



Intermediates, catalysts, persistence, and boundary steady states

Michael Marcondes de Freitas¹ · Elisenda Feliu¹ ·
Carsten Wiuf¹ 

Received: 15 September 2015 / Revised: 20 July 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract For dynamical systems arising from chemical reaction networks, persistence is the property that each species concentration remains positively bounded away from zero, as long as species concentrations were all positive in the beginning. We describe two graphical procedures for simplifying reaction networks without breaking known necessary or sufficient conditions for persistence, by iteratively removing so-called intermediates and catalysts from the network. The procedures are easy to apply and, in many cases, lead to highly simplified network structures, such as monomolecular networks. For specific classes of reaction networks, we show that these conditions for persistence are equivalent to one another. Furthermore, they can also be characterized by easily checkable strong connectivity properties of a related graph. In particular, this is the case for (conservative) monomolecular networks, as well as cascades of a large class of post-translational modification systems (of which the MAPK cascade and the n -site futile cycle are prominent examples). Since one of the aforementioned sufficient conditions for persistence precludes the existence of boundary steady states, our method also provides a graphical tool to check for that.

Keywords Reaction network theory · Model reduction · Persistence · Boundary steady states · Intermediates · Catalysts · Post-translational modification

Mathematics Subject Classification 34C99 · 80A30 · 92C42

✉ Carsten Wiuf
wiuf@math.ku.dk

¹ Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark

1 Introduction

Since the seminal works of Horn, Jackson and Feinberg in the 70s (Feinberg 1980; Gunawardena 2003; Horn and Jackson 1972; and references therein), chemical reaction network theory (CRNT) has provided a fruitful framework to study the dynamical systems describing how the concentrations of the involved chemical species evolve over time. Of great interest has been the long-term behavior of these systems, for example, whether they may exhibit oscillatory behavior (Feinberg 1987), local asymptotic stability (Anderson 2008, 2011; Feinberg 1987; Gopalkrishnan et al. 2014; Sontag 2001), or persistence (Angeli et al. 2007, 2011; Craciun et al. 2013; Deshpande and Gopalkrishnan 2014; Gnacadja 2011a, b; Gopalkrishnan et al. 2014).

The mathematical concept of persistence models the property that every species concentration remains above a certain threshold, as long as there were positive amounts of each species in the beginning. Besides its intrinsic relevance to the applied sciences, most notably in population biology (Smith and Thieme 2011), the concept of persistence has also drawn attention in the context of CRNT on account of its connection with the global attractor conjecture (Gopalkrishnan et al. 2014).

It can be difficult to determine if the solutions to a system of ordinary differential equations are persistent case by case. A recent contribution was given by Angeli et al. (2007), who provided two checkable conditions, one sufficient, and the other one necessary, for the persistence of conservative reaction networks. Their sufficient conditions were further developed and relaxed by Deshpande and Gopalkrishnan (2014). These criteria work under fairly general assumptions on the reaction kinetics. But perhaps unsurprisingly, reaction networks become more difficult to analyze the larger they are, often times exponentially so (Cordone et al. 2005). Thus, criteria for persistence in terms of a simplified “skeleton” of the given network are desirable. More importantly, simplified versions retaining the properties of interest of the original network may also give insight into the underlying biological mechanism, suggesting what might be the leading causes of the presence (or absence) of said properties. For example, for the class of post-translational modification (PTM) systems of Thomson and Gunawardena (2009), or cascades of PTM systems, persistence can be characterized in terms of strong connectedness of the underlying substrate network at each layer of the cascade, as we shall see.

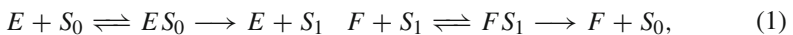
That is the motivation for our model simplification approach to study persistence. In this work we describe a process through which one may simplify a reaction network by iteratively removing “intermediates” (Feliu and Wiuf 2013), and/or “catalysts”. Intuitively speaking, an intermediate is a transient species appearing in the middle of a chain of reactions. Catalysts, on the other hand, are reactants which remain unchanged in every reaction, except possibly for interactions exclusively with other catalysts. Our main contribution is to show that the removal of intermediates and/or catalysts does not break the conditions for persistence given in Angeli et al. (2007) and Deshpande and Gopalkrishnan (2014). Our main results in this work may be informally stated as follows.

Theorem 1 *The conditions for persistence of reaction networks in Angeli et al. (2007) and Deshpande and Gopalkrishnan (2014) are invariant under the removal of intermediate species.*

Theorem 2 *The conditions for persistence of reaction networks in Angeli et al. (2007) and Deshpande and Gopalkrishnan (2014) are invariant under the removal of catalysts.*

Theorem 3 *The same minimally simplified reaction network is always obtained by iteratively removing intermediates and catalysts until none can be found, independently of the order in which they are removed.*

As shown by various examples throughout this work taken from the systems biology literature, reaction networks naturally exhibit many intermediate complexes and catalysts. So, their removal will often reduce dramatically the size of the network, facilitating its inspection for persistence. To illustrate this, consider a simple one-site phosphorylation process, which can be modeled by the reaction network



where S_0 , S_1 represent, respectively, the dephosphorylated and phosphorylated forms of a substrate, E acts as a kinase, F acts as a phosphatase, and ES_0 and FS_1 are intermediate protein complexes in the phosphorylation/dephosphorylation mechanism. Using our results, one may show that necessary or sufficient conditions for persistence for (1) are a consequence of the same necessary or sufficient conditions for its much simpler underlying substrate model,



For monomolecular models such as (2), the necessary or sufficient conditions for persistence are actually equivalent, and, furthermore, characterized by the strong connectedness of each connected component. In fact, (1) will turn out to be a special case of PTM system.

We emphasize that iteratively removing intermediates and catalysts—and, if eventually obtaining a monomolecular network, then checking it for strong connectedness of its connected components—is essentially a graphical procedure.

This paper is organized as follows. In Sect. 2, we review the basic formalism of reaction networks. We present the conditions for persistence in Angeli et al. (2007) and Deshpande and Gopalkrishnan (2014) in the form we shall use in this work, and discuss their relationship with boundary steady states. A few trivial but notable examples we shall refer to several times throughout the work are given, and persistence is characterized for monomolecular networks in terms of strong connectedness of its connected components. In Sect. 3, we define the concepts of intermediates and catalysts. We describe the networks obtained from their removal, and formally state our main results (Theorems 1, 2, 3), concerning how these operations do not break the aforementioned conditions for persistence. Some biologically relevant examples are presented in Sect. 4, the most important of which being cascades of a class of PTM systems. In Sect. 5 we return to our main results, giving the details of the proofs. A short appendix with some auxiliary technical results is presented at the end.

2 Reaction networks

In what follows we denote the set of nonnegative real (respectively, integer) numbers by $\mathbb{R}_{\geq 0}$ (respectively, $\mathbb{Z}_{\geq 0}$), and denote the set of strictly positive real (respectively, integer) numbers by $\mathbb{R}_{> 0}$ (respectively, $\mathbb{Z}_{> 0}$). We denote the boundary of the nonnegative orthant by $\partial\mathbb{R}_{\geq 0}^n$. Given $x \in \mathbb{R}^n$, for some $n \in \mathbb{Z}_{> 0}$, we write $x \geq 0$ to mean that $x \in \mathbb{R}_{\geq 0}^n$, that is, each coordinate of x is nonnegative. We write $x > 0$ to mean that $x \geq 0$, and at least one coordinate of x is positive, and write $x \gg 0$ to mean that $x \in \mathbb{R}_{> 0}^n$, in other words, each coordinate of x is strictly positive. For any finite set X , the notation $|X|$ represents the number of elements of X . Given $n \in \mathbb{Z}_{> 0}$, we write $[n] := \{1, \dots, n\}$. By convention $[0] := \emptyset$.

2.1 Basic formalism

In this work we take the approach of defining reaction networks from their reaction graphs. Thus, a *reaction network* is an ordered triple $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ in which \mathcal{S} is a finite, possibly empty set, \mathcal{C} is a finite subset of $\mathbb{R}_{\geq 0}^n$, where $n := |\mathcal{S}|$, and $(\mathcal{C}, \mathcal{R})$ is a digraph with no self-loops. The set \mathcal{S} is called the *species* set of the reaction network. Its elements are tacitly assumed to be ordered in some fixed way, say,

$$\mathcal{S} = \{S_1, \dots, S_n\}.$$

We identify the elements $(\alpha_1, \dots, \alpha_n)$ of \mathcal{C} , called the *complexes* of the reaction network, with the formal linear combinations of species

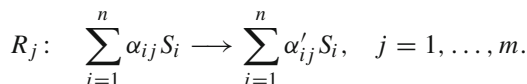
$$\alpha_1 S_1 + \dots + \alpha_n S_n.$$

The digraph $(\mathcal{C}, \mathcal{R})$ is called the *reaction graph* of G , and its edges are referred to as the *reactions* of the network. We further assume that each complex takes part in at least one reaction, and that each species is part of at least one complex. Formally, this means that each vertex of $(\mathcal{C}, \mathcal{R})$ has indegree or outdegree at least one, and that for each $i \in [n]$, there exists $(\alpha_1, \dots, \alpha_n) \in \mathcal{C}$ such that $\alpha_i > 0$. It follows that $\mathcal{S} = \emptyset \Leftrightarrow \mathcal{C} = \emptyset \Leftrightarrow \mathcal{R} = \emptyset$, in which case the network is referred to as the *empty reaction network*.

The reactions are also tacitly assumed to be ordered in some fixed way, say,

$$\mathcal{R} = \{R_1, \dots, R_m\},$$

where $m := |\mathcal{R}|$. We often express the reaction $R_j = ((\alpha_{1j}, \dots, \alpha_{nj}), (\alpha'_{1j}, \dots, \alpha'_{nj}))$ as



The complex on the lefthand side is referred to as the *reactant* of the reaction, while the complex on the righthand side is referred to as its *product*. The species S_i such that $\alpha_{ij} > 0$ are, accordingly, called the *reactants* of R_j , while the species S_i for which $\alpha'_{ij} > 0$ are called the *products* of the reaction.

A *reaction path* in G is a directed path in the digraph $(\mathcal{C}, \mathcal{R})$, that is, a sequence of reactions

$$y_0 \rightarrow y_1 \rightarrow \cdots \rightarrow y_{k-1} \rightarrow y_k$$

such that $y_{j-1} \rightarrow y_j \in \mathcal{R}$ for all $j \in [k]$ and all complexes are different. Similarly, an *undirected reaction path* in G is a path in the undirected graph underlying $(\mathcal{C}, \mathcal{R})$. In this case, we write $y_0 - y_1 - \cdots - y_k$, where each ‘ $-$ ’ can either be ‘ \leftarrow ’ or ‘ \rightarrow ’ in $(\mathcal{C}, \mathcal{R})$. By abuse of terminology, we refer to the connected components of the reaction graph $(\mathcal{C}, \mathcal{R})$ as the connected components of G .

