)}, X_m \in \mathscr{S}^{(\beta)}$ or $X_m \in \mathscr{S}^{(\alpha)}, X_\ell \in \mathscr{S}^{(\beta)}$.

Notice that a MESSI network is defined once the partition of $\mathscr{S}$ is given and all conditions and rules in Definition 2.3 are verified. It is important to point out that even if in the chemical setting there are natural partitions of the set of species given by the different types of molecules, there can be many ways to define a partition which defines a MESSI structure. We can define a partition order in the set of all possible partitions of the species of a given biochemical network.

Definition 2.5. *Given two partitions $\mathscr{S} = \mathscr{S}^{(0)} \sqcup \mathscr{S}^{(1)} \sqcup \mathscr{S}^{(2)} \sqcup \cdots \bigsqcup \mathscr{S}^{(m)}$ and $\mathscr{S} = \mathscr{S}'^{(0)} \sqcup \mathscr{S}'^{(1)} \sqcup \mathscr{S}'^{(2)} \sqcup \cdots \bigsqcup \mathscr{S}'^{(m')}$, we say that the first partition refines the second one if and only if $\mathscr{S}^{(0)} \supseteq \mathscr{S}'^{(0)}$ and for any $\alpha \geq 1$, there exists $\alpha' \geq 1$ such that $\mathscr{S}^{(\alpha)} \subseteq \mathscr{S}'^{(\alpha')}$. With this partial order we have the notion of a minimal partition.*

Before presenting our two running examples, we define enzyme behavior and swaps.

Definition 2.6. *A species X_j that satisfies $X_i + X_j \to_\circ X_\ell + X_j$ for some X_i, X_ℓ is said to act as an* enzyme. *In this case, we call X_i the* substrate *and X_ℓ the* product. *A reaction via intermediates is called a* swap *if $X_i + X_j \to_\circ X_\ell + X_m$, and $i, j \notin \{\ell, m\}$ (so, neither X_i nor X_j acts as an enzyme in $X_i + X_j \to_\circ X_\ell + X_m$).*

Notice that if a species X_j in a MESSI network *only* acts as an enzyme, we can consider a singleton subset $\mathscr{S}^{(\alpha)} = \{X_j\}$.

Example 2.7 (first running example). Consider the network in Figure 2 (A), with digraph

$$S_0 + E \underset{\kappa_2}{\overset{\kappa_1}{\rightleftarrows}} ES_0 \overset{\kappa_3}{\to} S_1 + E, \qquad S_1 + F \underset{\kappa_5}{\overset{\kappa_4}{\rightleftarrows}} FS_1 \overset{\kappa_6}{\to} S_0 + F,$$

$$P_0 + S_1 \underset{\kappa_8}{\overset{\kappa_7}{\rightleftarrows}} S_1P_0 \overset{\kappa_9}{\to} P_1 + S_1, \qquad P_1 + F \underset{\kappa_{11}}{\overset{\kappa_{10}}{\rightleftarrows}} FP_1 \overset{\kappa_{12}}{\to} P_0 + F.$$

We can consider the partition $\mathscr{S}^{(0)} = \{ES_0, FS_1, S_1P_0, FP_1\}$ (intermediate species), and $\mathscr{S}^{(1)} = \{S_0, S_1\}$, $\mathscr{S}^{(2)} = \{P_0, P_1\}$, $\mathscr{S}^{(3)} = \{E\}$, $\mathscr{S}^{(4)} = \{F\}$ (partition of the core species). The intermediate complexes correspond to the intermediate species, and the remaining complexes are core complexes. This partition defines a MESSI structure in the network. In fact, there is another possible choice of partition which also gives a MESSI structure to the network, considering $\mathscr{S}^{(0)}$, $\mathscr{S}^{(1)}$, and $\mathscr{S}^{(2)}$ as before, but $\mathscr{S}^{(3)}$ and $\mathscr{S}^{(4)}$ are replaced by their union $\{E, F\}$. We can see in this example that species E and F only act as enzymes, while species S_1 acts as an enzyme in the second layer but in the first one it plays the role of a substrate of F and of a product of E.

Example 2.8 (second running example). An example of swap can be the seen in the transfer of a modifier molecule, such as a phosphate group in a two-component system, from one molecule to another. We consider as our second running example the EnvZ/OmpR system. The corresponding digraph G is featured in Figure 1(D). The only possible partition for this network to be a MESSI network is $\mathscr{S}^{(0)} = \{\mathrm{X}_p\mathrm{Y}, \mathrm{XTY}_p\}$, $\mathscr{S}^{(1)} = \{X, XT, X_p\}$, $\mathscr{S}^{(2)} = \{Y, Y_p\}$. The reaction via intermediates in the second connected component of the graph of reactions is a swap. On the other hand, XT acts as an enzyme in the last component of G.

In Example 2.7, there are two different partitions, but the first one is a refinement of the second one. However, there might be noncomparable partitions, as we show in the following example.

Example 2.9 (noncomparable partitions). Consider the following network:

$$X_1 + X_2 \to X_3 + X_4, \quad X_4 + X_5 \to X_6 + X_1.$$

Set $\mathscr{S}^{(0)} = \emptyset$, $\mathscr{S}^{(1)} = \{X_1, X_4\}$, and $\mathscr{S}^{(2)} = \{X_2, X_3, X_5, X_6\}$. We can refine $\mathscr{S}^{(2)}$ into $\mathscr{S}'^{(2)} = \{X_2, X_3\}$ and $\mathscr{S}'^{(3)} = \{X_5, X_6\}$. In both cases, we get the structure of a MESSI network. If we instead consider $\mathscr{S}''^{(0)} = \emptyset$, $\mathscr{S}''^{(1)} = \{X_1, X_3, X_5\}$, and $\mathscr{S}''^{(2)} = \{X_2, X_4, X_6\}$, there is no possible way of refining $\mathscr{S}''^{(2)}$ without violating $(\mathcal{R}_3)$. The second and third partitions are not comparable, and both are minimal in the poset of partitions of the species set which yield a MESSI structure on the given network.

The main focus of this work is the properties of MESSI networks endowed with kinetics. Throughout this text we will always assume mass-action kinetics.

Definition 2.10. *We call a* MESSI system *the mass-action kinetics dynamical system as in* (2) *associated with a MESSI network.*

3. Conservation relations and persistence in MESSI systems. We first describe the equations of the stoichiometric subspace of a MESSI system, which give linear conservation relations along the trajectories. We then focus on the steady states of MESSI systems. We give sufficient conditions for MESSI systems to be persistent.

Conservation relations. A chemical reaction system is said to be *conservative* if there exists a linear combination of the species in the network with all *positive* coefficients which is constant along each trajectory (i.e., for all time t). Clearly, for any trajectory starting at a positive point, this constant is a positive real number. In this case, all stoichiometric compatibility classes are compact. In this section we show that MESSI systems are conservative, by exhibiting natural conservation relations. This implies that all trajectories are bounded and defined for any positive time.

Notation 3.1. *We denote the concentration of the species with small letters. For example, u_i denotes the concentration of U_i and x_j denotes the concentration of X_j.*

Given a MESSI network and a partition of the species set as in Definition 2.3, we define for any $\alpha \geq 1$ the set of indices

$$\text{Int}(\alpha) = \{k : \text{there exists } y_{ij} \text{ with either } X_i \in \mathscr{S}^{(\alpha)} \text{ or } X_j \in \mathscr{S}^{(\alpha)} \text{ such that } y_{ij} \to_\circ y_k\}. \tag{5}$$

We also denote by $\mathscr{S}\text{Int}(\alpha)$ the set of species with indices in $\text{Int}(\alpha)$. Note that the subsets $\text{Int}(\alpha)$ are in general not disjoint, but condition $(\mathcal{C})$ implies that $\cup_{\alpha\geq 1}\mathscr{S}\text{Int}(\alpha) = \mathscr{S}^{(0)}$. It is straightforward to see that the conditions imposed on a MESSI network ensure that for any $\alpha \geq 1$ the set of variables $\mathscr{S}^{(\alpha)} \cup \mathscr{S}\text{Int}(\alpha)$ is a siphon [1]. We will show in Theorem 3.2 below that the following explicit linear conservation relations with $\{0, 1\}$ coefficients hold:

$$\ell_\alpha(u, x) = C_\alpha, \text{ where } \ell_\alpha(u, x) = \sum_{X_i \in \mathscr{S}^{(\alpha)}} x_i + \sum_{k \in \text{Int}(\alpha)} u_k, \tag{6}$$

for some constant C_α, which is positive if the trajectory intersects the positive orthant. This is a direct consequence of Theorem 2.1 in [14] and of Theorem 5.3 in [17]. The second part of Theorem 3.2 gives sufficient conditions for these relations to generate *all* the equations defining a stoichiometric compatibility class. We show in Example 3.4 that if we relax any of these conditions, the result is not true. See also Proposition 3.6 on the conditions to ensure that the kinetic and the stoichiometric subspaces coincide.

Theorem 3.2. *Given a chemical reaction network G and a partition of the set of species $\mathscr{S}$ as in (4) that defines a MESSI structure, for each subset of species $\mathscr{S}^{(\alpha)}$, $1 \leq \alpha \leq m$, the linear form ℓ_α in (6) defines a conservation relation of the system. In particular, all MESSI systems are conservative.*

Furthermore, if there are no swaps in G, and the partition is minimal *in the poset of partitions defining a MESSI system structure on G, then* $\dim(S^\perp) = m$.

If, moreover, the stoichiometric subspace coincides with the kinetic subspace, then the only possible conservation relations in the system are linearly generated by the conservations (6) for $1 \leq \alpha \leq m$.

Example 3.3 (Examples 2.7 and 2.8, continued). For the cascade with one phosphatase in Example 2.7, the hypotheses in Theorem 3.2 are satisfied and the conservation relations are the following:

$$s_0 + s_1 + u_1 + u_2 + u_3 = S_{tot}, \quad p_0 + p_1 + u_3 + u_4 = P_{tot},$$
$$e + u_1 = E_{tot}, \quad f + u_2 + u_4 = F_{tot},$$

where we use small letters for the concentration of the corresponding species. The concentrations of the intermediates species es_0, fs_1, s_1p_0, fp_1 are denoted by u_1, u_2, u_3, u_4, respectively. In Example 2.8, the conservation relations are

$$x + x_t + x_p + x_py + x_ty_p = X_{tot}, \quad y + y_p + x_py + x_ty_p = Y_{tot}.$$

Example 3.4 (necessity of the hypotheses in Theorem 3.2). The following is Example 22 from [38]. It satisfies the hypotheses in Theorem 3.2 except for the absence of swaps:

$$X_1 + X_5 \to X_2 + X_6,$$
$$X_3 + X_6 \to X_4 + X_5,$$
$$X_4 + X_6 \to X_3 + X_7.$$

It is straightforward to see that the only possible *minimal* partition is $\mathscr{S}^{(1)} = \{X_1, X_2\}$, $\mathscr{S}^{(2)} = \{X_3, X_4\}$, $\mathscr{S}^{(3)} = \{X_5, X_6, X_7\}$, which gives three linearly independent conservation relations ℓ_1, ℓ_2, ℓ_3. However, there is a fourth independent conservation relation:

$$x_1 + x_4 + x_6 + 2x_7 = C.$$

Before stating the sufficient conditions to ensure that the kinetic and the stoichiometric subspaces coincide, we recall some concepts from graph theory that will be useful in the rest of the article.

Given a directed graph $G = (\mathcal{V}, \mathcal{E})$, define the following equivalence relation between the vertices: two vertices $i, j \in \mathcal{V}$ are related if and only if there is a directed path from i to j, and a directed path from j to i. Equivalence classes of vertices define the vertices of the *strongly connected* components of G. Thus, a directed graph is strongly connected when for each ordered pair of vertices there is a directed path from the first vertex to the second one. Note that the underlying undirected graph of a strongly connected graph is connected. If one strongly connected component has no edges from any node in the component to a node in a different strongly connected component, it is called a *terminal strongly connected component*.

A directed graph G is said to be *weakly reversible* if each connected component is strongly connected. This means that if there is a directed path from a vertex i to another vertex j, there is also a directed path from j to i, but it could happen that no path exists in any of the two directions. Thus G is strongly connected if and only if it is weakly reversible and connected, and the connected components of a weakly reversible graph are strongly connected.

Example 3.5. The underlying directed graph of the chemical reaction network

$$X_3 \overset{\kappa_1}{\leftarrow} X_1 \underset{\kappa_3}{\overset{\kappa_2}{\rightleftarrows}} X_2 \overset{\kappa_4}{\to} X_4$$

is connected but not weakly reversible. It has three strongly connected components: the node X_3 (with no arrows), the node X_4 (again, with no arrows), which are terminal strongly connected components, and the subgraph $X_1 \rightleftarrows X_2$, which is not terminal.

The following result is from [12].

Proposition 3.6. *If G has only one terminal strongly connected component in each connected component, the number of generators of the conservation relations is $s - \dim(S)$, where s is the total number of species and S is the stoichiometric subspace. In this case, the stoichiometric and the kinetic subspaces coincide.*

When there is more than one terminal strongly connected component in one connected component, even if there are no swaps, we can find other conservation relations. For instance, consider the chemical reaction network in Example 3.5 and the partition of the set of species: $\mathscr{S}^{(0)} = \emptyset$ and $\mathscr{S}^{(1)} = \{X_1, X_2, X_3, X_4\}$. Besides the linear relation $x_1 + x_2 + x_3 + x_4 = C_1$, we get another independent relation: $\kappa_4\kappa_1 x_2 - \kappa_4\kappa_2 x_3 + \kappa_1(\kappa_3 + \kappa_4)x_4 = C_2$.

The associated digraphs. Consider a directed graph $G = (\mathcal{V}, \mathcal{E}, \boldsymbol{\kappa})$ with a partition of the set of species which defines a MESSI structure in the network. We associate to G three other digraphs, denoted by G_1, G_2, G_E.

Definition 3.7. *Given a chemical reaction network with directed graph $G = (\mathcal{V}, \mathcal{E}, \boldsymbol{\kappa})$, together with a partition of the set of species $\mathscr{S}$ which defines a MESSI structure in the network with p intermediate species and n core species as in (4), we associate a digraph $G_1 = (\mathcal{V}_1, \mathcal{E}_1)$ with a set of n species consisting of the core species in G and with the inherited partition:*

$$\mathscr{S}_1 = \mathscr{S}^{(1)} \bigsqcup \mathscr{S}^{(2)} \bigsqcup \cdots \bigsqcup \mathscr{S}^{(m)} = \mathscr{S} \setminus \mathscr{S}^{(0)}. \tag{7}$$

The vertex set $\mathcal{V}_1$ consists of all the core complexes y_{ij} and the edge set is equal to $\mathcal{E}_1 = \{y_{ij} \to y_{\ell m} : y_{ij}, y_{\ell m} \in \mathcal{V}_1$ and $y_{ij} \to_\circ y_{\ell m}$ in $G\}$.

Note that G_1 might have loops. It is easy to check that partition (7) defines a MESSI structure on G_1 for any choice of positive labels in $\mathbb{R}^{\#\mathcal{E}_1}_{>0}$.

We now define a chemical reaction network on G_1 by decorating the edges $\mathcal{E}_1$ with labels $\boldsymbol{\tau}(\boldsymbol{\kappa})$, which are rational functions of the original rate constants $\boldsymbol{\kappa}$, following [14, Theorem 3.1].

Definition 3.8. *The map $\boldsymbol{\tau} : \mathbb{R}^{\#\mathcal{E}}_{>0} \to \mathbb{R}^{\#\mathcal{E}_1}_{>0}$ is defined as follows. For each $X_i + X_j \to_\circ X_\ell + X_m$ in G the reaction constant τ in G_1 which gives the label $X_i + X_j \xrightarrow{\tau} X_\ell + X_m$ has the form*

$$\tau = \kappa + \sum_{k=1}^{p} \kappa_k \mu_k, \tag{8}$$

where $\kappa \geq 0$ is positive when $X_i + X_j \xrightarrow{\kappa} X_\ell + X_m$ in G (and $\kappa = 0$ otherwise), and $\kappa_k \geq 0$ is positive if $U_k \xrightarrow{\kappa_k} X_\ell + X_m$ and $X_i + X_j \to_\circ U_k$ in G (and $\kappa_k = 0$ otherwise). The explicit expression of the coefficients μ_k is given in display (15) in the proof of Theorem 3.1 in the electronic supplementary material (ESM) of [14]; we will describe them for particular cases of interest to us in section 4.

It is straightforward to see that $\boldsymbol{\tau}$ defines a rational map (that is, $\mathbb{Q}(\tau) \subset \mathbb{Q}(\boldsymbol{\kappa})$). The main property of this assignment is the following.

Remark 3.9. When we label the edges in G_1 with the real constants $\boldsymbol{\tau}(\boldsymbol{\kappa}) \in \mathbb{R}_{>0}^{\#\mathcal{E}_1}$, the steady states of the mass-action chemical reaction systems defined by G and G_1 are in one-to-one correspondence. We refer the reader to the proof of Theorem 3 in the ESM of [14] and to the more recent article [28].

We now introduce a new associated labeled digraph G_2.

Definition 3.10. *Consider a chemical reaction network with directed graph $G = (\mathcal{V}, \mathcal{E}, \boldsymbol{\kappa})$, together with a partition of the set of species $\mathscr{S}$ which defines a MESSI structure in the network, and its associated labeled digraph $G_1 = (\mathcal{V}_1, \mathcal{E}_1, \boldsymbol{\tau})$ from Definition 3.7. We first define a labeled multidigraph where we "hide" the concentrations of some of the species in the labels. The species set $\mathcal{V}_2$ of $G_2 = (\mathcal{V}_2, \mathcal{E}_2, \boldsymbol{\tau}_x)$ is again equal to the set of core species $\mathscr{S}_1$, with the induced partition.*

The edge set $\mathcal{E}_2$ is defined as follows. We keep all monomolecular reactions $X_i \to X_j$ in $\mathcal{E}_1$ and for each reaction $X_i + X_\ell \xrightarrow{\tau} X_j + X_m$ in $\mathcal{E}_1$, with $X_i, X_j \in \mathscr{S}^{(\alpha)}$, $X_\ell, X_m \in \mathscr{S}^{(\beta)}$, we consider two reactions $X_i \xrightarrow{\tau x_\ell} X_j$ and $X_\ell \xrightarrow{\tau x_i} X_m$. We obtain in principle a multidigraph MG_2 that might contain loops or parallel edges between any pair of nodes (i.e., directed edges with the same source and target nodes). We define the digraph G_2 by collapsing into one edge all parallel edges in MG_2, and we define the labels $\boldsymbol{\tau}_x$ of the edges in $\mathcal{E}_2$ as the sum of the labels of the corresponding collapsed edges in MG_2.

We will moreover denote by G_2° the digraph obtained from the deletion of loops and isolated nodes of G_2.

By rules $(\mathcal{R}_1)$, $(\mathcal{R}_2)$, and $(\mathcal{R}_3)$, G_2 is a *linear* graph (its vertices are labeled by a single species). The labels on the edges of MG_2 (and of G_2) depend on the rate constants but might also depend on the concentrations $x_1, \dots, x_n$.

Example 3.11 (Examples 2.7 and 2.8, continued). The graphs G_1 and G_2° associated to the networks in Examples 2.7 and 2.8 are depicted in Figure 3.

Remark 3.12. We get the following important fact from the definition of the associated digraphs and networks for any MESSI network with digraph G: the networks of the associated digraphs G_1 and G_2 *determine the same polynomial equations.* They moreover define, together with the corresponding equations of the intermediate species, the steady states of G. We have already observed in Remark 3.9 that this is the case for G_1. Indeed, if we consider G_2 in a mass-action fashion, we can see that the same terms are added and substracted, obtaining the same equations associated to G_1. However, we cannot recover the dynamical properties of G_1 (nor G) from G_2 since we admit species (concentrations) as both vertices and edge labels.

Note that for each $\alpha \geq 1$, if one species of $\mathscr{S}^{(\alpha)}$ appears on a vertex of G_2, by $(\mathcal{R}_2)$ and $(\mathcal{R}_3)$ and the construction of G_2, all the species in the vertices of the corresponding connected component of G_2 belong to the *same* subset $\mathscr{S}^{(\alpha)}$ in the original partition (4). In fact, the same partition (7) defines a MESSI structure on G_2. Moreover, we have the following.

Lemma 3.13. *The partition of the set of species $\mathscr{S}$ of G in (4) is minimal in the poset of*

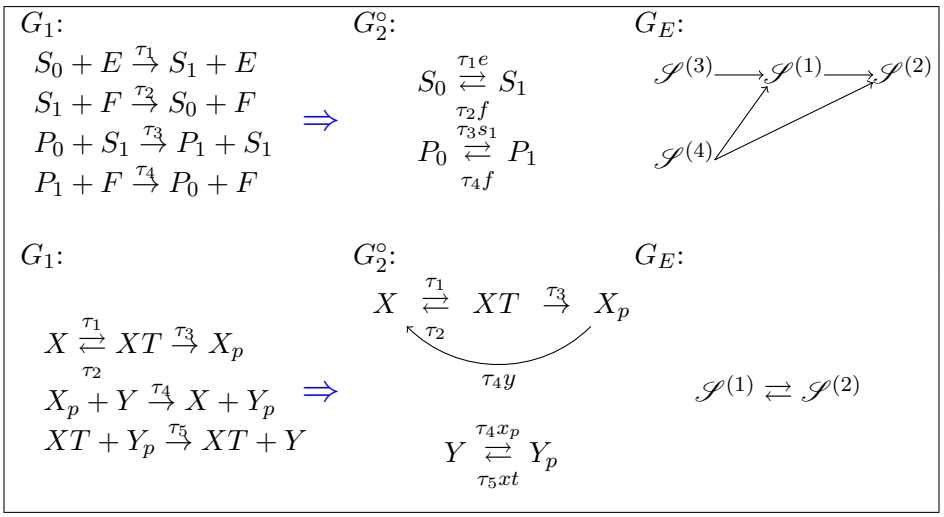


Figure 3. *The graphs G_1, G_2°, and G_E for the running examples. The corresponding sets $\mathscr{S}^{(\alpha)}$ can be found in Example 3.17.*

partitions defining a MESSI structure on the network if and only if the set of intermediate species is maximal, the connected components of G_2 are in bijection with the subsets $\mathscr{S}^{(\alpha)}$, and the set of nodes of the corresponding component equals $\mathscr{S}^{(\alpha)}$. Thus, by considering the connected components in G_2 we can refine any partition of the species set $\mathscr{S}$ to a minimal one defining a MESSI structure on G.

We finally define the associated digraph G_E.

Definition 3.14. *Consider a MESSI network with directed graph G, together with a* minimal *partition of the set of species as in (4). Let G_2 and G_2^0 be as in Definition 3.10. We define a new digraph $G_E = (\mathcal{V}_E, \mathcal{E}_E)$. The set of vertices equals $\mathcal{V}_E = \{\mathscr{S}^{(\alpha)}, \alpha \geq 1\}$. The pair $(\mathscr{S}^{(\alpha)}, \mathscr{S}^{(\beta)})$ lies in $\mathcal{E}_E$ when there is a species in $\mathscr{S}^{(\alpha)}$ in a label of an edge in G_2^0 between (different) species of $\mathscr{S}^{(\beta)}$.*

Example 3.17 below shows the corresponding digraphs G_E for our two running examples.

Persistence. As MESSI systems are conservative by Theorem 3.2, we know by Theorem 2 in [1] that a MESSI system is persistent when there are no *relevant boundary steady states*. This means that there are no steady states in the intersection of the boundary $\partial(\mathbb{R}^s_{\geq 0})$ of the nonnegative orthant with a stoichiometric compatibility class through a point in $\mathbb{R}^s_{>0}$. Persistence means that any trajectory starting from a point with positive coordinates stays at a positive distance from any point in the boundary.

Note that a necessary condition for system (2) to have a positive steady state is the existence of a positive relation among the vectors $y' - y$, that is, a positive vector λ such that $\sum_{y \to y'} \lambda_{yy'}(y' - y) = 0$. If this is satisfied, we will say that the system is *consistent*.

We give in Theorem 3.15 combinatorial conditions which ensure the persistence and consistency of MESSI systems. This result rules out relevant boundary steady states in many enzymatic examples—for instance, in those in [1].

Recall that a digraph is weakly reversible if any connected component is strongly connected, that is, when for any pair of nodes in the same connected component there is a directed path joining them. We have the following persistence result.

Theorem 3.15. *Let G be the underlying digraph of a MESSI system. Assume that the associated digraph G_2 is weakly reversible and the associated digraph G_E has no directed cycles. Then G has no relevant boundary steady states and so the system is persistent. Moreover, the system is consistent.*

Remark 3.16. The absence of directed cycles in G_E precludes the existence of swaps. On the other side, note that if G_2 is weakly reversible, then the stoichiometric and the kinetic subspaces coincide by Proposition 3.6.

Example 3.17 (Examples 2.7 and 2.8, continued). The MESSI network in Example 2.7 from Figure 2(A) (with partition $\mathscr{S}^{(1)} = \{S_0, S_1\}$, $\mathscr{S}^{(2)} = \{P_0, P_1\}$, $\mathscr{S}^{(3)} = \{E\}$, $\mathscr{S}^{(4)} = \{F\}$) is persistent since there are no directed cycles in G_E (depicted at the upper right in Figure 3). However, this is not the case in Example 2.8 from Figure 1(D); $x_p = X_{tot}, y_p = Y_{tot}, x = x_t = x_p y = x_t y_p = y = 0$ is a boundary steady state in the stoichiometric compatibility class defined by X_{tot}, Y_{tot}. Recall that we are considering the (minimal) partition $\mathscr{S}^{(1)} = \{X, XT, X_p\}$, $\mathscr{S}^{(2)} = \{Y, Y_p\}$. The associated graph G_E has a cycle (depicted at the lower right in Figure 3).