With the above notation, we define the $n \times m$ matrix N ,

$$N_{ij} := \alpha'_{ij} - \alpha_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m,$$

known as the *stoichiometric matrix* of the network. The column-space of N , which is a subset of \mathbb{R}^n , is called the *stoichiometric subspace* of G , and denoted by Γ . The sets $(s_0 + \Gamma) \cap \mathbb{R}_{\geq 0}^n$, $s_0 \in \mathbb{R}_{\geq 0}^n$, are called the *stoichiometric compatibility classes* of G . Let

$$Q_j := \{i \in [n] \mid \alpha_{ij} > 0\}, \quad j = 1, \dots, m,$$

be the subset of indices corresponding to the reactants of R_j .

The system of differential equations governing the evolution of the concentrations of the species of the network is given by

$$\frac{ds}{dt} = Nr(s(t)), t \geq 0, s \geq 0, \tag{3}$$

where $r: \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}_{\geq 0}^m$ is a vector-valued function modeling the kinetic rates of each reaction as functions of the reactant species, henceforth referred to simply as the *vector of reaction rates*. We shall assume throughout this work that the vector of reaction rates satisfies the following hypotheses:

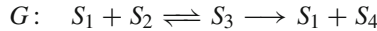
- (r1) $r = (r_1, \dots, r_m): \mathcal{O} \rightarrow \mathbb{R}^m$ is continuously differentiable on a neighborhood \mathcal{O} of $\mathbb{R}_{\geq 0}^n$, and $r(s) \geq 0$ for every $s \geq 0$.
- (r2) For each $j \in [m]$, and for each $s = (s_1, \dots, s_n) \in \mathbb{R}_{\geq 0}^n$,

$$r_j(s) = 0 \Leftrightarrow s_i = 0 \quad \text{for some } i \in Q_j.$$

- (r3) The flow of (3) is forward-complete; in other words, for any initial state, the (unique) maximal solution of the corresponding initial value problem in (3) is defined for all $t \geq 0$.

We note that (r1)–(r2) are satisfied under the most common kinetic assumptions in the literature, namely, mass-action, or more general power-law kinetics, Michaelis–Menten kinetics, or Hill kinetics, as well as combinations of these (Angeli et al. 2010, pages 585–586). We also note that it follows from (r1) and (r2) that the non-negative and positive orthants, $\mathbb{R}_{\geq 0}^n$ and $\mathbb{R}_{> 0}^n$, are forward invariant for the flow of (3) [see, for instance, (Sontag 2001, Section VII) or (Amann 1990, Section 16)].

We will often give a reaction network by simply listing all the reactions in the network. When we do so, the sets of species and complexes will be tacitly implied. For instance,



is the reaction network $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ obtained by setting

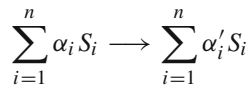
$$\mathcal{S} := \{S_1, S_2, S_3, S_4\}, \quad \mathcal{C} := \{S_1 + S_2, S_3, S_1 + S_4\}$$

and

$$\mathcal{R} := \{S_1 + S_2 \rightarrow S_3, S_3 \rightarrow S_1 + S_3, S_3 \rightarrow S_1 + S_4\}$$

in the formalism above.

Definition 1 (*Implied subnetworks*) Let $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network, and $\mathcal{E} \subseteq \mathcal{S}$ be a subset of species. We define the *subnetwork implied by \mathcal{E}* as the network $G_{\mathcal{E}} = (\mathcal{S}_{\mathcal{E}}, \mathcal{C}_{\mathcal{E}}, \mathcal{R}_{\mathcal{E}})$ consisting of reactions of G which involve exclusively species in \mathcal{E} . More precisely, $\mathcal{R}_{\mathcal{E}} \subseteq \mathcal{R}$ is the subset of reactions



such that $\alpha_i = \alpha'_i = 0$ for every $i \in [n]$ such that $S_i \notin \mathcal{E}$. We then define $\mathcal{C}_{\mathcal{E}} \subseteq \mathcal{C}$ to be the subset of complexes that appear as reactant or product of some reaction in $\mathcal{R}_{\mathcal{E}}$. Finally, $\mathcal{S}_{\mathcal{E}} \subseteq \mathcal{S}$ is defined as the subset of species which are part of some complex in $\mathcal{C}_{\mathcal{E}}$.

Although it is always true that $\mathcal{S}_{\mathcal{E}} \subseteq \mathcal{E}$, it may be the case that $\mathcal{S}_{\mathcal{E}} \neq \mathcal{E}$. To see this, consider the reaction network G with $\mathcal{R} = \{S_1 + S_2 \rightarrow S_3 + S_4, S_4 \rightarrow S_2\}$ and set $\mathcal{E} := \{S_1, S_2, S_4\}$. Then $G_{\mathcal{E}}$ consists of the reaction $S_4 \rightarrow S_2$. In particular, $\mathcal{S}_{\mathcal{E}} = \{S_2, S_4\} \subsetneq \{S_1, S_2, S_4\} = \mathcal{E}$.

2.2 Siphons, P- and T-semiflows, drainable sets and self-replicable sets

A few more concepts pertaining to reaction networks are needed. Some of the terminology below is adapted from Petri net theory. See Angeli et al. (2007) for the context. But since no results from Petri net theory itself are needed, we chose to define these

concepts as directly pertaining to their respective reaction networks, rather than the Petri nets associated with them.

Definition 2 (Siphons) A nonempty subset of species $\Sigma \subseteq \mathcal{S}$ is called a *siphon* if every reaction which has a product in Σ also has a reactant in Σ . A siphon is said to be *minimal* if it does not properly contain any other siphon.

Example 1 (Single phosphorylation mechanism) The minimal siphons of the single phosphorylation mechanism from the Introduction (1) are $\{E, ES_0\}$, $\{F, FS_1\}$, and $\{S_0, S_1, ES_0, FS_1\}$. □

Remark 1 Let $y \rightarrow y_1 \rightarrow \dots \rightarrow y_k \rightarrow y'$ be a reaction path in a reaction network G , and suppose Σ is a siphon containing some species S' that is part of y' . Then each of the complexes y, y_1, \dots, y_k must have at least one of its species in Σ .

Given a vector $\omega = (\omega_1, \dots, \omega_n) \in \mathbb{R}_{\geq 0}^n$ associated with the species set \mathcal{S} of a reaction network $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$, its *support* is defined to be the subset of species $\text{supp } \omega := \{S_i \in \mathcal{S} \mid \omega_i > 0\}$. Similarly, given a vector $v = (v_1, \dots, v_m) \in \mathbb{R}_{\geq 0}^m$ associated with the reaction set \mathcal{R} of G , its *support* is defined to be the subset of reactions $\text{supp } v := \{R_j \in \mathcal{R} \mid v_j > 0\}$. Although we use the same notation in both cases, it will be clear from the context whether the underlying vector is associated with the species or the reaction set.

Definition 3 (P- and T-semiflows) A *P-semiflow* or *positive conservation law* of a reaction network is any nonzero vector $\omega \in \mathbb{R}_{\geq 0}^n$ such that $\omega^T N = 0$. We say that a reaction network is *conservative* if it has a strictly positive P-semiflow $\omega \gg 0$, that is, $\text{supp } \omega = \mathcal{S}$. A *T-semiflow* of a reaction network is any nonzero vector $v \in \mathbb{R}_{\geq 0}^m$ such that $Nv = 0$. We say that a reaction network is *consistent* if it has a strictly positive T-semiflow $v \gg 0$, that is, if $\text{supp } v = \mathcal{R}$.

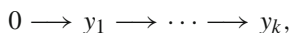
Definition 4 (Siphon/P-semiflow property) We say that a reaction network has the *siphon/P-semiflow property* if every siphon contains the support of a P-semiflow.

Nonempty sets of species not containing the support of a P-semiflow are also known in the literature as *critical* (Deshpande and Gopalkrishnan 2014). So, a reaction network has the siphon/P-semiflow property if, and only if every siphon is noncritical.

Note that, since every siphon is either itself minimal, or else contains a minimal siphon, we need only check whether every minimal siphon contains the support of a P-semiflow. We give a couple more trivial examples. Besides further illustrating the scope of the concepts just introduced, they will be used several times in the analysis of more elaborate examples further down.

Example 2 (Empty networks) Our formalism allows for reaction networks to be empty. Any such network is vacuously conservative, consistent, and also has the siphon/P-semiflow property.

Example 3 (Inflows) Consider a reaction network $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$. If one can find a reaction path in G of the form



then by Remark 1, none of the species that are a part of any of the complexes y_1, \dots, y_k belongs to a siphon because no species is part of the complex 0. This observation may drastically reduce the number of species one is concerned about in checking the siphon/P-semiflow property.

In particular, if G is such that $0 \rightarrow S \in \mathcal{R}$ for each $S \in \mathcal{S}$, then G has no siphons. In this case, G has vacuously the siphon/P-semiflow property.

We next introduce the concepts of drainable and self-replicable siphons. In [Deshpande and Gopalkrishnan \(2014\)](#), Definition 3.1(2–3), these concepts were defined in terms of “ G -reaction pathways”. We show in Proposition 6 in Sect. 1 in the appendix that both definitions are equivalent. This equivalence is already implicitly used in the proofs of the results in [Deshpande and Gopalkrishnan \(2014\)](#).

Definition 5 (*Drainable and self-replicable sets*) Let $G = (S, \mathcal{C}, \mathcal{R})$ be a reaction network. A nonempty subset of species $\Sigma \subseteq S$ is said to be *drainable* if there exists a sequence of reactions $y_1 \rightarrow y'_1, \dots, y_k \rightarrow y'_k \in \mathcal{R}$ such that

$$\left(\sum_{j=1}^k (y'_j - y_j) \right)_i < 0, \quad \forall i \in [n] : S_i \in \Sigma.$$

If there exists one such a sequence of reactions such that

$$\left(\sum_{j=1}^k (y'_j - y_j) \right)_i > 0, \quad \forall i \in [n] : S_i \in \Sigma,$$

then Σ is said to be *self-replicable*. In either case, the reactions need not be pairwise distinct.

We summarize some properties of critical, drainable and self-replicable siphons we will need further down.