4. Parametrizing the steady states.

A wide class of MESSI systems admits a rational parametrization. As we recalled in Remark 3.12, it is shown in [14] that the values of the intermediate species at steady state can be rationally written in terms of the core species in an algorithmic way. The following result (with the same assumptions as Theorem 3.15) extends Theorem 4 in [45].

Theorem 4.1. *Let G be the underlying digraph of a MESSI system. Assume that the associated digraph G_2 is weakly reversible and the associated digraph G_E has no directed cycles. Then, $V_f \cap \mathbb{R}^s_{>0}$ admits a rational parametrization, which can be algorithmically computed. More explicitly, it is possible to define levels for the subsets $\mathscr{S}^{(\alpha)}, \alpha \geq 1$, according to indegree. Then, given any choice of one index i_α in each $\mathscr{S}^{(\alpha)}$, the concentration of any core species x_i in a subset $\mathscr{S}^{(\beta)}$ can be rationally expressed in an effective way in terms of x_{i_β} and the variables x_{i_α} for which the indegree of $\mathscr{S}^{(\alpha)}$ is strictly smaller than the indegree of $\mathscr{S}^{(\beta)}$.*

Moreover, if the partition is minimal with m subsets of core species, the dimension of $V_f \cap \mathbb{R}^s_{>0}$ equals m and $m = \dim(S^\perp)$.

Recall that a binomial is a polynomial with two terms and that a Laurent monomial is a monomial with integer exponents, which can be negative.

Definition 4.2. *A* toric MESSI system *is a MESSI system whose positive steady states $V_f \cap \mathbb{R}^s_{>0}$ can be described with* binomials.

It is well known that the real positive points of a nonempty algebraic variety described by binomials can always be parametrized by Laurent monomials. This implies that if the MESSI system is toric, there exists a rational parametrization even if G_E has directed cycles, as long as the system is consistent.

We now show that many common MESSI systems are toric in an explicit way coming from the structure of the network, which we call s-toric.

In order to define s-toric MESSI systems, we need to use some concepts from graph theory. A spanning tree of a digraph is a subgraph that contains all the vertices and is connected and acyclic as an undirected graph. An i-tree of a graph is a spanning tree where the ith vertex is its unique sink (equivalently, it is the only vertex of the tree with no edges leaving from it). For an i-tree T, call c^T the product of the labels of all the edges of T. For the associated graph G_2 of a MESSI network G, the products c^T are monomials depending in principle on both the rate constants τ and the x-variables.

Definition 4.3. *A* structurally toric*, or* s-toric MESSI system*, is a MESSI system whose digraph G satisfies the following conditions:*
($\mathcal{C}'$) *Condition ($\mathcal{C}$) holds, and moreover, for every intermediate complex y_k there exists a* unique *core complex y_{ij} such that $y_{ij} \to_\circ y_k$ in G.*
($\mathcal{C}''$) *The associated multidigraph MG_2 does not have parallel edges, and the digraph G_2 is weakly reversible.*
($\mathcal{C}'''$) *For each $i \in \{1, \dots, n\}$ and any choice of i-trees T, T' of G_2°, the quotient $c^T/c^{T'}$ only depends on the rate constants τ.*

Examples of networks satisfying condition ($\mathcal{C}'''$) are the phosphorylation cascades, as there is a unique i-tree for each i. Our second running Example 2.8 also has this property (see Example 4.4). Moreover, phosphorylation cascades, the multisite sequential distributive phosphorylation system, the multisite processive phosphorylation system, and the bacterial EnvZ/OmpR network depicted in Figure 1 are s-toric MESSI systems.

Example 4.4 (running Example 2.8, continued). For the system in Example 2.8, the graph G_2° is

$$X \underset{\tau_2}{\overset{\tau_1}{\rightleftarrows}} XT \xrightarrow{\tau_3} X_p \qquad Y \underset{\tau_5 x_T}{\overset{\tau_4 x_p}{\rightleftarrows}} Y_p.$$

In this case, there are two X-trees: $\tau_4 y$

$$T_1: \; X \xleftarrow[\tau_4 y]{\tau_2} XT \; X_p, \qquad T_2: \; X \; XT \xrightarrow[\tau_4 y]{\tau_3} X_p.$$

However, $c^{T_1} = \tau_2\tau_4 y$, $c^{T_2} = \tau_3\tau_4 y$, and $c^{T_1}/c^{T_2} = \tau_2/\tau_3$, which only depends on the rate constants τ_i. For the other vertices, the corresponding tree is unique, and therefore this MESSI network is s-toric.

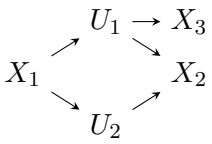
We now clarify the meaning of condition ($\mathcal{C}'$).

Example 4.5. Network (A) on the left of Figure 4 satisfies condition ($\mathcal{C}'$), while network (B) on the right does not since both core complexes X_1 and X_2 react via intermediates to the intermediate complex U_2.

We will use the following notation.

Notation 4.6. *Given an intermediate complex y_k of an s-toric MESSI system, denote by y_{ij} the unique core complex reacting through intermediates to y_k and denote by $x^{\varphi(k)}$ the monomial*

$$x^{\varphi(k)} = \begin{cases} x_i x_j & \text{if } \; y_{ij} = X_i + X_j, \\ x_i & \text{if } \; j = 0 \text{ and } y_{ij} = X_i. \end{cases} \tag{9}$$

(A) 

(B) $X_1 \rightleftarrows U_1 \to U_2 \rightleftarrows X_2.$

Figure 4. *Validity of condition* $(\mathcal{C}')$.

As we recalled in Remark 3.9, the rational map $\boldsymbol{\tau} : \mathbb{R}^{\#\mathcal{E}}_{>0} \to \mathbb{R}^{\#\mathcal{E}_1}_{>0}$ in Definition 3.8 verifies that the steady states of the mass-action chemical reaction systems defined by G with rate constants $\boldsymbol{\kappa}$ and G_1 with rate constants $\boldsymbol{\tau}(\boldsymbol{\kappa})$ are in one-to-one correspondence via the projection $\pi(u, x) = x$. We now give conditions for the inverse of this projection to be a monomial map in the concentrations of the core species.

Proposition 4.7. *Given a MESSI network G that satisfies condition $(\mathcal{C}')$ in Definition 4.3, there are (explicit) rational functions $\mu_k \in \mathbb{Q}(\boldsymbol{\kappa}), 1 \le k \le p$, such that for any steady state $x \in \mathbb{R}^n_{>0}$ of the associated MESSI network G_1, the steady state $\pi^{-1}(x) = (u(x), x)$ of G is given by the monomial map:*

$$u_k(x) = \mu_k\, x^{\varphi(k)}, \quad k = 1, \dots, p. \tag{10}$$

The rational functions μ_k are in simple cases the usual Michaelis–Menten constants associated with the original rate constants $\boldsymbol{\kappa}$.

It holds that an s-toric MESSI system is toric, and, moreover, its positive steady states can be described by explicit binomials.

Theorem 4.8. *Any s-toric MESSI system is toric. Moreover, we can choose $s - m'$ explicit binomials with coefficients in $\mathbb{Q}(\boldsymbol{\kappa})$ which describe the positive steady states, where m' is the number of connected components of G_2.*

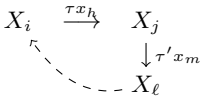
In particular, given a MESSI network G with a partition of the set of species as in (4), assume that for each $\alpha \ge 1$ and $X_i \ne X_j \in \mathscr{S}^{(\alpha)}$ in the same connected component of G_2 there exists a unique simple path P_{ji} in G_2° from X_j to X_i.[2] *Then, the associated dynamical system is s-toric and there exist explicit μ_k and η_{ij} in $\mathbb{Q}(\boldsymbol{\kappa})$ such that the $s - m'$ binomials describing the positive steady states can be chosen from the following:*

$$u_k - \mu_k x^{\varphi(k)} = 0 \tag{11}$$

for each intermediate U_k $(1 \le k \le p)$,

$$x_h x_i - \eta_{ij} x_m x_j = 0 \tag{12}$$

if $X_i \xrightarrow{\tau x_h} X_j$ is in G_2° and $X_j \xrightarrow{\tau' x_m} X_\ell$ is in P_{ji}.



Example 4.9 (running Example 2.7, continued). Recall that the graph G_2° for the cascade in Example 2.7 is

$$S_0 \underset{\tau_2 f}{\overset{\tau_1 e}{\rightleftarrows}} S_1, \quad P_0 \underset{\tau_4 f}{\overset{\tau_3 s_1}{\rightleftarrows}} P_1,$$

[2] A *simple path* is a path that visits each vertex exactly once.

and the graph G_2 has two extra connected components, corresponding to the isolated nodes E and F. Clearly, for each vertex in G_2° there is only one simple directed path from the other vertex in the same connected component. For example, the only S_1-tree, T, is $S_0 \overset{\tau_1 e}{\to} S_1$ and $c^T = \tau_1 e$.

We denote the concentration of the intermediate species es_0, fs_1, s_1p_0, fp_1 by u_1, u_2, u_3, u_4, respectively. The corresponding rational functions $\mu_1, \ldots, \mu_4$ in the statement of Proposition 4.7 equal

$$\mu_1 = \frac{\kappa_1}{\kappa_2 + \kappa_3}, \quad \mu_2 = \frac{\kappa_4}{\kappa_5 + \kappa_6}, \quad \mu_3 = \frac{\kappa_7}{\kappa_8 + \kappa_9}, \quad \mu_4 = \frac{\kappa_{10}}{\kappa_{11} + \kappa_{12}}.$$

We further denote $\eta_1 = \frac{\tau_2}{\tau_1}, \eta_2 = \frac{\tau_4}{\tau_3}$. According to Theorem 4.8, the following $6 = 10 - 4$ binomials describe the positive steady states of the associated MESSI system:

$$u_1 - \mu_1 e.s_0 = u_2 - \mu_2 f.s_1 = u_3 - \mu_3 s_1.p_0 = u_4 - \mu_4 f.p_1 = e.s_0 - \eta_1 f.s_1 = s_1.p_0 - \eta_2 f.p_1 = 0.$$

The first four binomials correspond to (11), and the last two occur in (12).

5. Toric MESSI systems and multistationarity.

We present in this section a necessary and sufficient criterion to decide whether a system is multistationary, which holds for toric MESSI systems (see Definitions 4.2 and 4.3). Again, the assumptions we make seem to be very restrictive. Nevertheless, it can be easily seen that all standard phosphorylation cascades, multisite sequential phosphorylation networks, and many two component bacterial networks are of this form, so there is a wide range of applications. This is summarized in Theorems 5.4 and 5.8. We implemented this result by means of Algorithm 1, which certifies mono- or multistationarity and in this last case provides different choices of rate constants for which multistationarity occurs.

Necessary and sufficient conditions.

Theorem 5.4 below gives a necessary and sufficient criterion to detect the capacity for multistationarity of a toric MESSI system. It is deduced from results in [32] and [34]. Then, we give in Proposition 5.6 checkable conditions that ensure the validity of the hypotheses of Theorem 5.4. When the system is not monostationary, we finally show in Theorem 5.8 how to choose rate constants for which the system shows multistationarity (see also [4, 10]).

Notation 5.1. *Let G be a MESSI network. Assume the positive steady states of the associated dynamical system are described by binomials $x^{v'} - \eta x^v$. We call T the subspace of $\mathbb{R}^n$ generated by all these vectors $v' - v$. Choose any matrix B whose columns form a basis of T. For a positive vector x write $(x^B)_j = x^{B_j}$, where B_j denotes the jth column of B. Then, there exists a constant vector $\boldsymbol{\eta}$ such that x is a positive steady state of the associated system if and only if $x^B = \boldsymbol{\eta}$. Considering the orthogonal complement of T in $\mathbb{R}^s$, we construct another matrix $B^\perp$ whose rows form a basis of the orthogonal subspace $T^\perp$. We can choose both B and $B^\perp$ with integer entries. We consider also a matrix M whose columns form a basis of the stoichiometric subspace S. Again, we construct a matrix $M^\perp$ whose rows form a basis of the orthogonal complement $S^\perp$. Thus, when the stoichiometric and the kinetic spaces coincide, the row vectors of $M^\perp$ are the coefficients of a basis of linear conservation relations. For any natural number s we denote $[s] = \{1, \ldots, s\}$. Given a matrix $A \in \mathbb{R}^{d \times s}$ with $s \geq d$ and a*

subset $J \subseteq [s]$, we denote by A_J the submatrix of A with column indices in J. We furthermore denote J^c the complement of J in $[s]$ and $\nu(J) = \sum_{j\in J} j$. An orthant $\mathcal{O} \subset \mathbb{R}^s$ is defined by the signs of the coordinates of its points and it will be identified with a vector in $\{-1, 0, 1\}^s$.