Proposition 1 *Let $G = (S, \mathcal{C}, \mathcal{R})$ be a reaction network, and $\Sigma \subseteq S$ a nonempty subset. Then,*

- (i) *if Σ is drainable or self-replicable, then it is critical; and*
- (ii) *if Σ is a minimal critical siphon, then it is drainable or self-replicable.*

Proof See [Deshpande and Gopalkrishnan \(2014\)](#), Theorem 5.3. □

Corollary 1 *A reaction network G has the siphon/P-semiflow property if and only if G does not have any drainable or self-replicable siphons.*

2.3 Persistence and boundary steady states

The existence of drainable siphons and the siphon/P-semiflow property are linked to persistence and the existence of boundary steady states. The connection is made precise

in this subsection, where we compile results from [Deshpande and Gopalkrishnan \(2014\)](#), [Angeli et al. \(2007\)](#) and [Shiu and Sturfels \(2010\)](#).

Intuitively, persistence (of a reaction network) is the property that no species concentration goes below a certain threshold as the system evolves, as long as they were initially all positive. This threshold may depend on the initial conditions though. In order to formulate this more precisely, let $\sigma : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}_{\geq 0}^n$ be the semiflow of (3). In other words, for each initial state $s_0 \in \mathbb{R}_{\geq 0}^n$, $\sigma(\cdot, s_0) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}^n$ is the unique, solution of (3). The solution is unique in virtue of (r1), and defined for all $t \geq 0$ on account of (r3).

Definition 6 (Persistence) A reaction network (3) is said to be *persistent* if

$$\liminf_{t \rightarrow \infty} \sigma_i(t, s_0) > 0, \quad \forall i \in [n], \tag{4}$$

for every initial state $s_0 \gg 0$.

We also introduce a weaker notion of persistence. First, recall that, for each $s_0 \geq 0$, the ω -limit set of s_0 is the set

$$\omega(s_0) := \bigcap_{\tau \geq 0} \overline{\bigcup_{t \geq \tau} \{\sigma(t, s_0)\}}.$$

Note that $s \in \omega(s_0)$ if, and only if there exists a sequence $(t_k)_{k \in \mathbb{N}}$ going to infinity in $\mathbb{R}_{\geq 0}$ such that

$$\lim_{k \rightarrow \infty} \sigma(t_k, s_0) = s.$$

Definition 7 (Bounded-persistence) A reaction network (3) is said to be *bounded-persistent* if $\omega(s_0) \cap \partial \mathbb{R}_{\geq 0}^n = \emptyset$ for each $s_0 \gg 0$.

A *steady state* of a reaction network G is any point $s_0 \geq 0$ such that $Nr(s_0) = 0$.

Definition 8 (Boundary steady state) A *boundary steady state* is any point $s_0 \in \partial \mathbb{R}_{\geq 0}^n$ such that $Nr(s_0) = 0$, in other words, any steady state that lies on the boundary.

In the following proposition we collect relationships among persistence, bounded-persistence, consistence, drainable siphons and the siphon/P-semiflow property. Details of the proof are given in Sect. 1 in the appendix.

Proposition 2 Consider a reaction network G .

- (i) If G is persistent, then it is bounded-persistent.
- (ii) If G is conservative and bounded-persistent, then it is persistent.
- (iii) If G is conservative and persistent, then it is consistent.
- (iv) If G has no drainable siphons, then it is bounded-persistent.
- (v) If G has the siphon/P-semiflow property, then the stoichiometric compatibility classes of G that are not entirely contained in the boundary do not contain any boundary steady states.

Conservative networks are a special case of dissipative networks (Definition 15), for which bounded-persistence is also equivalent to persistence. These will be discussed in Sect. 4.2.

Remark 2 In view of Corollary 1, if a reaction network has the siphon/P-semiflow property, then it has no drainable siphons and therefore it is bounded-persistent by Proposition 2(iv).

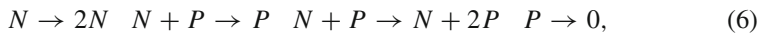
The next example shows that not having any drainable siphons is not in general a necessary condition for the bounded-persistence of reaction networks.

Example 4 (Lotka–Volterra predator–prey model) The Lotka–Volterra equations,

$$\frac{dN}{dt} = N(t)(a - bP(t)) \quad \frac{dP}{dt} = P(t)(cN(t) - d), \quad (5)$$

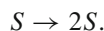
where a, b, c, d are positive parameters, model the population sizes at time $t \geq 0$ of a predator species, $P(t)$, and its prey, $N(t)$, under the assumptions that $N(t)$ grows exponentially in the absence of predators, $P(t)$ decays exponentially in the absence of prey, and that both the growth rate of $P(t)$ and the depletion rate of $N(t)$ on account of predation are directly proportional to the population counts $N(t)$ and $P(t)$.

Equations (5) can be derived as (3) from the reaction network



under mass-action kinetics (see, for instance, Gunawardena 2003 for an account of mass-action kinetics). Solutions of (5) are known to be uniformly bounded away from zero. In fact, they are periodic (James 2002, Section 3.1). In particular, (6) is bounded-persistent. However, the minimal siphons of (6) are $\{N\}$ and $\{P\}$, both of which are drainable on account of reactions $N + P \rightarrow P$ and $P \rightarrow 0$, respectively.

Example 5 (Non-drainable siphons and boundary steady states) The absence of drainable siphons does not in general preclude boundary steady states in stoichiometric compatibility classes that meet the interior of the positive orthant. For example, consider the reaction network with the reaction



This reaction network has one stoichiometric compatibility class, namely $\mathbb{R}_{\geq 0}$, and a boundary steady state. However, it has no drainable siphons.

2.4 Monomolecular networks

Iterating the simplification procedures discussed in this work will often result in what we shall refer to as monomolecular networks. Intuitively, these are reaction networks in which each reactant or product consists of at most a single species. The precise definition is given below in Definition 9. For conservative monomolecular networks,

the necessary and sufficient conditions for persistence given in Proposition 2(iii) and (iv) are actually equivalent, and characterized by the strong connectedness of the connected components of the network (Proposition 3).

Definition 9 (*Monomolecular networks*) A reaction network $G = (S, C, \mathcal{R})$ is said to be *monomolecular* if, for each $y \in C$, either $y = 0$ or $y = S_i$ for some $i \in [n]$. In this case, we identify the nonzero ‘complexes’ of G with the corresponding ‘species.’

Proposition 3 Consider a monomolecular reaction network $G = (S, C, \mathcal{R})$ and the following seven properties.

- (i) G is consistent.
- (ii) Each connected component of G is strongly connected.
- (iii) G has the siphon/ P -semiflow property.
- (iv) G has no drainable siphons.
- (v) G is bounded-persistent.
- (vi) G is persistent.

Then the following implications hold:

$$(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (v) \Leftarrow (vi).$$

If the reaction network is conservative, then the six properties are equivalent.

Proof Proposition 2 and Corollary 1 guarantee that $(iii) \Rightarrow (iv) \Rightarrow (v) \Leftarrow (vi)$ for any reaction network. Furthermore, $(v) \Rightarrow (vi) \Rightarrow (i)$ for conservative networks, also by Proposition 2.

Thus, it is sufficient to show that $(i) \Rightarrow (ii)$ and $(ii) \Rightarrow (iii)$ for arbitrary monomolecular networks.

$(i) \Rightarrow (ii)$. Since G is consistent by hypothesis, there exists a strictly positive T -semiflow $v \in \mathbb{R}_{>0}^m$, that is $Nv = 0$. We prove below that v is in the kernel of the incidence matrix of the reaction graph of G . Strong connectness of each connected component of G then follows, for example, from (Feinberg 1995, Remark 6.1.1).

The incidence matrix C_G of the reaction graph (C, \mathcal{R}) has m columns and one row for each complex. The entries of the j -th column, corresponding to a reaction $R_j = y \rightarrow y'$, are all zero except for the entry corresponding to y , which is -1 and the entry corresponding to y' , which is 1 . If $C = \{S_1, \dots, S_n\}$, then $C_G = N$ by definition, hence $C_G v = 0$. If $C = \{S_1, \dots, S_n, 0\}$, then the first n rows of C_G agree with N . Since the sum of the rows of C_G is zero, we have

$$(C_G)_{n+1} \cdot v = - \sum_{i=1}^n (C_G)_i \cdot v = - \sum_{i=1}^n N_i \cdot v = 0.$$

We conclude once again that $C_G v = 0$.

$(ii) \Rightarrow (iii)$. Let $(C_1, \mathcal{R}_1), \dots, (C_J, \mathcal{R}_J)$ be the connected components of (C, \mathcal{R}) , and denote the canonical basis of \mathbb{R}^n by $\{e_1, \dots, e_n\}$. Let $j \in [J]$. We have two possibilities.

If $0 \notin \mathcal{C}_j$, then any siphon of G containing some species $S' \in \mathcal{C}_j$ contains \mathcal{C}_j . Indeed, for any other $S \in \mathcal{C}_j$, there exists a reaction path connecting S to S' . Thus, S belongs to any siphon containing S' . Furthermore,

$$\sum_{i: S_i \in \mathcal{C}_j} \mathbf{e}_i$$

is a P-semiflow of G . This follows from the fact that, for each reaction $S \rightarrow S' \in \mathcal{R}_j$, the column of N corresponding to $S \rightarrow S'$ has exactly two nonzero entries, namely, a 1 in the row corresponding to S' , and a -1 in the row corresponding to S .

If $0 \in \mathcal{C}_j$, then by strong connectedness, there is a reaction path from 0 to any species $S \in \mathcal{C}_j$. By Example 3, S cannot belong to any siphon of G and thus \mathcal{C}_j contains no siphons. We conclude that every siphon of G contains the support of a P-semiflow. \square

The property that every connected component of the reaction graph is strongly connected is also known in the literature as *weak reversibility* (see Gunawardena 2003, Definition 6.1). Thus, Proposition 3, as well as other results further down, could well have been stated in these terms. In this work, the property of weak reversibility only comes up in the context of monomolecular networks. Thus, we chose to use the more informative, explicit description in terms of strong connectivity of the connected components.

Example 6 (Persistence without conservativity) The hypothesis that the network is conservative in Proposition 3 is not superfluous for the full equivalence of the six statements. For example, the reaction network $0 \rightarrow A$ with, say, mass-action kinetics is persistent but not consistent and the implication (vii) \Rightarrow (i) does not hold.