Definition 5.2. *Given matrices $M^\perp$ and $B^\perp$ as above, with $d = \operatorname{rank}(M^\perp) = \operatorname{rank}(B^\perp)$, we define the following sets of signed products:*

$$\begin{aligned}
\Sigma &= \{\operatorname{sign}(\det(M_I^t)\det(B_I^t)) : I \subseteq [s], \#I = s-d\},\\
\Sigma^\perp &= \{\operatorname{sign}((-1)^{\nu(J)}\det(M_J^\perp)\det(B_{J^c}^t)) : J \subseteq [s], \#J = d\},\\
\Sigma_\perp &= \{\operatorname{sign}((-1)^{\nu(J)}\det(M_{J^c}^t)\det(B_J^\perp)) : J \subseteq [s], \#J = d\},\\
\Sigma_\perp^\perp &= \{\operatorname{sign}(\det(M_J^\perp)\det(B_J^\perp)) : J \subseteq [s], \#J = d\}.
\end{aligned}$$

We say that a set $\sigma \neq \{0\}$ of signs is mixed *if $\{-,+\} \subset \sigma$ and* unmixed *otherwise.*

The following lemma is a consequence of Lemma 2.10 in [32] (and the references therein).

Lemma 5.3. *With the notation of Definition 5.2, if any of the four signs sets $\Sigma, \Sigma^\perp, \Sigma_\perp, \Sigma_\perp^\perp$ is different from $\{0\}$, the four of them are, and if so, if any of the four is mixed, all of them are mixed.*

The following theorem gives a necessary and sufficient criterion to determine if the toric MESSI system is monostationary, based on [32] and [34].

Theorem 5.4. *Let G be a toric MESSI network with matrices M and B as above, which verifies that $\operatorname{rank}(M) = rank(B) = d$ and the signs sets $\Sigma, \Sigma^\perp, \Sigma_\perp, \Sigma_\perp^\perp$ are different from $\{0\}$. Then, the following statements are equivalent:*

1. *The associated MESSI system is monostationary.*
2. *The signs sets $\Sigma, \Sigma^\perp, \Sigma_\perp, \Sigma_\perp^\perp$ are unmixed.*
3. *For all orthants $\mathcal{O} \in \{-1, 0, 1\}^s, \mathcal{O} \neq \mathbf{0}$, either $S \cap \mathcal{O} = \emptyset$ or $T^\perp \cap \mathcal{O} = \emptyset$.*

Example 5.5 (Example 2.7, continued). Consider the two phosphorylation cascades in Figure 2. Both cascades differ in the phosphatases: the cascade in Figure 2(B) has different phosphatases for each layer, while the cascade (A) does not. The set Σ corresponding to the cascade in (B) is unmixed, which according to Theorem 5.4 implies that the system is monostationary. In contrast, the set Σ for the cascade in (A) is mixed, and the system has the capacity for multistationarity. For instance, if we consider J the set of indices corresponding to S_0, P_0, ES_0, and FP_1, and $\tilde{J}$ the set of indices corresponding to S_0, P_1, ES_0, and FP_1 (where $4 = \operatorname{rank}(M^\perp) = \operatorname{rank}(B^\perp)$), $\operatorname{sign}(\det(M_J^\perp)\det(B_J^\perp)) \neq \operatorname{sign}(\det(M_{\tilde{J}}^\perp)\det(B_{\tilde{J}}^\perp))$, and they are both nonzero.

If we add the reactions $P_1 + D \rightleftarrows P_1D$, which represent a drug interacting with the phosphorylated form P_1, we can check that this new system remains multistationary for the cascade (A). The new matrices $\hat{M}$ and $\hat{B}$ can be obtained in the following way:

$$\hat{M}^t = \left(\begin{array}{ccccc|ccc} & & & & & P_1 & D & P_1D \\ & & & & & * & 0 & 0 \\ & & M^t & & & \vdots & \vdots & \vdots \\ & & & & & * & 0 & 0 \\ \hline 0 & \dots & 0 & 1 & & 1 & 1 & -1 \end{array}\right), \qquad \hat{B}^t = \left(\begin{array}{ccccc|ccc} & & & & & P_1 & D & P_1D \\ & & & & & * & 0 & 0 \\ & & B^t & & & \vdots & \vdots & \vdots \\ & & & & & * & 0 & 0 \\ \hline 0 & \dots & 0 & 1 & & 1 & 1 & -1 \end{array}\right).$$

Both sets of indices J and $\tilde{J}$ witnessing multistationarity do not contain P_1. Then, from the structure of the matrix $\operatorname{sign}(\det(\hat{M}^t_{J\cup\{P_1\}})\det(\hat{B}^t_{J\cup\{P_1\}})) \neq \operatorname{sign}(\det(\hat{M}^t_{\tilde{J}\cup\{P_1\}})\det(\hat{B}^t_{\tilde{J}\cup\{P_1\}}))$, which by Theorem 5.4 ensures that the cascade with the drug is multistationary.

For s-toric MESSI systems we give in Proposition 5.6 below sufficient conditions for the hypothesis in Theorem 5.4 that the ranks of M and B coincide. These conditions are not necessary, but if any of them is not satisfied, the ranks might be different.

Proposition 5.6. *Let G be an s-toric MESSI network G. Assume that the partition is minimal with m subsets of core species and the associated digraph G_E has no directed cycles. Then,* $\operatorname{rank}(B^\perp) = \operatorname{rank}(M^\perp) = m$.

Example 5.7 (necessity of the hypothesis about G_E in Proposition 5.6). If there are directed cycles in G_E, we cannot assert that $\operatorname{rank}(M^\perp) = \operatorname{rank}(B)$. Consider, for instance, the following MESSI network without intermediate complexes:

$$\begin{aligned} S_0 + R_1 &\overset{\kappa_1}{\to} S_1 + R_1, & S_1 + R_0 &\overset{\kappa_2}{\to} S_0 + R_0,\\ P_0 + S_0 &\overset{\kappa_3}{\to} P_1 + S_0, & P_1 + S_1 &\overset{\kappa_4}{\to} P_0 + S_1,\\ R_0 + P_0 &\overset{\kappa_5}{\to} R_1 + P_0, & R_1 + P_1 &\overset{\kappa_6}{\to} R_0 + P_1, \end{aligned}$$

where $\mathscr{S}$ is the disjoint union of $\mathscr{S}^{(1)} = \{S_0, S_1\}$, $\mathscr{S}^{(2)} = \{P_0, P_1\}$, and $\mathscr{S}^{(3)} = \{R_0, R_1\}$. The corresponding digraph G_2 equals

$$S_0 \underset{\kappa_2 r_0}{\overset{\kappa_1 r_1}{\rightleftarrows}} S_1, \qquad P_0 \underset{\kappa_4 s_1}{\overset{\kappa_3 s_0}{\rightleftarrows}} P_1, \qquad R_0 \underset{\kappa_6 p_1}{\overset{\kappa_5 p_0}{\rightleftarrows}} R_1,$$

and the digraph G_E is a cycle: $\mathscr{S}^{(1)} \to \mathscr{S}^{(2)} \to \mathscr{S}^{(3)} \to \mathscr{S}^{(1)}$.

We call s_0, s_1 the concentrations of S_0, S_1 (respectively), p_0, p_1 the concentrations of P_0, P_1, and r_0, r_1 the concentrations of R_0, R_1. There are three linearly independent conservation relations:

$$s_0 + s_1 = C_1, \quad p_0 + p_1 = C_2, \quad r_0 + r_1 = C_3.$$

We expect the rank of B to be 3. But the system equals

$$ds_0/dt = -\kappa_1 s_0 r_1 + \kappa_2 s_1 r_0,\ dp_0/dt = -\kappa_3 s_0 p_0 + \kappa_4 s_1 p_1,\ dr_0/dt = -\kappa_5 p_0 r_0 + \kappa_6 p_1 r_1,$$

and so we can choose B to be the matrix

$$\begin{pmatrix} -1 & 1 & 0 & 0 & 1 & -1 \\ -1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 \end{pmatrix},$$

which has rank 2.

Assume there exists a positive steady state. Then, we deduce that

$$\kappa_1\kappa_4\kappa_5 = \kappa_2\kappa_3\kappa_6. \tag{13}$$

So, when (13) is not satisfied, there are no positive steady states, and when it is satisfied, any of the three steady state equations is a consequence of the other two, and when we intersect with the linear variety defined by the conservation relations, we get a variety of dimension 1, with an infinite number of positive steady states (there are 5 equations in 6 variables).

If a consistent toric MESSI system is not monostationary, we can effectively construct two different steady states x^1 and x^2 and a reaction rate constant vector $\boldsymbol{\kappa}$ that witness multistationarity based on item (3) in the statement of Theorem 5.4, following the arguments in [34] (see also [4, 10]).

Theorem 5.8. *Let G be a consistent MESSI network which satisfies the hypotheses of Theorem 5.4, such that the associated system is toric and it is not monostationary. Then, for any choice of $\mathbf{w} \in S, \mathbf{v} \in T^\perp$ in the same orthant, the positive vectors x^1 and x^2 defined as*

$$\begin{aligned} \left(x_i^1\right)_{i=1,\dots,s} &= \begin{cases} \frac{w_i}{e^{v_i}-1} & \text{if } v_i \neq 0, \\ \text{any } \bar{x}_i > 0 & \text{otherwise,} \end{cases} \\ \mathbf{x}^2 &= \operatorname{diag}(e^{\mathbf{v}})\, \mathbf{x}^1 \end{aligned}$$

are two different steady states of the given toric MESSI system for any vector of rate constants $\boldsymbol{\kappa}$ which is a positive solution of the linear system $f(x^1, \boldsymbol{\kappa}) = 0$, with $f(x^1, \boldsymbol{\kappa})$ as in (2).

An algorithm to find different steady states in multistationary toric MESSI systems. We present here an algorithm based on Theorems 5.4 and 5.8 which checks whether a consistent toric MESSI system has the capacity for multistationarity. In this case, it looks for orthants where S and $T^\perp$ meet and finds two different steady states in the same stoichiometric compatibility class, together with a corresponding set of reaction constants (based on [4, 10, 34]).

The algorithm to find these orthants relies on the theory of oriented matroids [2, 36, 37]. Recall that the *support* of a vector is defined as the set of its nonzero coordinates. A circuit of a real matrix A is a nonzero element $r \in \operatorname{rowspan}(A)$ with minimal support (with respect to inclusion). Given an orthant $\mathcal{O}$ (resp., a vector v), a circuit r is said to be *conformal* to $\mathcal{O}$ (resp., v) if for any index i *in its support*, $\operatorname{sign}(r_i) = \mathcal{O}_i$ (resp., $\operatorname{sign}(r_i) = \operatorname{sign}(v_i)$). A key result is that every vector $v \in \operatorname{rowspan}(A)$ is a nonnegative sum of circuits conformal to v [37]. All the circuits of A can be described in terms of vectors of maximal minors of A (see Lemma A.5 in Appendix A), and one can thus compute all orthants containing vectors in $\operatorname{rowspan}(A)$ as those orthants $\mathcal{O}$ whose support equals the union of the supports of the circuits conformal to $\mathcal{O}$. These arguments also allow us to check the consistency of a given network, that is, whether there is a positive element in the kernel of a matrix with columns given by the reaction vectors $y' \to y$.

Efficiency can certainly be improved at any step of the algorithm, mainly to avoid unnecessary computations. The rows of $M^\perp$ usually present some nice structure that minimizes the search for orthants containing a circuit, because in the conditions of Theorem 3.2 all columns corresponding to the same set in the partition of the species are equal, which produces many zero minors that can be predicted. In Step 5, infinitely many different choices of $\mathbf{v}$ and $\mathbf{w}$ can be obtained by considering positive linear combinations of all circuits which are conformal to the orthant $\mathcal{O}$ (one circuit per support).

We implemented this algorithm in Octave [8] for the cascades in Figure 2. In the multistationary case of only one phosphatase F, we obtained two different orthants $\mathcal{O}_1, \mathcal{O}_2$ where S and $T^\perp$ meet. In both cases, we computed for $i = 1, 2$ a choice of corresponding rate constants $\boldsymbol{\kappa}(i)$ and two steady states $x^1(i)$ and $x^2(i)$ in the same stoichiometric compatibility class. We

Algorithm 1 Test for multistationarity.

Given a consistent toric MESSI system with network G, the following procedure finds, if they exist, multistationarity parameters $\boldsymbol{\kappa}$ or decides that the system is monostationary.