3 Intermediates and catalysts

In this section we define the concepts of intermediate and catalyst of a reaction network. We also describe the reaction networks that are obtained from their removal. After establishing these concepts and underlying terminology in Sects. 3.1 and 3.2, we state our main results in Sect. 3.3.

3.1 Intermediates

Consider a reaction network $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$. Let \mathcal{Y} be a nonempty subset of \mathcal{S} , and write

$$\mathcal{Y} = \{Y_1, \dots, Y_p\}, \quad \text{and} \quad \mathcal{S} \setminus \mathcal{Y} = \{S_1, \dots, S_q\}.$$

Consider the following two properties.

- (I1) Each complex $y \in \mathcal{C}$ is either of the form $y = \alpha_1 S_1 + \dots + \alpha_q S_q$ for some $\alpha_1, \dots, \alpha_q \geq 0$, or of the form $y = Y_i$ for some $i \in [p]$. In particular, we identify the ‘complexes’ and ‘species’ Y_1, \dots, Y_p . (See also Definition 9).

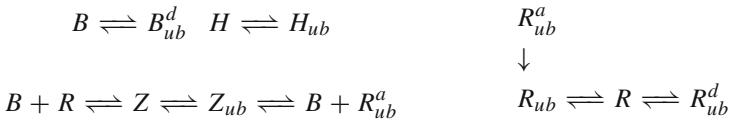
(I2) For each $Y \in \mathcal{Y}$, there exist $y, y' \in \mathcal{C} \setminus \mathcal{Y}$, and reaction paths from y to Y and from Y to y' such that all their non-endpoints are in \mathcal{Y} .

If (I1) and (I2) hold, then we may construct a reaction network $G^* = (\mathcal{S}^*, \mathcal{C}^*, \mathcal{R}^*)$ as follows. We define $\mathcal{R}^* := \mathcal{R}_c^* \cup \mathcal{R}_Y^*$, where $\mathcal{R}_c^* \subseteq \mathcal{R}$ is the set of reactions $y \rightarrow y' \in \mathcal{R}$ such that $y, y' \in \mathcal{C} \setminus \mathcal{Y}$, and \mathcal{R}_Y^* is constructed as the set of reactions $y \rightarrow y' \in \mathcal{R}$ such that $y, y' \in \mathcal{C} \setminus \mathcal{Y}$, $y \neq y'$, and there is a reaction path in G connecting y to y' such that all their non-endpoints are in \mathcal{Y} . We set \mathcal{C}^* to be the set of reactant and product complexes in the reactions in \mathcal{R}^* , and we set \mathcal{S}^* to be the set of species that are part of some complex in \mathcal{C}^* . Note that \mathcal{S}^* does not always coincide with $\mathcal{S} \setminus \mathcal{Y}$, as illustrated in Example 7 below. In the above description, we think of the reactant and product sides of a reaction $y \rightarrow y' \in \mathcal{R}^*$ as the formal linear combinations of participating species alluded to before.

Definition 10 (*Intermediates*) Let $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network and \mathcal{Y} be a nonempty subset of \mathcal{S} . We call \mathcal{Y} a *set of intermediate species* of G , if (I1) and (I2) hold. In this case, the reaction network $G^* = (\mathcal{S}^*, \mathcal{C}^*, \mathcal{R}^*)$ defined as above is called the *reduction of G by the removal of the set of intermediates \mathcal{Y}* . The elements of \mathcal{Y} are then referred to as the *intermediate species* of G .

For brevity, we will often write simply intermediates instead of intermediate species.

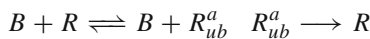
Example 7 (A ubiquitination model) Consider the following reaction network model for Ring1B/Bmi1 ubiquitination (Nguyen et al. 2011).



Note that

$$\mathcal{Y} := \{B_{ub}^d, H, R_{ub}, R_{ub}^d, Z, Z_{ub}\}$$

is a set of intermediate species of the network. This network can be reduced to



by removing these intermediates and collapsing the paths in which they appear, as described above.

We emphasize that \mathcal{S}^* does not always coincide with $\mathcal{S} \setminus \mathcal{Y}$. In this example, H_{ub} is in $\mathcal{S} \setminus \mathcal{Y}$, but not in \mathcal{S}^* . We also note that the same network G^* may arise from removing a different set of intermediates. For instance, in this example, we could have set H_{ub} as an intermediate in place of H .

Removing one intermediate at a time

In the proofs of some of our results concerning the removal of intermediates, we use induction on the number of intermediates removed. Thus, a discussion of how the intermediates in a set of intermediates may be iteratively removed, one at a time, is warranted.

Let $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network, and suppose $\mathcal{Y} = \{Y_1, \dots, Y_p\}$ is a set of intermediates of G . Set $G_p := G$. It follows directly from the definition that any nonempty subset of \mathcal{Y} is a set of intermediates of G . In particular, $\{Y_p\}$ is a set of intermediates of G_p . Let G_{p-1} be the reduction of G_p by the removal of the set of intermediates $\{Y_p\}$. Now $\{Y_1, \dots, Y_{p-1}\}$ is a set of intermediates of G_{p-1} . In particular, $\{Y_{p-1}\}$ is a set of intermediates of G_{p-1} . We define G_{p-2} to be the reduction of G_{p-1} by the removal of the set of intermediates $\{Y_{p-1}\}$. Iterating this process p times, we obtain a sequence G_p, \dots, G_1, G_0 such that $G_p = G$, and G_{i-1} is the reduction of G_i by the removal of the set of intermediates $\{Y_i\}, i = p, \dots, 1$.

Lemma 1 *If G, \mathcal{Y} , and G_p, \dots, G_1, G_0 are like in the above construction, and G^* is the reduction of G by the removal of the set of intermediates \mathcal{Y} , then $G_0 = G^*$.*

Proof We use induction on p . The claim is trivial for $p = 1$. So, suppose it has been proven to be true for the removal of up to p intermediates, for some $p \geq 1$. Let $\mathcal{Y} = \{Y_1, \dots, Y_p, Y_{p+1}\}$ be a set of intermediates of G . As noted above, $\{Y_2, \dots, Y_{p+1}\}$ is a set of intermediates of G . Let G_1^* be the reduction of G obtained by their removal. By the induction hypothesis, $G_1^* = G_1$, and so $\mathcal{R}_1^* = \mathcal{R}_1$. We want to show that $\mathcal{R}_0 = \mathcal{R}^*$.

$\mathcal{R}^* \subseteq \mathcal{R}_0$. Let $y \rightarrow y'$ be any reaction in \mathcal{R}^* . If $y \rightarrow y' \in \mathcal{R}$, then $y \rightarrow y' \in \mathcal{R}_1^*$, and so $y \rightarrow y' \in \mathcal{R}_0$. So, suppose $y \rightarrow y' \notin \mathcal{R}$. Then there exist $Y^{(1)}, \dots, Y^{(\ell)} \in \mathcal{Y}$ such that

$$y \longrightarrow Y^{(1)} \longrightarrow \dots \longrightarrow Y^{(\ell)} \longrightarrow y'$$

is a reaction path in G . If $Y^{(1)}, \dots, Y^{(\ell)} \in \{Y_2, \dots, Y_{p+1}\}$, then $y \rightarrow y' \in \mathcal{R}_1^*$, and so $y \rightarrow y' \in \mathcal{R}_0$ like in the previous case. Otherwise, we have $Y_1 = Y^{(i)}$ for some $i \in [\ell]$. But now

$$y \longrightarrow Y_1 \longrightarrow y' \tag{7}$$

is a reaction path in G_1^* , and so $y \rightarrow y' \in \mathcal{R}_0$ once again.

$\mathcal{R}_0 \subseteq \mathcal{R}^*$. Let $y \rightarrow y'$ be any reaction in \mathcal{R}_0 . If $y \rightarrow y' \in \mathcal{R}_1^*$, then there exists a reaction path connecting y to y' in G such that all its non-endpoints belong to $\{Y_2, \dots, Y_{p+1}\}$. In this case, $y \rightarrow y' \in \mathcal{R}^*$. If $y \rightarrow y' \notin \mathcal{R}_1^*$, then (7) is a reaction path in G_1^* . In this case there are reaction paths in G connecting y to Y_1 and Y_1 to y' , all non-endpoints of which belong to $\{Y_2, \dots, Y_{p+1}\}$. Concatenating these two reaction paths we obtain a reaction path in G connecting y and y' such that all its non-endpoints belong to \mathcal{Y} . It follows once again that $y \rightarrow y' \in \mathcal{R}^*$. □

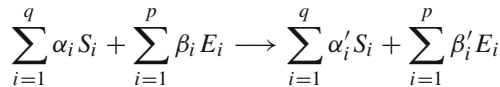
3.2 Catalysts

Consider a reaction network $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$. Let \mathcal{E} be a nonempty subset of \mathcal{S} , and write

$$\mathcal{E} = \{E_1, \dots, E_p\}, \quad \text{and} \quad \mathcal{S} \setminus \mathcal{E} = \{S_1, \dots, S_q\}.$$

Consider the following two properties.

(C1) For each reaction

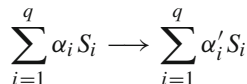


in \mathcal{R} , we have

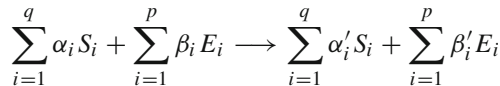
$$\sum_{i=1}^p \beta_i E_i = \sum_{i=1}^p \beta'_i E_i \quad \text{or} \quad \alpha_1 = \alpha'_1 = \dots = \alpha_q = \alpha'_q = 0.$$

(C2) The subnetwork $G_{\mathcal{E}} = (\mathcal{S}_{\mathcal{E}}, \mathcal{C}_{\mathcal{E}}, \mathcal{R}_{\mathcal{E}})$ implied by \mathcal{E} (refer to Definition 1) has no drainable or self-replicable siphons (equivalently, has the siphon/P-semiflow property).

If (C1) and (C2) hold, then we may construct a reaction network $G^* = (\mathcal{S}^*, \mathcal{C}^*, \mathcal{R}^*)$ as follows. We set \mathcal{R}^* to be the set of reactions



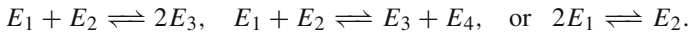
such that



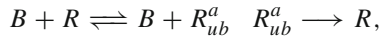
belongs to \mathcal{R} , and $\alpha_{i_0} > 0$ or $\alpha'_{i_0} > 0$ for some $i_0 \in [q]$. We then set \mathcal{C}^* to be the set of reactants and products in these reactions, and set \mathcal{S}^* to be the set of species that are part of some complex in \mathcal{C}^* . Contrary to what happened with intermediates, \mathcal{S}^* always agrees with $\mathcal{S} \setminus \mathcal{E}$.

Definition 11 (*Catalysts*) Let $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network and \mathcal{E} be a nonempty subset of \mathcal{S} . We call \mathcal{E} a *set of catalysts* of G if (C1) and (C2) hold. In this case, the reaction network $G^* = (\mathcal{S}^*, \mathcal{C}^*, \mathcal{R}^*)$ defined as above is called a *reduction of G by the removal of the set of catalysts \mathcal{E}* . The elements of \mathcal{E} are then referred to as the *catalysts of G* .

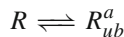
Typically (C2) is checked via Proposition 3, by showing that $G_{\mathcal{E}}$ is a monomolecular network and each connected component of its reaction graph is strongly connected, as we shall see in some of the examples in the next section. However, the theory allows for catalysts to interact in more complex, yet still biologically meaningful ways, for instance, in reversible reactions of the forms



Example 8 (A ubiquitination model (continued)) Consider the network



obtained from the ubiquitination model in Example 7 after intermediates were removed. Note that $\mathcal{E} := \{B\}$ is a set of catalysts. Thus, this network can be further reduced to



by removing B and projecting the reactions as described above.

Note that B is not a catalyst of the original ubiquitination model in Example 7. In realistic biochemical models, it is often the case that catalysts in the sense of Definition 11 only emerge after some intermediates are removed.

3.3 Main results

We are now ready to precisely restate Theorems 1, 2 and 3 in the introduction, and consider a few examples. The proofs will be given in Sect. 5.

Theorem 1 (Removal of intermediates) *Suppose a reaction network G^* is obtained from a reaction network G by the removal of a set of intermediates. Then,*

- (i) G has no drainable (respectively, self-replicable) siphons if, and only if G^* has no drainable (respectively, self-replicable) siphons;
- (ii) G has the siphon/ P -semiflow property if, and only if G^* has the siphon/ P -semiflow property;
- (iii) G is consistent if, and only if G^* is consistent; and
- (iv) if G is conservative, then G^* is conservative; conversely, if G^* is conservative and $0 \notin \mathcal{C}$, then G is also conservative.

Theorem 2 (Removal of catalysts) *Suppose a reaction network G^* is obtained from a reaction network G by the removal of a set of catalysts \mathcal{E} . Then,*

- (i) G has no drainable (respectively, self-replicable) siphons if, and only if G^* has no drainable (respectively, self-replicable) siphons;
- (ii) G has the siphon/ P -semiflow property if, and only if G^* has the siphon/ P -semiflow property;

- (iii) if G is consistent, then G^* is consistent; conversely, if G^* is consistent and $G_{\mathcal{E}}$ is conservative, then G is consistent; and
- (iv) if G is conservative, then G^* is conservative; conversely, if G^* is conservative and $G_{\mathcal{E}}$ is conservative, then G is conservative.

Combining Theorems 1 and 2 with Proposition 3 we obtain the following corollary.

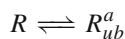
Corollary 2 *Suppose a monomolecular reaction network G^* is obtained by iteratively removing sets of intermediates and catalysts from a reaction network G . If each of the connected components of G^* is strongly connected, then G is bounded-persistent and has no boundary steady states in any stoichiometric compatibility class that is not contained in the boundary of the positive orthant. Furthermore, if G is conservative, then G is persistent and consistent.*

Definition 12 (*Primitive networks*) A reaction network $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ is said to be *primitive* (with respect to the removal of catalysts or intermediates) if no subset of \mathcal{S} is a set of catalysts or intermediates of G . If iteratively removing sets of intermediates and catalysts of a reaction network G results in a primitive reaction network G^* , then we refer to G^* as a *primitive reduction* of G .

Theorem 3 (*Uniqueness of the primitive reduction*) *Let G be a reaction network, and suppose G_1^* and G_2^* are primitive reductions of G . Then $G_1^* = G_2^*$.*

Observe that Theorem 3 is more than just a theoretical curiosity. As noted in Example 7, choosing a set of intermediates or catalysts to remove is not something that can always be done in a unique way at each stage of the simplification process. Thus, knowing that one would always obtain the same minimally simplified reaction network regardless of the order in which catalysts and intermediates are removed has also practical relevance.

Example 9 (*A ubiquitination model (concluded)*) The network

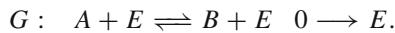


is a strongly connected monomolecular network. By Corollary 2, so long as the reaction rates of the ubiquitination model from Example 7 satisfy our hypotheses, we conclude that the network is persistent.

We emphasize that the procedures of removal of intermediates and catalysts carried out in Examples 7 and 8, as well as the analysis of the emerging underlying substrate network for strong connectedness in Example 9, are essentially graphical. More specifically, one need not do any calculations with the stoichiometric matrix or the reaction rates.

In Theorem 2(iii), the hypothesis that $G_{\mathcal{E}}$ be conservative is not superfluous. If that is not the case, then it might happen that G^* is consistent and G is not, as shown in Example 10 below. However, if G is consistent, then G^* is consistent regardless of whether $G_{\mathcal{E}}$ is conservative or not, as shown later in Lemma 12.

Example 10 (Non-conservative $G_{\mathcal{E}}$) Consider the reaction network



The singleton $\mathcal{E} := \{E\}$ is a set of catalysts of G , the removal of which yields the reaction network



By Proposition 3, G^* is consistent. The stoichiometric matrix of G is

$$N = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Any T-semiflow of G must have its third coordinate equal to zero, so G is not consistent.

4 Examples

We shall apply Theorems 1 and 2 to two main classes of reaction networks. In Sect. 4.1, we give necessary and sufficient conditions for cascades of a class of post-translational modification (PTM) systems to be persistent. The reaction network (1) in the introduction, as well as the ubiquitination model discussed in Examples 7, 8 and 9, will turn out to be special cases of PTM systems. In Sect. 4.2, we argue that a nonconservative reaction network may still be shown to be persistent when it has no drainable siphons as long as it can be also shown to be dissipative. Finally, in Sect. 4.3, we apply our results to a model of Wnt signaling that focuses on shuttling and degradation (Shiu and Sturmfels 2010).

4.1 Cascades of PTM systems

In this subsection, we study the persistence of cascades of a class of PTM systems. Combining Theorems 1 and 2 with Propositions 2, 3 and Corollary 1, we will achieve necessary and sufficient conditions for persistence of cascades of PTM systems in terms of strong connectedness of the connected components of the underlying substrate network of each layer.

4.1.1 PTM systems

Consider a reaction network $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$. Let

$$\mathcal{S} = \text{Enz} \cup \text{Sub} \cup \text{Int}$$

be a partition of the species set. Thus, Enz, Sub, and Int are pairwise disjoint. Consider the following properties.

(M1) The reactions set \mathcal{R} can be partitioned into a disjoint union of subsets

$$\mathcal{R} = \mathcal{R}_{\text{Sub}} \sqcup \mathcal{R}_{\text{Sub+Enz}} \sqcup \mathcal{R}_{\rightarrow\text{Int}} \sqcup \mathcal{R}_{\text{Int}\rightarrow} \sqcup \mathcal{R}_{\text{Int}},$$

which are uniquely determined from the partition $\mathcal{S} = \text{Enz} \cup \text{Sub} \cup \text{Int}$ by the inclusions

$$\begin{aligned} \mathcal{R}_{\text{Sub}} &\subseteq \{S \rightarrow S' \mid S, S' \in \text{Sub}\}, \\ \mathcal{R}_{\text{Sub+Enz}} &\subseteq \{S + E \rightarrow S' + E \mid E \in \text{Enz}, \text{ and } S, S' \in \text{Sub}\}, \\ \mathcal{R}_{\rightarrow\text{Int}} &\subseteq \{S + E \rightarrow Y' \mid E \in \text{Enz}, S \in \text{Sub}, \text{ and } Y' \in \text{Int}\}, \\ \mathcal{R}_{\text{Int}\rightarrow} &\subseteq \{Y \rightarrow S' + E \mid E \in \text{Enz}, S' \in \text{Sub}, \text{ and } Y \in \text{Int}\}, \\ \mathcal{R}_{\text{Int}} &\subseteq \{Y \rightarrow Y' \mid Y, Y' \in \text{Int}\}. \end{aligned}$$

(M2) Int is either empty or a set of intermediates of G .

(M3) If



is a reaction path in G for some $E, E' \in \text{Enz}$, some $S, S' \in \text{Sub}$, and some $Y^{(1)}, \dots, Y^{(\ell)} \in \text{Int}$, then $E = E'$.

Definition 13 (*PTM systems*) Let $G = (S, \mathcal{C}, \mathcal{R})$ be a reaction network, and let

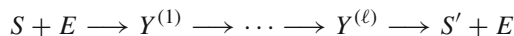
$$\mathcal{S} = \text{Enz} \cup \text{Sub} \cup \text{Int}$$

be a partition of the species set. We say that G is a *PTM system* with *enzyme* set Enz , *substrate* set Sub , and *intermediates* set Int if it has properties (M1)–(M3) above.

Let $G = (S, \mathcal{C}, \mathcal{R})$ be a PTM system. If $\text{Int} = \emptyset$, then set $G^* = (S^*, \mathcal{C}^*, \mathcal{R}^*) := G$. Otherwise, let $G^* = (S^*, \mathcal{C}^*, \mathcal{R}^*)$ be the network obtained from G by the removal of the set of intermediates Int . Thus,

$$\mathcal{R}^* = \mathcal{R}_{\text{Sub}} \cup \mathcal{R}_{\text{Sub+Enz}} \cup \mathcal{R}_{\text{Sub+Enz}}^Y,$$

where $\mathcal{R}_{\text{Sub+Enz}}^Y$ is the set of reactions of the form $S + E \rightarrow S' + E$ such that



is a reaction path in G for some $E \in \text{Enz}$, some $S, S' \in \text{Sub}$ such that $S \neq S'$, and some $Y^{(1)}, \dots, Y^{(\ell)} \in \text{Int}$.