Input: A toric MESSI network G.

Step 0: Compute matrices $M^\perp$ (or M) and B (or $B^\perp$) for G.

Step 1: Compute $\Sigma^\perp$ (or any of the sets $\Sigma, \Sigma_\perp, \Sigma_\perp^\perp$). Check if $\Sigma^\perp$ is mixed. If it is unmixed, stop and assert that the system is monostationary.

Step 2: Compute the circuits for $B^\perp$ and find an orthant whose support equals the union of the circuits conformal to it.

Step 3: For the orthant computed in Step 2, check if there is a conformal circuit of M contained in this orthant. In this case, check whether its support equals the union of the circuits of M conformal to it. Otherwise, ignore it, and go back to Step 2.

Step 4: For each orthant $\mathcal{O}$ with $S \cap \mathcal{O} \neq \emptyset$ and $T^\perp \cap \mathcal{O} \neq \emptyset$, keep the conformal circuits.

Step 5: Build vectors $\mathbf{v} \in T^\perp$ and $\mathbf{w} \in S$, for example, as the sum of the corresponding conformal circuits.

Step 6: Output x^1, x^2 and $\boldsymbol{\kappa}$ that witness multistationarity, as in Theorem 5.8.

ordered the species S_0, S_1, P_0, P_1, $ES0$, $FS1$, $S1P0$, $FP1$, E, F. We considered in both cases two sets of initial conditions (on the same stoichiometric compatibility class); first we set initial states $S_0 = S_{tot}$, $P_0 = P_{tot}$, $E = E_{tot}$, $F = F_{tot}$ and then initial states $S_0 = S_{tot}$, $P_1 = P_{tot}$, $E = E_{tot}$, $F = F_{tot}$, and all the other species equal to zero. We simulated the system and we depicted the output in Figure 5, which confirms the occurrence of two stoichiometrically compatible steady states for $\boldsymbol{\kappa}(1)$ and $\boldsymbol{\kappa}(2)$. Approximate values are as follows:

$$\boldsymbol{\kappa}(1) \cong (25.46, 0.86, 0.86, 11, 0.86, 0.86, 0.14, 0.21, 0.21, 37.47, 0.21, 0.21),$$

$$x^1(1) \cong (0.037, 3.47, 4.07, 1.02, 1.16, 1.16, 4.75, 4.75, 2.1, 0.052),$$

$$x^2(1) \cong (2.04, 0.47, 11.07, 0.019, 3.16, 3.16, 1.7, 1.75, 0.1, 1.05),$$

and

$$\boldsymbol{\kappa}(2) \cong (101.86, 1.72, 1.72, 33, 0.86, 0.86, 37.47, 0.13, 0.13, 0.42, 0.63, 0.63),$$

$$x^1(2) \cong (0.019, 0.052, 1.02, 4.07, 0.58, 1.16, 7.91, 1.58, 1.05, 1.16),$$

$$x^2(2) \cong (1.02, 1.05, 0.019, 11.07, 1.58, 3.16, 2.9, 0.58, 0.052, 0.16).$$

6. Discussion. Our contribution to the study of many different important biological systems modeled with mass-action kinetics is the identification of a common underlying structure in quite diverse networks. We call this a MESSI structure, since it describes Modifications of type Enzyme-Substrate or Swap with Intermediates. The mathematical formulation of the distinguished properties of MESSI biological systems allows us to prove general results on their dynamics from the structure of the network. We give very precise hypotheses that ensure the validity of our statements and which can be easily verified in common networks of biological interest.

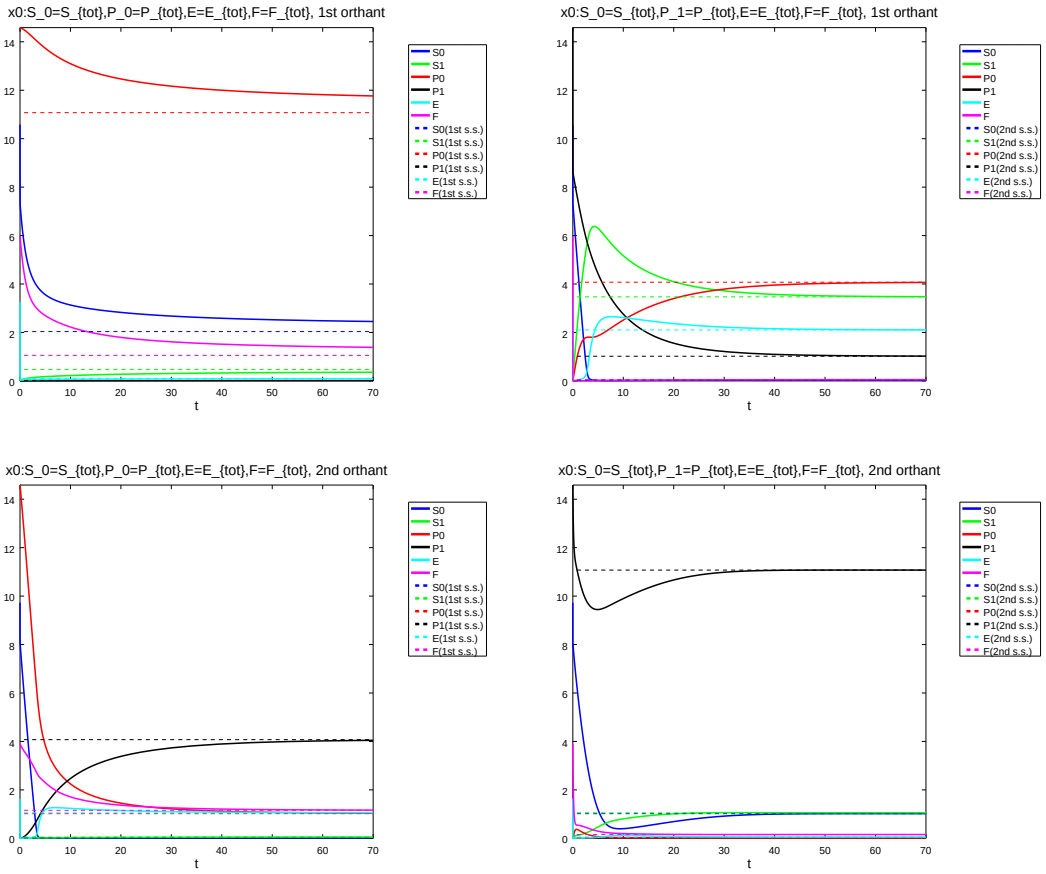


Figure 5. *Witnesses for multistationarity of the phosphorylation cascade with one phosphatase F, with reaction constants and total amounts obtained from two orthants $\mathcal{O}_1, \mathcal{O}_2$ given by Algorithm 1. Upper plots depict the two different steady states constructed from $\mathcal{O}_1$ (dashed lines) along with the simulated trajectories of S_0, S_1, P_0, P_1, E, and F. The initial state on the left is $S_0 = S_{tot}$, $P_0 = P_{tot}$, $E = E_{tot}$, $F = F_{tot}$, and the initial state on the right is $S_0 = S_{tot}$, $P_1 = P_{tot}$, $E = E_{tot}$, $F = F_{tot}$. The lower plots correspond to $\mathcal{O}_2$, with the same initial conditions. We used the function* `ode23s` *from the package* odepkg *version* 0.8.5 *in Octave* [8].

It is important to observe that all the conditions and hypotheses in our paper can be algorithmically checked. In particular, it is possible to devise an algorithm to check whether a given network has a MESSI structure, to prove that a given partition is minimal, to construct the associated digraphs and networks, including the corresponding labels, and to check the hypotheses of all our statements. The construction of the rational parametrization in Theorem 4.1 is also algorithmic. Note also that the sufficient conditions which ensure persistence in Theorem 3.15 are independent of the conditions to have a toric MESSI system or even an s-toric MESSI system, including the criterion for multistationarity given in Theorem 5.4. However, the hypotheses in Proposition 5.6 to ensure the validity of the hypotheses in Theorem 5.4 also imply persistence. This does not mean that multistationarity is related to

persistence, but when there are boundary steady states the hypotheses of Theorem 5.4 should be verified in an ad hoc manner.

Appendix A. Proofs. We assume the reader is familiar with the notion of the Laplacian $\mathcal{L}(G)$ of a digraph G and its main properties. One key observation is that mass-action kinetics associated with a linear digraph G with variables $x = (x_1, \dots, x_s)$ equals $\dot{x} = \mathcal{L}(G)x$. A second key observation is that the fact that the rows of $\mathcal{L}(G)$ add up to zero translates into $\sum_{i=1}^{s} \dot{x}_i = 0$, and so $\sum_{i=1}^{s} x_i$ is a conserved quantity. The last key observation is that when G is strongly connected, the kernel of $\mathcal{L}(G)$ has dimension one and there is a known generator $\rho(G)$ with positive entries described as follows. Recall that an i-tree T of a graph is a spanning tree where the ith vertex is its unique sink (equivalently, the ith is the only vertex of the tree with no edges leaving from it), and we call c^T the product of the labels of all the edges of T. Then, the ith coordinate of $\rho(G)$ equals

$$\rho(G)_i = \sum_{T \text{ an } i-tree} c^T. \tag{14}$$

We refer the reader to [30, 47] for a detailed account.

Conservation relations and persistence. In order to prove Theorem 3.2, we need to first introduce a remark. We call S_1 the stoichiometric subspace of the biochemical network defined by the associated digraph G_1 of a MESSI reaction network G (with stoichiometric subspace S). We denote by $\tilde{S_1} = \{(0, \dots, 0, w) \in \mathbb{R}^{p+n} : w \in S_1 \subseteq \mathbb{R}^n\}$ the lifting of S_1 to $\mathbb{R}^{p+n}$.

Remark A.1. With the previous notation, the following equality of dimensions is an immediate consequence of Lemma 1 in the ESM of [14]:

$$\dim(S) = \dim(S_1) + p. \tag{15}$$

We will also need Lemma 3.13 in the main text.

Proof of Lemma 3.13. Each vertex in the associated digraph G_2 to the digraph G is labeled by only one species. If one species of $\mathscr{S}^{(\alpha)}$ appears on a vertex of G_2, by $(\mathcal{R}_2)$ and $(\mathcal{R}_3)$ and the construction of G_2, all the species in the vertices of the corresponding connected component of G_2 belong to the same $\mathscr{S}^{(\alpha)}$. Moreover, if two core species X_i, X_h in the same subset $\mathscr{S}^{(\alpha)}$ correspond to different connected components of G_2, then for any complex y_{ij} containing X_i and any complex $y_{h\ell}$ containing X_h, the relation $y_{ij} \to_\circ y_{h\ell}$ does *not* hold. It follows that we can refine each subset $\mathscr{S}^{(\alpha)}$ as the disjoint union of the subsets of species in each connected component of G_2 which consists of species in $\mathscr{S}^{(\alpha)}$, and no further refinement is possible if the set of intermediate species is maximal. ∎

We are ready to prove Theorem 3.2. We will mainly adapt the results in [14] (Theorem 2.1) to our setting.

Proof of Theorem 3.2. Given a chemical reaction network G and a partition of the set of species $\mathscr{S}$ that leads to a MESSI system with the given complexes and reactions, consider the mass-action system defined by G_1, with species $X_1, \dots, X_n$. By Theorem 2.1 in [14], the conservation relations in G are in one-to-one correspondence with the conservation relations of

G_1 in an explicit way that we detail below after our hypotheses. Recall that by Remark 3.12, the associated graph G_2 determines the same equations.

Fix $\alpha \geq 1$. As we remarked in the proof of Lemma 3.13, each subset $\mathscr{S}^{(\alpha)}$ coincides with the variables in the vertices of some of the connected components of the associated digraph G_2. Given such a connected component H, let $\mathscr{S}_H^{(\alpha)}$ be its set of vertex labels. As G_2 is a linear digraph, H is also linear, and so the matrix of the associated (linear) system is given by its Laplacian $\mathcal{L}(H)$. Therefore, the sum of its rows equals zero, which means that $\sum_{X_i \in S_H^{(\alpha)}} \dot{x}_i = 0$ and a fortiori $\sum_{X_i \in S^{(\alpha)}} \dot{x}_i = 0$ for the mass-action system defined by G_1. We find now the corresponding linear combination which includes the concentrations of the intermediate species by adapting Lemma 1 in the ESM of [14].