Now $S^* \subseteq \text{Enz} \cup \text{Sub}$, and $\text{Enz}^* := \text{Enz} \cap S^*$, if nonempty, is a set of catalysts of G^* . Indeed, it follows directly from the form of the reactions that (C1) holds, and the subnetwork of G^* implied by Enz^* is the empty network, so (C2) also holds. If $\text{Enz}^* = \emptyset$, then we set $G^{**} = (S^{**}, \mathcal{C}^{**}, \mathcal{R}^{**}) := G^*$. Otherwise, let $G^{**} = (S^{**}, \mathcal{C}^{**}, \mathcal{R}^{**})$ be the network obtained from G^* by the removal of the set of catalysts

Enz*. Then G^{**} is a monomolecular network consisting of the reactions $S \rightarrow S'$ such that $S + \alpha E \rightarrow S' + \alpha E \in \mathcal{R}^*$ for some $E \in \text{Enz}^*$, some $\alpha \in \{0, 1\}$, and some $S, S' \in \text{Sub}$ such that $S \neq S'$. We refer to G^{**} as the *underlying substrate network of G* .

Regardless of whether Int or Enz are empty or nonempty, we shall abuse the terminology and refer to the reaction network G^* as the reaction network obtained from G by the removal of the set of intermediates Int and to G^{**} as the reaction network obtained from G^* by the removal of the set of catalysts Enz^* , for simplicity.

Note that G^* and G^{**} are themselves PTM systems with an empty set of intermediates and empty sets of intermediates and catalysts respectively.

By Thomson and Gunawardena (2009), Equations (16) and (17), any PTM system is conservative (see also Lemma 2 below). In particular, persistence and bounded-persistence are equivalent for PTM systems in view of Proposition 2.

Proposition 4 *Let G be a PTM system. Then the following properties are equivalent.*

- (i) G is consistent.
- (ii) Each connected component of the underlying substrate network G^{**} is strongly connected.
- (iii) G has the siphon/P-semiflow property.
- (iv) G has no drainable siphons.
- (v) G is persistent.

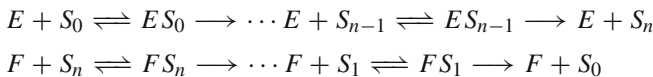
Proof Using that G is conservative, Proposition 2 and Corollary 1 give the following implications: (iii) \Rightarrow (iv) \Rightarrow (v) \Rightarrow (i). Thus, it remains to show that (i) \Rightarrow (ii) \Rightarrow (iii).

(i) \Rightarrow (ii). It follows from Theorems 1(iii) and 2(iii) that G^{**} is consistent. Since G^{**} is conservative, Proposition 3 gives that each connected component of G^{**} is strongly connected.

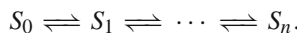
(ii) \Rightarrow (iii). By Proposition 3, G^{**} has the siphon/P-semiflow property. It then follows by Theorems 2(ii) and 1(ii), respectively, that G^* and, consequently, G have the siphon/P-semiflow property. □

Remark 3 In view of Proposition 3, statement (ii) in Proposition 4 is equivalent to each of the statements that the underlying substrate network G^{**} of G is consistent, has the siphon/P-semiflow property, has no drainable siphons, or is persistent. Thus, either of these properties could also be checked to establish the persistence of G .

Example 11 (An n -site phosphorylation mechanism) The sequential and distributive n -site phosphorylation mechanism given by



is a PTM system with $\text{Int} = \{ES_0, ES_1, \dots, ES_{n-1}, FS_n, FS_{n-1}, \dots, FS_1\}$, $\text{Enz} = \{E, F\}$, and $\text{Sub} = \{S_0, S_1, \dots, S_n\}$. The underlying substrate network is



It consists of a single strongly connected component, so the PTM system is persistent by Proposition 4.

Example 12 Consider the PTM system



The underlying substrate network, $S_0 \longrightarrow S_1$, is not strongly connected. We conclude that the PTM system is not persistent.

4.1.2 Signaling cascades of PTM systems

We now discuss a formalism for cascades of PTM systems. Intuitively, a signaling cascade of PTM systems is a reaction network that can be decomposed into a hierarchy of PTM systems in such a way that substrates at a certain level, or layer, may act as enzymes in lower levels (but not in higher levels).

Consider a reaction network $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$, and write the species, complex and reaction sets of the network as (not necessarily disjoint) unions,

$$\mathcal{S} = \bigcup_{i=1}^T \mathcal{S}_i, \quad \mathcal{C} = \bigcup_{i=1}^T \mathcal{C}_i, \quad \text{and} \quad \mathcal{R} = \bigcup_{i=1}^T \mathcal{R}_i. \tag{8}$$

Consider the following properties.

(F1) For each $i \in [T]$, $G_i := (\mathcal{S}_i, \mathcal{C}_i, \mathcal{R}_i)$ is a PTM system with enzyme, substrate, and intermediates sets, respectively, Enz_i , Sub_i , and Int_i .

(F2) $\text{Sub}_j \cap \left(\bigcup_{i=1}^{j-1} \text{Sub}_i \right) = \emptyset, j = 2, \dots, T.$

(F3) $\text{Enz}_j \cap \left(\bigcup_{i=1}^j \text{Sub}_i \right) = \emptyset, j = 1, \dots, T.$

(F4) $\left(\bigcup_{i=1}^T \text{Int}_i \right) \cap \left(\bigcup_{i=1}^T (\text{Enz}_i \cup \text{Sub}_i) \right) = \emptyset.$

Definition 14 (*Signaling cascades of PTM systems*) Let $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network. We say that G is a *signaling cascade of PTM systems* if there is a decomposition of the species, complex and reaction sets as in (8) that satisfy properties (F1)–(F4). In this case, we set

$$\text{Enz} := \bigcup_{i=1}^T \text{Enz}_i, \quad \text{Sub} := \bigcup_{i=1}^T \text{Sub}_i, \quad \text{and} \quad \text{Int} := \bigcup_{i=1}^T \text{Int}_i,$$

and the PTM systems $G_1 = (\mathcal{S}_1, \mathcal{C}_1, \mathcal{R}_1), \dots, G_T = (\mathcal{S}_T, \mathcal{C}_T, \mathcal{R}_T)$ are referred to as the *layers* of the cascade.

Remark 4 If $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ is a signaling cascade of PTM systems, then by (F1) and (F4), the species set \mathcal{S} can be partitioned into the two subsets $\text{Enz} \cup \text{Sub}$ and Int . That is, $\mathcal{S} = (\text{Enz} \cup \text{Sub}) \sqcup \text{Int}$. Furthermore, Int is a set of intermediates of G , provided that it is nonempty.

Observe that (F3) implies that any enzyme that is a substrate in some layer may appear in any layer below it, and not just the one immediately below the layer where it acts as a substrate. Thus, the layer hierarchy implied in the definition of signaling cascades of PTM systems may be a tree, in other words, it is not constrained to linear, sequential relationships where each layer can only provide the layer immediately after with enzymes.

Signaling cascades of PTM systems are always conservative.

Lemma 2 *Any signaling cascade of PTM systems is conservative.*

Proof Let $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a cascade of PTM systems. Write $\mathcal{S} = \{S_1, \dots, S_n\}$. With the notation in Definition 14 and using Remark 4, for each $i \in [n]$, set

$$\omega_i := \begin{cases} 1, & \text{if } S_i \in \text{Enz} \cup \text{Sub} \\ 2, & \text{if } S_i \in \text{Int}. \end{cases}$$

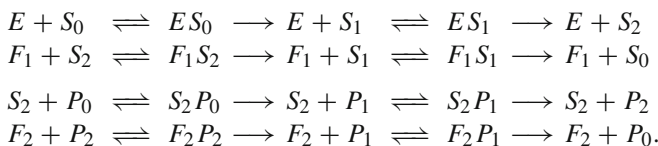
Then $\omega := (\omega_1, \dots, \omega_n)$ is a conservation law of G . This can be readily seen from the possible forms a reaction in \mathcal{R} may take. Since every entry of ω is strictly positive, this means G is conservative. □

Proposition 5 *Let G be a signaling cascade of PTM systems. Then the following properties are equivalent.*

- (i) G is consistent.
- (ii) The connected components of the underlying substrate network of each layer of G are strongly connected.
- (iii) G has the siphon/ P -semiflow property.
- (iv) G has no drainable siphons.
- (v) G is persistent.

The proof of Proposition 5 will be given in the next subsection.

Example 13 (Double phosphorylation cascade) Consider the concatenation of double phosphorylation mechanisms from Example 11 given by the reaction network



The double phosphorylation of a substrate S_0 is catalyzed by a kinase E , and the dephosphorylation of its singly and doubly phosphorylated forms is catalyzed by a

phosphatase F_1 . The doubly phosphorylated form S_2 of S_0 then acts as a kinase in a similar double phosphorylation/dephosphorylation mechanism for another substrate P_0 . This is a signaling cascade of PTM systems with

$$\begin{aligned} \text{Enz}_1 &= \{S_2, F_2\}, & \text{Sub}_1 &= \{P_0, P_1, P_2\}, & \text{Int}_1 &= \{S_2P_0, S_2P_1, F_2P_2, F_2P_1\}, \\ \text{Enz}_2 &= \{E, F_1\}, & \text{Sub}_2 &= \{S_0, S_1, S_2\}, & \text{Int}_2 &= \{ES_0, ES_1, F_1S_2, F_1S_1\}. \end{aligned}$$

Each of the layers of the cascade coincides with the double phosphorylation mechanism in Example 11 with $n = 2$. In particular, Proposition 5(ii) holds, hence the network is persistent.

In Gnacadja (2011b), the persistence of a class of cascades of PTM systems (there called *cascaded binary enzymatic networks*) is studied under mass-action kinetics. There is an overlap between the class of networks studied in Gnacadja (2011b) and the class of cascades of PTM systems considered here, although neither is more general than the other, nor do they agree. For instance, we allow for individual enzymes to take part in reactions in more than one layer of the cascade. In Gnacadja (2011b), sufficient conditions for a stronger concept of persistence (*vacuous persistence*) are given in terms of the so-called *futility* of the network (Gnacadja 2011b, Theorem 6.7). For conservative networks, vacuous persistence is equivalent to persistence together with the absence of boundary steady states in the stoichiometric compatibility classes of G that are not entirely contained in the boundary (Gnacadja 2011a, Proposition 5.2). In view of Proposition 2(v) and Proposition 5, persistence of a cascade of PTM systems in our setting is equivalent to vacuous persistence. Futility implies that the connected components of the underlying substrate network of the cascaded PTM system are strongly connected (Gnacadja 2011b, Remark 4.6). However, the condition is not necessary for futility. Therefore, our results establish that, for the overlapping class of cascades of PTM systems, strong connectedness of the components of the underlying substrate network is also necessary for vacuous persistence. We also note that our results are stated under more general kinetic assumptions.

4.1.3 Proof of Proposition 5

Since G is conservative, Proposition 2 and Corollary 1 give the implications (iii) \Rightarrow (iv) \Rightarrow (v) \Rightarrow (i).