Let $\omega^\alpha \in \{0,1\}^n$ be the characteristic vector of $\mathscr{S}^{(\alpha)}$, so that $\langle \omega^\alpha, \dot{x} \rangle = \sum_{X_i \in S^{(\alpha)}} \dot{x}_i$. For any complex y^j of G_1, we know from $(\mathcal{R}_2)$ and $(\mathcal{R}_3)$ that it has at most one species in $\mathscr{S}^{(\alpha)}$. Then,

$$\omega^\alpha \cdot y^j = \begin{cases} 1 & \text{if there is a species of } \mathscr{S}^{(\alpha)} \text{ in } y^j, \\ 0 & \text{otherwise.} \end{cases}$$

Define the $(p+n)$-vector:

$$\widetilde{\omega}_i^\alpha = \begin{cases} \omega_i^\alpha & \text{for } i = p+1, \ldots, p+n, \\ 1 & \text{if } i \in \mathrm{Int}(\alpha), \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathrm{Int}(\alpha)$ is as in (5). Lemma 1 in the ESM of [14] asserts precisely that the linear form defined by $\widetilde{\omega}^\alpha$ leads to the conservation of the whole network associated with the linear form defined by ω^α on the variables in $\mathscr{S}_1 = \mathscr{S} \setminus \mathscr{S}^{(0)}$. But this linear form is precisely ℓ_α, as we wanted to prove. Since we are assuming that all species participate in at least one reaction and intermediate species satisfy condition $(\mathcal{C})$, we have that $\mathscr{S}^{(0)} = \cup_{\alpha=1}^m \mathscr{S}\mathrm{Int}(\alpha)$. Therefore, all coefficients of the conservation relation $\sum_{\alpha=1}^m \ell_\alpha$ are positive, and we get that any MESSI system is conservative.

To see the second part of the statement, note that $\ell_1, \ldots, \ell_m$ are linearly independent conservation relations and so $\dim(S^\perp) = s - \dim(S) \geq m$. It only remains to prove that, if G has no swaps, then $s - \dim(S) \leq m$. By Remark A.1 it holds that $\dim(S) = \dim(\tilde{S}_1) + p$ because clearly $\dim(S_1) = \dim(\tilde{S}_1)$. It is then enough to show that $\dim(\tilde{S}_1) \geq n - m$. If $X_i + X_j \to X_\ell + X_k$ in G_1, and there are no swaps in G, either $i \in \{\ell, k\}$ or $j \in \{\ell, k\}$. Assume, without loss of generality, that $j = k$. Then $e_\ell - e_i \in S_1$, for e_i is the ith canonical vector of $\mathbb{R}^n$. As $\mathscr{S}$ is minimal, if $X_i, X_\ell \in \mathscr{S}^{(\alpha)}$, necessarily X_i and X_ℓ belong to the same connected component of G_2. Then there is an undirected path between X_i and X_ℓ in G_2. By a telescopic sum, as in the proof of Lemma A.4 below, we have that each vector $e_\ell - e_i \in S_1$ for each $X_i, X_\ell \in \mathscr{S}^{(\alpha)}$. Fix $X_i \in \mathscr{S}^{(\alpha)}$; then for all $\ell \neq i$, $e_\ell - e_i \in S_1$. This gives us $n_\alpha - 1$ linearly independent vectors for each $\alpha \geq 1$, which are in turn linearly independent from the corresponding vectors obtained from each β, $\beta \neq \alpha$, $1 \leq \beta \leq m$ (when $n_\alpha > 1$). Adding over $\alpha \geq 1$, we obtain $n - m$ linearly independent vectors in S. (Notice that if $\mathscr{S}^{(\alpha)}$ is a singleton, $n_\alpha - 1 = 0$.) Therefore, $\dim(S) \geq p + n - m = s - m$, which is what we wanted to prove. The total number of conservation relations in a system is equal to the codimension of the kinetic subspace. If, moreover, the kinetic subspace equals S, then $\dim(S^\perp) = m$, as claimed. ∎

We now focus on the occurrence of boundary steady states. Both Theorem 3.15 and Proposition 4.7 below are based on the proof of Theorem 3.1 in [14] (Theorem 2 in their ESM).

Proof of Theorem 3.15. Assume there is a boundary steady state in some stoichiometric compatibility class that intersects the positive orthant.

Following the proof of Theorem 2 in the ESM of [14], it can be seen that at steady state the concentration of an intermediate species u_k is a nonnegative linear combination of monomials in the concentrations of the core species in the complexes that react via intermediates to it. Then, if there is an intermediate species U_k such that $u_k = 0$ at steady state, there is at least one core species (in a core complex that reacts via intermediates to U_k) that vanishes at steady state. Therefore, if there is a boundary steady state, there is a core species X_i such that $x_i = 0$ at steady state.

By Lemma 3.13, we can refine the given MESSI structure in such a way that subsets of core species are in bijection with the connected components of G_2. In order to avoid unnecessary notation, we will assume in what follows that the partition is minimal. Recall that a vertex in a directed graph has *indegree zero* if it is not the head of any directed edge. Let us define the subsets of indices

$$L_0 = \{\beta \geq 1 : \text{indegree of } \mathscr{S}^{(\beta)} \text{ is } 0\},$$

$$L_k = \{\beta \geq 1 : \text{for any edge } \mathscr{S}^{(\gamma)} \to \mathscr{S}^{(\beta)} \text{ in } G_E \text{ it holds that } \gamma \in L_t, \text{ with } t < k\} \setminus \bigcup_{t=0}^{k-1} L_t, k \geq 1.$$

The main observation that makes the following inductive argument work is that as $\mathscr{S}$ is finite and there are no directed cycles in G_E, there must exist a subset $\mathscr{S}^{(\beta)}$ with $1 \leq \beta \leq m$ such that its indegree in G_E is zero. This means that $L_0 \neq \emptyset$.

Let $\ell \geq 0$ be minimal with the property that there exist $\alpha \in L_\ell$ and a core species $X_i \in \mathscr{S}^{(\alpha)}$ such that $x_i = 0$ at steady state. Denote by H_α the connected component of G_2 with vertices the species in $\mathscr{S}^{(\alpha)}$. Let $\rho(H_\alpha)$ be the generator of the kernel of $\mathcal{L}(H_\alpha)$ as in (14). Its entries are nonnegative sums of terms involving the rate constants τ and concentrations of species in L_j with $j < \ell$. Then, $\rho(H_\alpha)$ has nonzero coordinates since H_α is strongly connected because G_2 is weakly reversible and ℓ is minimal. Moreover, the following equation is satisfied at steady state for any $X_j \in \mathscr{S}^{(\alpha)}$:

$$\rho(H_\alpha)_j \, x_i - \rho(H_\alpha)_i \, x_j = 0. \tag{16}$$

Then the corresponding concentrations x_j vanish at steady state for any $X_j \in \mathscr{S}^{(\alpha)}$. Take $k \in \mathrm{Int}(\alpha)$. The concentration of the intermediate species u_k is a nonnegative linear combination of monomials in the concentrations of the core species that react via intermediates to it. By condition $(\mathcal{C})$ and rule $(\mathcal{R}_3)$, any such monomial contains one variable indexed by a species in $\mathscr{S}^{(\alpha)}$. As $x_j = 0$ for all $j \in \mathscr{S}^{(\alpha)}$, we get that $u_k = 0$. This gives a contradiction by (6) in Theorem 3.2 since C_α is a nonzero constant.

As MESSI systems are conservative, the existence of nonnegative steady states is guaranteed by fixed-point arguments. Indeed, a version of the Brouwer fixed-point theorem ensures that a nonnegative steady state exists in each compatibility class. As the system has no

boundary steady states, we deduce the existence of a positive steady state in each compatibility class, and, in particular, the consistency of the system. ∎

Parametrizing the steady states. We first prove the existence of rational parametrizations under the hypotheses of Theorem 4.1.

Proof of Theorem 4.1. The arguments of the proof are similar to those in the proof of Theorem 3.15. Again, we will assume that the partition is minimal to ease the notation. Recall the sets L_k in that proof and the crucial remark that $L_0 \neq \emptyset$ because the graph G_E has no directed cycles.

For each $\alpha \geq 1$, fix $X_{i_\alpha} \in \mathscr{S}^{(\alpha)}$. Because of the minimality of the partition, any other $X_i \in \mathscr{S}^{(\alpha)}$ lies in the connected component H_α of G_2 containing X_{i_α}. We can then parametrize all the species in $\mathscr{S}^{(\alpha)}$ for $\alpha \in L_k$ in terms of x_{i_α} and the species in L_j for $j < k$, recursively using (16) to write

$$x_i = \frac{\rho(H_\alpha)_i}{\rho(H_\alpha)_{i_\alpha}} x_{i_\alpha}$$

at steady state. Moreover, the concentrations of intermediate species can be rationally written in terms of all $x_{i_\alpha}, \alpha = 1, \ldots, m$ (see Definition 3.8 and Remark 3.9). Thus, $\dim(V_f \cap \mathbb{R}^s_{>0}) = m$. The last equality $\dim S^\perp = m$ in the statement follows from Theorem 3.2 using Remark 3.16. ∎

We show now that the *positive* steady states of s-toric MESSI systems can be described by binomials, and we postpone the proof of the choice of very explicit binomials when any pair of nodes in the same component are connected by a single simple path.

Proof of Proposition 4.7. Following the arguments in [14], we first build a new labeled directed graph $\widehat{G}$ with node set $\mathscr{S}^{(0)} \cup \{*\}$, which consists of collapsing all core complexes into the vertex $*$, and label directed edges that are obtained from hiding the core complexes in the labels. For example, $X_i + X_j \xrightarrow{\kappa} U_k$ becomes $* \xrightarrow{\kappa x_i x_j} U_k$ and $U_k \xrightarrow{\kappa'} X_i + X_j$ becomes $U_k \xrightarrow{\kappa'} *$. This new graph is linear and satisfies that $\dot{\mathbf{u}} = 0$ is equivalent to $\mathcal{L}(\widehat{G})\,\widetilde{\mathbf{u}} = 0$, where $\widetilde{\mathbf{u}} = (u_1, \ldots, u_p, 1)^t$ (this last coordinate stands for "the concentration" of the node $*$). It is important to notice that the graph $\widehat{G}$ is strongly connected by condition $(\mathcal{C})$.

Then, at steady state we obtain that $\widetilde{\mathbf{u}}$ is proportional to the vector $\rho_{\widetilde{G}} = (\rho_1, \ldots, \rho_p, \rho)$ defined in (14), so that $u_k = \rho_k/\rho$ for any $k = 1, \ldots, p$. It is straightforward to check that every $*$-tree involves labels in $\mathbb{Q}[\kappa]$. On the other hand, for every U_k, as by condition $(\mathcal{C}')$ there is a unique core complex $y_{i_k j_k}$ such that $y_{i_k j_k} \to_\circ y_k$, every k-tree involves labels in $\mathbb{Q}[\kappa, x_{i_k} x_{j_k}]$. Moreover, as there must be a path from $*$ to U_k in each k-tree, $x_{i_k} x_{j_k}$ necessarily appears as a label on those trees. Then,

$$u_k = \mu_k\, x_{i_k} x_{j_k}, \quad k = 1, \ldots, p, \tag{17}$$

where

$$\mu_k = \frac{\rho_k}{x_{i_k} x_{j_k}} \frac{1}{\rho} \in \mathbb{Q}(\kappa).$$

∎

Proof of the first part of Theorem 4.8. Let x be a positive steady state and $X_i \neq X_j$ in $\mathscr{S}^{(\alpha)}$ in the same connected component H of G_2. Let $\rho(H)$ be the explicit generator of the

kernel of $\mathcal{L}(H)$ as in (14). Then, as in (16), $\rho(H)_j x_i - \rho(H)_i x_j = 0$. Fix a j-tree T_0. The product of the labels c^{T_0} of all the edges in T_0 is equal to a monomial x^{γ_j} times a polynomial in the rate constants τ. For any other j-tree T, condition $(\mathcal{C}''')$ ensures that $c^T = \mu_T(\tau)\, c^{T_0}$, with $\mu_T \in \mathbb{Q}(\tau)$. It follows that the quotient of the sum $\rho(H)_j$ by x^{γ_j} lies in $\mathbb{Q}(\tau)$ (and also there exists a monomial x^{γ_i} such that $\rho(H)_i/x^{\gamma_i} \in \mathbb{Q}(\tau)$). Call

$$\eta_{ij} = \rho(H)_i x^{\gamma_j}/\rho(H)_j x^{\gamma_i} \in \mathbb{Q}(\tau) \subset \mathbb{Q}(\boldsymbol{\kappa}). \tag{18}$$

Then, $x^{\gamma_j} x_i - \eta_{ij} x^{\gamma_i} x_j = 0$. Combining this with (17), the *positive* steady states can be described by the binomials:

$$u_k - \mu_k x^{\varphi(k)} \text{ for each intermediate species } U_k, \tag{19}$$

$$x^{\gamma_j} x_i - \eta_{ij} x^{\gamma_i} x_j \text{ if } X_i, X_j \text{ lie in the same connected component of } G_2. \tag{20}$$

We can fix one species X_{i_h} in each connected component H of G_2 and consider the binomial equations of the form in (20) where $i = i_h$. There are p further binomial equations in (19). These $p + n - m' = s - m'$ binomial equations cut out the positive steady states. ∎

To prove the second part of Theorem 4.8, we first need a combinatorial lemma.

Lemma A.2. *Assume H is a digraph with the property that there is a unique simple path P_{ij} from any node X_i to any node X_j in the same connected component of H. Then the following hold:*

(i) *For each vertex X_i of H there is only one i-tree, denoted by T_i.*

(ii) *Let $X_i \xrightarrow{\tau x_h} X_j$ be an edge in H. Then, T_i is obtained from T_j by deleting the edge $X_i \xrightarrow{\tau x_h} X_j$ and adding the edge $X_j \xrightarrow{\tau' x_m} X_\ell$, where X_ℓ is such that $X_j \xrightarrow{\tau' x_m} X_\ell$ is in P_{ji}.*

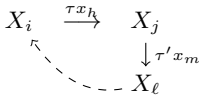
Proof. Proof of (i): Let X_j ($j \neq i$) be in the same connected component of H as X_i. In any i-tree there is an edge leaving from X_j; otherwise X_j would be another sink different from X_i. Moreover, there must be a path from X_j to X_i in any such i-tree. If the path visits some vertex two (or more) times, there would be a cycle in the underlying undirected graph of the tree, which is not possible. Hence, the path is simple. By hypothesis, there is only one choice for this path, and so there is only one i-tree in H.

Proof of (ii): Call T' the new digraph obtained from T_j by deleting the edge $X_i \xrightarrow{\tau x_h} X_j$ and adding the edge $X_j \xrightarrow{\tau' x_m} X_\ell$. T' still visits every vertex of the corresponding connected component of H, and the only vertex from which no arrows leave is X_i. We claim that there are no cycles in T'. In fact, the only possible cycle in T' must involve the new edge from X_j to X_ℓ. Then, there is a directed path in T' (and therefore in H) from X_ℓ to X_j. Moreover, as the paths in T_j are simple, this path from X_ℓ to X_j in T' is simple. But in H there is another simple path $P_{\ell i} \cup \{X_i \to X_j\}$ from X_ℓ to X_j, which is different from the one obtained in T' since the edge $X_i \to X_j$ does not exist in T'. This is a contradiction since by assumption there is only one simple path in H from X_ℓ to X_j. Then, $T' = T_i$. ∎

Proof of the second part of Theorem 4.8. If there is a unique simple path P_{ij} from each X_i to each X_j in the same connected component of G_2, and $X_i \xrightarrow{\tau x_h} X_j$ is in G_2, the binomial

in (19) involves the edges on T_i and the edges on T_j. But, from Lemma A.2, T_i and T_j only differ in the edges $X_i \xrightarrow{\tau x_h} X_j$ and $X_j \xrightarrow{\tau' x_m} X_\ell$, where X_ℓ is such that $X_j \xrightarrow{\tau' x_m} X_\ell$ is in P_{ji}. Then, after taking out a monomial, the following binomials define the positive steady states:

$$u_k - \mu_k x^{\varphi(k)} \text{ for each intermediate species } U_k,$$
$$\tau x_h x_i - \tau' x_m x_j \text{ if } X_i \xrightarrow{\tau x_h} X_j \text{ in } G_2^\circ \text{ and } X_j \xrightarrow{\tau' x_m} X_\ell \text{ is in } P_{ji}.$$



This completes the proof. ∎

Toric MESSI systems and multistationarity. We will prove Theorem 5.4 by adapting Proposition 3.9 and Corollary 2.15 in [32] and Theorem 5.5 in [34] to our setting. We recall that a chemical reaction system has the capacity for multistationarity if there exists a choice of rate constants such that there are two or more *positive* steady states in one stoichiometric compatibility class $(x^0 + S) \cap \mathbb{R}^s_{\geq 0}$ for some initial state $x^0 \in \mathbb{R}^s_{\geq 0}$ (and it is monostationary otherwise).

Remark A.3. Consider a toric MESSI system whose positive steady states can be described by binomial equations of the form $x^{y'} - \eta x^y = 0$. Equivalently, the positive steady states of the toric MESSI system can be described by the monomial equations $x^{y'-y} = \eta$, where we consider Laurent monomials. We construct now a matrix B whose columns form a basis of the subspace T generated by these difference vectors $y' - y$, and also the monomial map $x \mapsto x^B$, where $(x^B)_j = x^{B_j} = x_1^{B_{1j}} \cdot \ldots \cdot x_s^{B_{sj}}$, for each column B_j of B. Then x^* is a positive steady state of the system if and only if $x^{*B} = \tilde{\eta}$ for an appropriate vector $\tilde{\eta}$. Thus, the system is monostationary *for any choice of rate constants* if and only if the monomial map $x \mapsto x^B$ is injective on each stoichiometric compatibility class $(x^0 + S) \cap \mathbb{R}^s_{>0}$ for every $x^0 \in \mathbb{R}^s_{>0}$.

Proof of Theorem 5.4. Under the hypotheses in the statement, we want to prove the equivalence of the assertions:

(i) The associated MESSI system is monostationary.
(ii) The signs sets $\Sigma, \Sigma^\perp, \Sigma_\perp, \Sigma_\perp^\perp$ are unmixed.
(iii) For all orthants $\mathcal{O} \in \{-1, 0, 1\}^s, \mathcal{O} \neq \mathbf{0}$, either $S \cap \mathcal{O} = \emptyset$ or $T^\perp \cap \mathcal{O} = \emptyset$.

We first prove (i) ⇔ (ii) by adapting the results in [32]. We will see that (i) and (ii) are both equivalent to

$$\{\text{sign}(v) : v \in \ker(B^t)\} \cap \{\text{sign}(v) : v \in S\} = \{0\},$$

where $(\text{sign}(v))_i = \text{sign}(v_i)$ for $i = 1, \ldots, s$. This is also equivalent by the definition of $T^\perp$ to

$$\{\text{sign}(v) : v \in T^\perp\} \cap \{\text{sign}(v) : v \in S\} = \{0\}. \tag{21}$$

By Remark A.3, (i) is equivalent to the injectivity of the map $x \mapsto x^B$ on each stoichiometric compatibility class $(x^0 + S) \cap \mathbb{R}^s_{>0}$. We deduce from Proposition 3.9 in [32] that (i) is equivalent to (21). Previously, in Corollary 2.15 the authors had proved that (21) is in turn equivalent to asking that for all $J \subseteq [s]$, $\#J = s - d = \text{rank}(B) = \text{rank}(M)$, $\det(B_J)\det(M_J)$ is either zero or has the same sign as all other nonzero products, and, moreover, at least one such product is nonzero. In other words, (21) is equivalent to the set Σ being unmixed. By Lemma 5.3,

this is equivalent to (ii). To finish the proof, we just need to show that (21) $\Leftrightarrow$ (iii), but this is straightforward. ∎

We now prove Theorem 5.8, and we postpone the proof of Proposition 5.6, which needs an ancillary lemma.

Proof of Theorem 5.8. By Theorem 5.4, if the system is not monostationary, we know that there exists an orthant $\mathcal{O} \in \{-1, 0, 1\}^s, \mathcal{O} \neq \mathbf{0}$, such that $S \cap \mathcal{O} \neq \emptyset$ and $T^\perp \cap \mathcal{O} \neq \emptyset$. Then, there exist $\mathbf{w} \in S, \mathbf{v} \in T^\perp$ such that $\operatorname{sign}(\mathbf{w}) = \operatorname{sign}(\mathbf{v})$. Inspired by Theorem 5.5 in [34], for any index i not in the support of $\mathbf{v}$, we choose any positive real number h_i and we define positive vectors x^1 and x^2 as follows:

$$\begin{aligned} \left(x_i^1\right)_{i=1,\dots,s} &= \begin{cases} \frac{w_i}{e^{v_i}-1} & \text{if } v_i \neq 0, \\ h_i & \text{otherwise,} \end{cases} \\ x^2 &= \operatorname{diag}(e^{\mathbf{v}})\, x^1, \end{aligned}$$

where "e^x" for a vector $x \in \mathbb{R}^s_{>0}$ denotes the vector $(e^{x_1}, e^{x_2}, \dots, e^{x_s}) \in \mathbb{R}^s$ and $\operatorname{diag}(x)$ denotes the diagonal matrix whose diagonal is the vector x.

As the system is consistent, there exists a positive vector λ such that $\sum_{y \to y'} \lambda_{yy'}(y' - y) = 0$. For any edge $y \to y'$, take the (positive) rate constant

$$k_{yy'} = \lambda_{yy'} (x^1)^{-y},$$

which defines a positive vector $\boldsymbol{\kappa}$ satisfying

$$f(x^1, \boldsymbol{\kappa}) = \sum_{y \to y'} \kappa_{yy'} \, (x^1)^y \, (y' - y) = 0.$$

Then, x^1 is a positive steady state of the system for these reaction rate constants $\boldsymbol{\kappa}$. As the system is a toric MESSI system, x^1 is a solution of the binomial equations that describe the positive steady states. Call $\boldsymbol{\eta} := (x^1)^B$. Then, x is a positive steady state of the system if and only if $x^B = \boldsymbol{\eta}$. It can be checked that $((x^2)^B)_j = e^{\langle \mathbf{v}, B_j \rangle} (x^1)^{B_j}$, and, as $\mathbf{v} \in T^\perp$, we have $(x^1)^B = (x^2)^B = \boldsymbol{\eta}$. Therefore, x^2 is also a positive steady state of the system. Moreover, $x^2 - x^1 = \mathbf{w} \in S$, and so x^1 and x^2 belong to the same stoichiometric compatibility class. ∎

Recall the definitions of S_1 and $\tilde{S}_1$ before Remark A.1.

Lemma A.4. *Assume that condition* $(\mathcal{C}')$ *in Definition* 4.3 *holds, and consider the vectors*

$$v_k = y_k - y_{i_k j_k}. \tag{22}$$

Then, $S = \tilde{S}_1 \oplus \langle v_1, \dots, v_p \rangle$.

Proof. It is clear, from the definitions of $\tilde{S}_1$ and the vectors v_k, that $\tilde{S}_1 \cap \langle v_1, \dots, v_p \rangle = \{0\}$ (as no intermediate complex appears in the reactions of G_1). Moreover, the vectors v_k are linearly independent, and therefore $\dim(\langle v_1, \dots, v_p \rangle) = p$. By Remark A.1, we know that $\dim(S) = \dim(S_1) + p = \dim(\tilde{S}_1) + p$. Thus, we only need to show now that $S \supseteq \tilde{S}_1 \oplus \langle v_1, \dots, v_p \rangle$.

For simplicity, we will assume that all core complexes consist of two species, but it is easy to adapt the proof for the case where the core complexes consist of only one species. We first notice that $v_k \in S$ for all k. In fact, if $X_{i_k} + X_{j_k} \to_\circ U_k$, there exist $U_{k_1}, \dots, U_{k_t}$ intermediates such that the chain of reactions $X_{i_k} + X_{j_k} \to U_{k_1} \to \cdots \to U_{k_t} \to U_k$ is in G. Therefore, from the telescopic sum $y_k - y_{i_k j_k} = (y_k - y_{k_t}) + (y_{k_t} - y_{k_t - 1}) + \cdots + (y_{k_2} - y_{k_1}) + (y_{k_1} - y_{i_k j_k})$, we see that $v_k \in S$, as we wanted to prove. Given $X_i + X_j \to X_\ell + X_m$ in G_1, there exist intermediates $U_{k_1}, \dots, U_{k_t}$ such that the chain of reactions $X_i + X_j \to U_{k_1} \to \cdots \to U_{k_t} \to X_\ell + X_m$ is in G. As above, from a telescopic sum we deduce that $y_{\ell m} - y_{ij} \in S$. Hence, $\tilde{S}_1 \subseteq S$ and $\tilde{S}_1 \oplus \langle v_1, \dots, v_p \rangle \subseteq S$. ∎

Proof of Proposition 5.6. By Theorem 4.1, we know that $\operatorname{rank}(M^\perp) = m$. We show now that $\operatorname{rank}(B) = s - m$, or equivalently that $\dim(T) = s - m = p + n - m$. From (17) we see that the vectors v_k defined in (22) live in T for all $1 \le k \le p$ (recall that y_k denotes the vector corresponding to the monomolecular complex U_k). This implies that $\langle v_1, \dots, v_p \rangle \subseteq T$. As none of the exponents determined by (19) involves any variable u_i, it is enough to find $n - m$ linearly independent vectors in T that have support in the last n coordinates.

Call T_x the projection $\pi_x(T)$ of T onto the last n coordinates corresponding to $x_1, \dots, x_n$. We need to prove then that $\dim(T_x) = n - m$. For each $\alpha \ge 1$, fix $i_\alpha \in \mathscr{S}^{(\alpha)}$ and for each $X_j \in \mathscr{S}^{(\alpha)}$, $j \ne i_\alpha$, call $z_{i_\alpha j} = (\gamma_j + e_{i_\alpha}) - (\gamma_{i_\alpha} + e_j)$, the vector in $\mathbb{R}^n$ deduced from the exponents of the binomials in (19). Denote by T_α the linear subspace with generators $\{z_{i_\alpha j}\}_{j \ne i_\alpha}$. We claim that $\dim(T_\alpha) = n_\alpha - 1$ for any $\alpha \ge 1$ and that $T_x = T_1 \oplus T_2 \oplus \cdots \oplus T_m$.