It remains to prove (i) \Rightarrow (ii) and (ii) \Rightarrow (iii). We begin with a few simple observations about signaling cascades of PTM systems.

Let G be a signaling cascade of PTM systems with layers G_1, \dots, G_T . If the set of intermediates Int is nonempty, let G^* be the reaction network obtained by its removal. For each $i \in [T]$, let G_i^* be the reaction network obtained from G_i by the removal of the set of intermediates Int_i .

By the next lemma, we may assume without loss of generality that the cascade has no intermediates.

Lemma 3 *In the construction above, G^* is a signaling cascade of PTM systems with layers G_1^*, \dots, G_T^* . Furthermore,*

- (i) G^* has the siphon/P-semiflow property if, and only if G does also, and
- (ii) G^* is consistent if, and only if G is also.

Proof For each $i \in [T]$, set $\text{Sub}_i^* := \text{Sub}_i \cap \mathcal{S}_i^*$, $\text{Enz}_i^* := \text{Enz}_i \cap \mathcal{S}_i^*$, and $\text{Int}_i^* := \emptyset$. Then G_i^* is a PTM system with enzyme, substrate, and intermediates sets, respectively, Enz_i^* , Sub_i^* , and Int_i^* , thus satisfying (F1). Properties (F2) and (F3) are inherited directly from G , and (F4) is trivial. This proves the first statement. Statements (i) and (ii) then follow directly from Theorem 1(ii)–(iii). \square

Throughout the rest of this subsection, $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ will be assumed to be a signaling cascade of PTM systems with an empty set of intermediates.

Next, let $G^- = (\mathcal{S}^-, \mathcal{C}^-, \mathcal{R}^-)$ be the reaction network with

$$\mathcal{S}^- := \bigcup_{i=1}^{T-1} \mathcal{S}_i \quad \mathcal{C}^- := \bigcup_{i=1}^{T-1} \mathcal{C}_i \quad \text{and} \quad \mathcal{R}^- := \bigcup_{i=1}^{T-1} \mathcal{R}_i.$$

Set

$$\text{Enz}'_T := \text{Enz}_T \cap \left(\bigcup_{i=1}^{T-1} \text{Enz}_i \right).$$

If $\text{Enz}'_T \neq \emptyset$, then it is a set of catalysts of G^- . So, define G' to be the network obtained from G^- by the removal of the set of catalysts Enz'_T .

Lemma 4 *In the construction above, G' is a cascade of PTM systems with $T - 1$ layers G'_1, \dots, G'_{T-1} . Furthermore, for each $i \in [T - 1]$, the underlying substrate networks of G'_i and G_i coincide.*

Proof For each $i \in [T - 1]$, define \mathcal{R}'_i to be the set of reactions $S + \alpha E \rightarrow S' + \alpha E \in \mathcal{R}_i$ such that $S, S' \in \text{Sub}_i, S \neq S', E \in \text{Enz}_i \setminus \text{Enz}_T$, and $\alpha \in \{0, 1\}$, plus the reactions $S \rightarrow S'$ such that $S + E \rightarrow S' + E \in \mathcal{R}_i$ for some $S, S' \in \text{Sub}_i, S \neq S'$, and $E \in \text{Enz}_T$. Then define $G'_i = (\mathcal{S}'_i, \mathcal{C}'_i, \mathcal{R}'_i)$ to be the reaction network determined by \mathcal{R}'_i . We then have

$$\mathcal{S}' = \bigcup_{i=1}^{T-1} \mathcal{S}'_i, \quad \mathcal{C}' = \bigcup_{i=1}^{T-1} \mathcal{C}'_i, \quad \text{and} \quad \mathcal{R}' = \bigcup_{i=1}^{T-1} \mathcal{R}'_i.$$

Now G'_i is a PTM system with $\text{Enz}'_i = \text{Enz}_i \setminus \text{Enz}_T, \text{Sub}'_i = \text{Sub}_i$, and $\text{Int}'_i = \emptyset, i = 1, \dots, T - 1$. Indeed, (M1) is fulfilled by construction, and (M2) and (M3) hold vacuously. Thus, (F1) holds. Furthermore, properties (F2) and (F3) are inherited from G , and (F4) is fulfilled vacuously. This shows G' is a signaling cascade of PTM systems with layers G'_1, \dots, G'_{T-1} .

To establish the second statement, it is enough to show that $(\mathcal{R}'_i)^{**} = \mathcal{R}_i^{**}, i = 1, \dots, T - 1$. Let $i \in [T - 1]$, and $S \rightarrow S' \in (\mathcal{R}'_i)^{**}$. Then, by construction, $S + \alpha E \rightarrow S' + \alpha E \in \mathcal{R}_i$ for some $S, S' \in \text{Sub}_i, S \neq S', E \in \text{Enz}_i$, and $\alpha \in \{0, 1\}$,

and so $S \rightarrow S' \in \mathcal{R}_i^{**}$. Conversely, if $S \rightarrow S' \in \mathcal{R}_i^{**}$, then $S + \alpha E \rightarrow S' + \alpha E \in \mathcal{R}_i$ for some $S, S' \in \text{Sub}_i, S \neq S', E \in \text{Enz}_i$, and $\alpha \in \{0, 1\}$. If $E \in \text{Enz}_T$ and $\alpha = 1$, then we get $S \rightarrow S' \in \mathcal{R}'_i$ by construction, and so $S \rightarrow S' \in (\mathcal{R}'_i)^{**}$. Otherwise, $S + \alpha E \rightarrow S' + \alpha E \in \mathcal{R}'_i$, and so $S \rightarrow S' \in (\mathcal{R}'_i)^{**}$ after the removal of catalysts. \square

Finally, let \widehat{G}_T be the reaction network obtained from G by the removal of the set of catalysts Enz_T , and let G_T^{**} be the underlying substrate network of G_T . Upon ordering the species and reactions of \widehat{G}_T in such a way that all species belonging to Sub_T correspond to the bottom-most rows, and all monomolecular reactions between species in Sub_T correspond to the right-most columns, the stoichiometric matrix of \widehat{G}_T may be written as

$$\widehat{N}_T = \begin{bmatrix} N' & 0 \\ 0 & N_T^{**} \end{bmatrix}, \tag{9}$$

where N' is the stoichiometric matrix of the network G' introduced above, and N_T^{**} is the stoichiometric matrix of G_T^{**} . This decomposition will be used in the proofs of the next two results.

Proof of (i) \Rightarrow (ii) in Proposition 5. We use induction on the number T of layers. For $T = 1$, this follows from Proposition 4.

Now suppose the result holds for cascades of PTM systems with $T - 1$ layers for some $T \geq 2$, and let G be a cascade with T layers. By Theorem 2(iii), \widehat{G}_T is consistent. So, there exists a $\widehat{v}_T \gg 0$ such that $\widehat{N}_T \widehat{v}_T = 0$. We may write $\widehat{v}_T = (v', v_T^{**})$, where v' corresponds to the reactions of G' , and v_T^{**} corresponds to the reactions of G_T^{**} . From (9), we obtain

$$N'v' = 0 \quad \text{and} \quad N_T^{**}v_T^{**} = 0,$$

concluding that G' and G_T^{**} are consistent. It follows by the inductive hypothesis, Lemma 4, and Proposition 3 that $G_1^{**}, \dots, G_{T-1}^{**}, G_T^{**}$, the underlying substrate networks of G_1, \dots, G_{T-1}, G_T , respectively, are such that their connected components are strongly connected. This establishes the inductive step, proving the result. \square

Proof of (ii) \Rightarrow (iii) in Proposition 5. We use induction on the number T of layers. For $T = 1$, this follows from Proposition 4.

Now suppose the result holds for signaling cascades of PTM systems with $T - 1$ layers for some $T \geq 2$, and let G be a cascade with T layers. By Theorem 2(ii), it is enough to show that \widehat{G}_T has the siphon/P-semiflow property.

By construction, the species set \widehat{S}_T of \widehat{G}_T can be partitioned as the disjoint union $\widehat{S}_T = S' \cup S_T^{**}$ of the species sets of G' and G_T^{**} . We claim that every minimal siphon of \widehat{G}_T is entirely contained in either S' or S_T^{**} . To see this, let $\widehat{\Sigma}_T$ be any minimal siphon of \widehat{G}_T , and suppose it is not entirely contained in S' . So, $\Sigma_T \cap S_T^{**} \neq \emptyset$. By hypothesis, G_T^{**} is a monomolecular network with the property that each of its connected components is strongly connected. Thus, each of its connected components is a minimal siphon. We conclude that Σ_T contains one of the connected components of G_T^{**} and, by minimality, must be actually equal to it.

By the inductive hypothesis and Lemma 4, G' has the siphon/P-semiflow property. By Proposition 3, G_T^{**} also has the siphon/P-semiflow property. We conclude from the

block-diagonal decomposition in (9) and the claim above that \widehat{G}_T has the siphon/P-semiflow property. \square

4.2 Dissipative networks

In the next definition, we use the same notation as in Sect. 2.3.

Definition 15 (*Dissipative networks*) A reaction network (3) is said to be *dissipative* if its solutions are eventually uniformly bounded. More precisely, if there exists a constant $K \geq 0$ such that

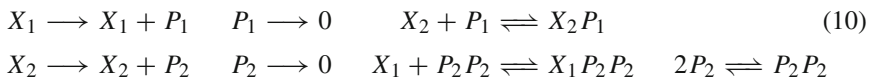
$$\limsup_{t \rightarrow \infty} |\sigma(t, s_0)| \leq K,$$

for each initial state $s_0 \geq 0$.

Corollary 3 *If a dissipative reaction network is bounded-persistent, then it is persistent.*

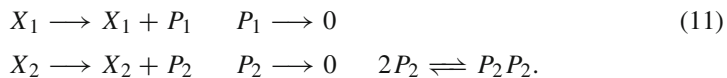
Proof Indeed, every solution of a dissipative reaction network is bounded. The conclusion then follows from Lemma 17 in Sect. 1 in the appendix. \square

Example 14 (Monomer–dimer toggle) Consider the monomer–dimer toggle model given by the reaction network

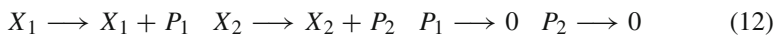


The leftmost four reactions model basal protein production and degradation. The $P_2 P_2$ represents a dimeric species, while $X_2 P_1$ and $X_1 P_2 P_2$ represent, respectively, monomers and dimers bound to gene promoters. See Siegal-Gaskins et al. 2015, page S1 for further contextualization.