To prove these claims, we need to recall the proof of Theorem 3.15. We consider again the subsets $L_0, L_1, \dots$, and we assume that $\alpha \in L_k$. Then, as remarked in the last paragraph of that proof, it holds that the connected component G_2^α with vertices in $\mathscr{S}^{(\alpha)}$ (ensured by Lemma 3.13 by our hypothesis of minimality of the partition) has labels in $\mathbb{Q}[\tau, x_\beta : \beta \in L_t, t < k]$. This implies that the jth coordinate of the vector $z_{i_\alpha h}$ equals -1 if $h = j$ and 0 otherwise. So the vectors $\{z_{i_\alpha j}\}_{j \ne i_\alpha}$ are linearly independent, that is, $\dim(T_\alpha) = n_\alpha - 1$, and by a similar argument we deduce that the sum is direct. Therefore, $\dim(T_x) = \sum_{\alpha=1}^m (n_\alpha - 1) = n - m$, as wanted. ∎

Algorithm. Step 1 in the algorithm follows directly from Theorem 5.4. Step 7 follows from [4, 10, 34] and Theorem 5.8. Theorem 4.8 explains how to find a matrix B for an s-toric MESSI system. The intermediate steps follow from the following considerations. Given a matrix A, every vector in $\operatorname{rowspan}(A)$ is a conformal sum of circuits. (We refer the reader to [31, 37, 44].) Moreover, the circuits of a matrix $A \in \mathbb{R}^{d \times s}$ of rank d are found in the following way. For $J \subseteq [s]$ with $\#J = d - 1$, define $r_J \in \operatorname{rowspan}(A)$ as the vector $r_{J,\ell} = (-1)^{\mu(\ell, J)} \det(A_{J \cup \{\ell\}})$, where $\mu(\ell, J)$ is the sign of the permutation of $J \cup \{\ell\}$ which takes ℓ followed by the ordered elements of J to the ordered elements of $J \cup \{\ell\}$ for all $\ell \in \{0, \dots, s\}$. The following lemma is straightforward and well known.

Lemma A.5. *Let $A \in \mathbb{R}^{d \times s}$ be a matrix of rank d and $J \subseteq [s]$ such that $\#J = d - 1$ and $\operatorname{rank}(A_J) = d - 1$. Then r_J is a circuit of A. Moreover, up to a multiplicative constant, these are all the circuits of A (possibly repeated).*

Acknowledgment. We would like to thank the referee for providing useful comments.

REFERENCES

[1] D. Angeli, P. De Leenher, and E. Sontag, *A Petri net approach to the study of persistence in chemical reaction networks*, Math. Biosci., 210 (2007), pp. 598–618.

[2] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. Ziegler, *Oriented Matroids*, 2nd ed., Encyclopedia Math. Appl. 46, Cambridge University Press, Cambridge, UK, 1999.

[3] S. Catozzi. J. P. Di-Bella, A. Ventura, and J.-A. Sepulchre, *Signaling cascades transmit information downstream and upstream but unlikely simultaneously*, BMC Syst. Biol., 16 (2016), pp. 1–20.

[4] C. Conradi, D. Flockerzi, and J. Raisch, *Multistationarity in the activation of a MAPK: Parametrizing the relevant region in parameter space*, Math. Biosci., 211 (2008), pp. 105–131.

[5] C. Conradi, and A. Shiu, *A global convergence result for processive multisite phosphorylation systems*, Bull. Math. Biol., 77 (2015), pp. 126–155.

[6] R. J. Davis, *Signal transduction by the JNK group of MAP kinases*, Cell, 103 (2000), pp. 239–252.

[7] R. J. Deshaies and J. E. Ferrell, *Multisite phosphorylation and the countdown to S phase*, Cell, 107 (2001), pp. 819–822.

[8] J. W. Eaton, D. Bateman, S. Hauberg, and R. Wehbring, *GNU Octave Version* 4.0.0 *Manual: A High-Level Interactive Language for Numerical Computations*, http://www.gnu.org/software/octave/doc/interpreter/, 2015.

[9] M. Feinberg, *Lectures on Chemical Reaction Networks*, University of Wisconsin, available at https://crnt.osu.edu/LecturesOnReactionNetworks, 1979.

[10] M. Feinberg, *Multiple steady states for chemical reaction networks of deficiency one*, Arch. Ration. Mech. Anal., 132 (1995), pp. 371–406.

[11] M. Feinberg et al., *Chemical Reaction Network Toolbox*, available at https://crnt.osu.edu/CRNTWin.

[12] M. Feinberg and F. Horn, *Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspaces*, Arch. Ration. Mech. Anal., 66 (1977), pp. 83–97.

[13] E. Feliu and C. Wiuf, *Enzyme-sharing as a cause of multi-stationarity in signalling systems*, J. R. Soc. Interface, 9 (2012), pp. 1224–1232.

[14] E. Feliu and C. Wiuf, *Simplifying biochemical models with intermediate species*, J. R. Soc. Interface, 10 (2013), 20130484.

[15] E. Feliu and C. Wiuf, *Variable elimination in post-translational modification reaction networks with mass-action kinetics*, J. Math. Biol., 66 (2013), pp. 281–310.

[16] G. Gnacadja, *Reachability, persistence, and constructive chemical reaction networks (part* II*): A formalism for species composition in chemical reaction network theory and application to persistence*, J. Math. Chem. 49 (2011), pp. 2137–2157.

[17] G. Gnacadja, *Reachability, persistence, and constructive chemical reaction networks (part* III*): A mathematical formalism for binary enzymatic networks and application to persistence*, J. Math. Chem., 49 (2011), pp. 2158–2176.

[18] N. Hermann-Kleiter and G. Baier, *NFAT pulls the strings during CD4+ T helper cell effector functions*, Blood, 115 (2010), pp. 2989–2997.

[19] P. G. Hogan, L. Chen, J. Nardone, and A. Rao, *Transcriptional regulation by calcium, calcineurin, and NFAT*, Genes Dev., 17 (2003), pp. 2205–2232.

[20] J. J. Hornberg, B. Binder, F. J. Bruggeman, B. Schoeber, R. Heinrich, and H. V. Westerhoff, *Control of MAPK signalling: From complexity to what really matters*, Oncogene 24, (2005), pp. 5533–5542.

[21] W. Hsing and T. J. Silhavy, *Function of conserved histidine-243 in phosphatase activity of EnvZ, the sensor for porin osmoregulation in Escherichia coli*, J. Bacteriol., 179 (1997), pp. 3729–3735.

[22] C.-Y. F. Huang and J. E. Ferrell, *Ultrasensitivity in the mitogen-activated protein kinase cascade*, Proc. Natl. Acad. Sci. USA, 93 (1996), pp. 10078–10083.

[23] M. M. Igo, A. J. Ninfa, J. B. Stock, and T. J. Silhavy, *Phosphorylation and dephosphorylation of a bacterial transcriptional activator by a transmembrane receptor*, Genes Dev. 3 (1989), pp. 1725–1734.

[24] B. N. Kholodenko, *Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades*, Eur. J. Biochem., 267 (2000), pp. 1583–1588.

[25] V. B. Kothamachu, E. Feliu, L. Cardelli, and O. S. Soyer, *Unlimited multistability and Boolean logic in microbial signalling*, J. R. Soc. Interface, 12 (2015), 20150234.
[26] J. M Kyriakis and J. Avruch, *Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation*, Physiol. Rev., 81 (2001), pp. 807–869.
[27] F. Macian, *NFAT proteins: Key regulators of T-cell development and function*, Nat. Rev. Immunol., 5 (2005), pp. 472–484.
[28] M. Marcondes de Freitas, E. Feliu, and C. Wiuf, *Intermediates, catalysts, persistence, and boundary steady states*, J. Math. Biol., 74 (2017), pp. 887–932.
[29] N. I. Markevich, J. B. Hoek, and B. N. Kholodenko, *Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades*, J. Cell Biol. 164 (2004), pp. 353–359.
[30] I. Mirzaev and J. Gunawardena, *Laplacian dynamics on general graphs*, Bull. Math. Biol., 75 (2013), pp. 2118–2149.
[31] S. Müller and G. Regensburger, *Elementary vectors and conformal sums in polyhedral geometry and their relevance for metabolic pathway analysis*, Front Genet., 7 (2016), p. 90.
[32] S. Müller, E. Feliu, G. Regensburger, C. Conradi, A. Shiu, and A. Dickenstein, *Sign conditions for injectivity of generalized polynomial maps with applications to chemical reaction networks and real algebraic geometry*, Found. Comput. Math., 16 (2016), pp. 69–97.
[33] G. Pearson, F. Robinson, T. Beers Gibson, B. E. Xu, M. Karandikar, K. Berman, and M. H. Cobb, *Mitogen-activated protein (MAP) kinase pathways: Regulation and physiological functions*, Endocr. Rev., 22 (2001), pp. 153–183.
[34] M. Pérez Millán, A. Dickenstein, A. Shiu, and C. Conradi, *Chemical reaction systems with toric steady states*, Bull. Math. Biol., 74 (2012), pp. 1027–1065.
[35] L. Pratt and T. J. Silhavy, *Porin regulation of* Escherichia coli, in Two-Component Signal Transduction, J. A. Hoch and T. J. Silhavy, eds., American Society for Microbiology, Washington, DC, 1995, pp. 105–127.
[36] J. Richter-Gebert and G. M. Ziegler, *Oriented matroids*, in Handbook of Discrete and Computational Geometry, J. E. Goodman and J. O'Rourke, eds., Chapman and Hall/CRC, Boca Raton, FL, 1997, pp. 111–132.
[37] R. T. Rockafellar, *The elementary vectors of a subspace of* R^N, in Combinatorial Mathematics and Its Applications (Proc. Conf., Univ. North Carolina, Chapel Hill, NC, 1967), University of North Carolina Press, Chapel Hill, NC, 1969, pp. 104–127.
[38] M. Sáez, C. Wiuf, and E. Feliu, *Graphical reduction of reaction networks by linear elimination of species*, J. Math. Biol., 74 (2017), pp. 195–237.
[39] H. J. Schaeffer and M. J. Weber, *Mitogen-activated protein kinases: Specific messages from ubiquitous messengers*, Mol. Cell Biol., 19 (1999), pp. 2435–2444.
[40] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Muller, *Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors*, Nat. Biotechnol., 20 (2002), pp. 370–375.
[41] Y. D. Shaul and R. Seger, *The MEK/ERK cascade: From signaling specificity to diverse functions*, Biochim. Biophys. Acta., 1773 (2007), pp. 1213–1226.
[42] G. Shinar and M. Feinberg, *Structural sources of robustness in biochemical reaction networks*, Science, 327 (2010), pp. 1389–1391.
[43] A. M. Stock, V. L. Robinson, and P. N. Goudreau, *Two-component signal transduction*, Annu. Rev. Biochem., 69 (2000), pp. 183–215.
[44] B. Sturmfels, *Gröbner Bases and Convex Polytopes*, University Lecture Ser. 8, AMS, Providence, RI, 1995.
[45] M. Thomson and J. Gunawardena, *The rational parametrisation theorem for multisite post-translational modification systems*, J. Theoret. Biol., 261 (2009), pp. 626–636.
[46] A. Turjanski, J. Vaqué, and J. Gutkind, *MAP kinases and the control of nuclear events*, Oncogene, 26 (2007), pp. 3240–3253.
[47] W. T. Tutte, *The dissection of equilateral triangles into equilateral triangles*, Proc. Cambridge Philos. Soc., 44 (1948), pp. 463–482.
[48] A. Vol'pert and S. Hudjaev, *Analysis in Classes of Discontinuous Functions and Equations of Mathematical Physics*, Mech. Anal. 8, Springer, Amsterdam, 1985.

[49] L. Wang and E. Sontag, *On the number of steady states in a multiple futile cycle*, J. Math. Biol., 57 (2008), pp. 29–52.
[50] C. Widmann, S. Gibson, M. B. Jarpe, and G. L. Johnson, *Mitogen-activated protein kinase conservation of a three-kinase module from yeast to human*, Physiol. Rev., 79 (1999), pp. 143–180.
[51] T. Zarubin and J. Han, *Activation and signaling of the p38 MAP kinase pathway*, Cell Res., 15 (2005), pp. 11–18.
[52] Y. Zhu, L. Qin, T. Yoshida, and M. Inouye, *Phosphatase activity of histidine kinase EnvZ without kinase catalytic domain*, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 7808–7813.