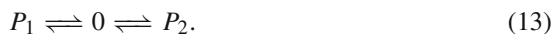
By removing the set of intermediates $\{X_2 P_1, X_1 P_2 P_2\}$, we obtain the network



Now $\{P_2 P_2\}$ constitutes a set of intermediates of (11). Its removal yields



Now $\{X_1, X_2\}$ is a set of catalysts of (12). Their removal leaves us with

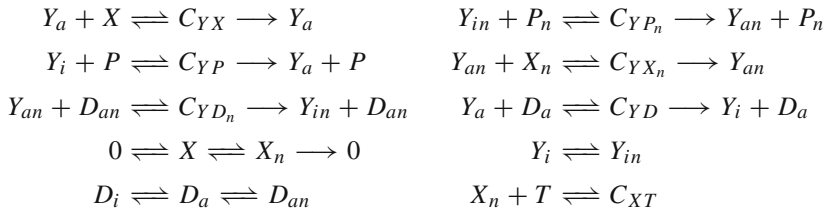


This is a non-conservative strongly connected monomolecular network. Thus, by Corollary 2, the network (10) is bounded-persistent and does not have boundary steady

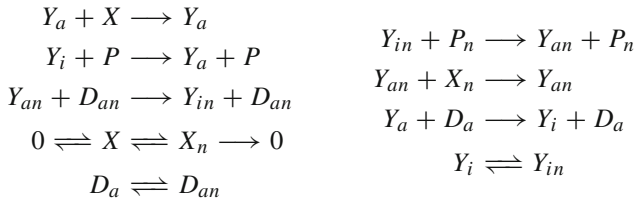
states in any stoichiometric compatibility class that is not contained in the boundary of the positive orthant. Under mass-action kinetics, (10) is dissipative (Siegal-Gaskins et al. 2015, pages S7–S8). Thus, it is also persistent by Corollary 3.

4.3 A shuttling and degradation focused Wnt model

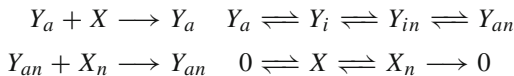
The following reaction network model for the Wnt pathway was proposed in MacLean et al. (2015).



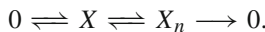
Note that $\{C_{YX}, C_{YP_n}, C_{YP}, C_{YX_n}, C_{YD_n}, C_{YD}, C_{XT}, D_i\}$ is a set of intermediates. Their removal yields the reaction network



Now $\{D_a, D_{an}, P_n, P\}$ constitutes a set of catalysts. After their removal, we obtain the reaction network



We may now remove $\{Y_i, Y_{in}\}$ as a set of intermediates, and then remove $\{Y_a, Y_{an}\}$ as a set of catalysts, thus obtaining



This is a non-conservative strongly connected monomolecular network. Thus, by Corollary 2, the network (10) is bounded-persistent and does not have boundary steady states in any stoichiometric compatibility class that is not already contained in the boundary of the positive orthant. \square

5 Proofs of Theorems 1, 2 and 3

In Sect. 5.1 we prove Theorem 1, items (i), (iii) and (iv). Likewise, the proof of Theorem 2, items (i), (iii) and (iv) is carried out in Sect. 5.2. Item (ii) in each result follows from (i) in the same result by Corollary 1. The proof of Theorem 3 is worked out in Sect. 5.3.

We begin with a general fact about reaction networks. Let $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network, and let $(\mathcal{C}_1, \mathcal{R}_1), \dots, (\mathcal{C}_J, \mathcal{R}_J)$ be the connected components of its reaction graph $(\mathcal{C}, \mathcal{R})$.

Lemma 5 *For each $j \in [J]$, $y' - y \in \Gamma$ for all $y, y' \in \mathcal{C}_j$.*

Proof Since y and y' are in the same connected component of $(\mathcal{C}, \mathcal{R})$, there exists an undirected reaction path $y - y_1 - \dots - y_k - y'$ in $(\mathcal{C}, \mathcal{R})$ connecting y and y' . Now

$$y' - y = (y' - y_1) + \sum_{i=2}^k (y_{i-1} - y_i) + (y_k - y) \in \Gamma,$$

establishing the lemma. □

5.1 Intermediates

Now suppose $G^* = (\mathcal{S}^*, \mathcal{C}^*, \mathcal{R}^*)$ is the reduction of G by the removal of a set of intermediates \mathcal{Y} . Recall that \mathcal{S}^* does not always agree with $\mathcal{S} \setminus \mathcal{Y}$. Let

$$\mathcal{X} := (\mathcal{S} \setminus \mathcal{Y}) \setminus \mathcal{S}^*,$$

and write

$$\mathcal{X} = \{X_1, \dots, X_\ell\}, \quad \text{and} \quad \mathcal{S}^* = \{S_1^*, \dots, S_n^*\}.$$

Thus,

$$\mathcal{S} = \mathcal{S}^* \cup \mathcal{X} \cup \mathcal{Y} = \{S_1^*, \dots, S_n^*, X_1, \dots, X_\ell, Y_1, \dots, Y_p\}.$$

This is the ordering we shall assume whenever working with the stoichiometric matrix or the stoichiometric subspace of G . Given a complex

$$y = (\alpha_1, \dots, \alpha_n, \gamma_1, \dots, \gamma_\ell, \beta_1, \dots, \beta_p) = \sum_{i=1}^n \alpha_i S_i^* + \sum_{i=1}^{\ell} \gamma_i X_i + \sum_{i=1}^p \beta_i Y_i,$$

in \mathcal{C} , we will denote its projection over the first n coordinates by

$$\widehat{y} := (\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i S_i^*. \tag{14}$$

Conversely, given a complex \widehat{y} in $\mathbb{R}_{\geq 0}^n$ as in (14), we denote its embedding in $\mathbb{R}^{n+\ell+p}$ by

$$y := (\alpha_1, \dots, \alpha_n, 0, \dots, 0) = \sum_{i=1}^n \alpha_i S_i^*.$$

Lemma 6 *For each $j \in [J]$, if $y, y' \in \mathcal{C}_j \setminus \mathcal{Y}$, then \widehat{y} and \widehat{y}' are in the same connected component of $(\mathcal{C}^*, \mathcal{R}^*)$.*

Proof If $y \neq y'$, then there exists an undirected reaction path

$$y - y_1 - \dots - y_k - y'$$

in $(\mathcal{C}, \mathcal{R})$ connecting y and y' . Let $i_1, \dots, i_d \in [k]$ be the indices such that $y_{i_1}, \dots, y_{i_d} \in \mathcal{C} \setminus \mathcal{Y}$, so that each non-endpoint in each of the paths

$$\begin{aligned} & y - y_1 - \dots - y_{i_1-1} - y_{i_1}, \\ & y_{i_1} - y_{i_1+1} - \dots - y_{i_2-1} - y_{i_2}, \\ & \quad \vdots \\ & y_{i_d} - y_{i_d+1} - \dots - y_k - y', \end{aligned} \tag{15}$$

is an intermediate. We may assume without loss of generality that, within each path, all arrows point in the same direction. To see this, suppose that this is not the case for, say, the first path. Suppose $y \rightarrow y_1$, and let $q_1 \in [i_1 - 1]$ be the index corresponding to the first (intermediate) complex where the arrows switch directions. So we have

$$y \longrightarrow y_1 \longrightarrow \dots \longrightarrow y_{q_1} \longleftarrow y_{q_1+1} \dots$$

By (I2), there exists $y^{(1)} \in \mathcal{C} \setminus \mathcal{Y}$, and $Y^{(1)}, \dots, Y^{(p_1)} \in \mathcal{Y}$ such that,

$$y_{q_1} \longrightarrow Y^{(1)} \longrightarrow \dots \longrightarrow Y^{(p_1)} \longrightarrow y^{(1)}$$

is a reaction path in G . We may then split the first path in (15) into the two paths

$$\begin{aligned} & y \longrightarrow y_1 \longrightarrow \dots \longrightarrow y_{q_1} \longrightarrow Y^{(1)} \longrightarrow \dots \longrightarrow Y^{(p_1)} \longrightarrow y^{(1)}, \\ & y^{(1)} \longleftarrow Y^{(p_1)} \longleftarrow \dots \longleftarrow Y^{(1)} \longleftarrow y_{q_1} \longleftarrow y_{q_1+1} \dots, \end{aligned}$$

where in each path we remove any loops starting and ending at an intermediate that might have been created. If there are other changes of direction between y_{q_1} and y_{i_1} , we may employ the same construction as many times as needed. If $y \leftarrow y_1$ instead, the argument is analogous, and the same construction applies also to any other path not having the property that all arrows point in the same direction.

This construction gives an undirected reaction path in $(\mathcal{C}^*, \mathcal{R}^*)$:

$$\widehat{y} - \widehat{y}_{i_1} - \cdots - \widehat{y}_{i_d} - \widehat{y}'.$$

We conclude that \widehat{y} and \widehat{y}' are in the same connected component of $(\mathcal{C}^*, \mathcal{R}^*)$. \square

Remark 5 We have that $\text{supp } y \cap \mathcal{X} = \emptyset$ whenever the connected component of G that contains y contains at least one more non-intermediate complex y' .

Conservation laws

We prove here Theorem 1(iv). In what follows, $\Gamma^* \subseteq \mathbb{R}^n$ is the stoichiometric subspace of G^* . Thus, its orthogonal complement $(\Gamma^*)^\perp$ is taken in \mathbb{R}^n .

Lemma 7 For each $j \in [J]$,

$$(\omega^*, x, 0) \cdot y = (\omega^*, x, 0) \cdot y', \quad \forall y, y' \in \mathcal{C}_j \setminus \mathcal{Y}, \quad \forall (\omega^*, x) \in (\Gamma^*)^\perp \times \mathbb{R}^\ell.$$

Proof Fix arbitrarily $\omega^* \in (\Gamma^*)^\perp, x \in \mathbb{R}^\ell, j \in [J]$, and $y, y' \in \mathcal{C}_j \setminus \mathcal{Y}$. The equality is trivial if $y = y'$, so, assume $y \neq y'$. By Lemma 6, \widehat{y} and \widehat{y}' are in the same connected component of $(\mathcal{C}^*, \mathcal{R}^*)$. By Lemma 5, we conclude that $\widehat{y} - \widehat{y}' \in \Gamma^*$. In view of Remark 5, we have now

$$(\omega^*, x, 0) \cdot (y - y') = \omega^* \cdot (\widehat{y} - \widehat{y}') = 0,$$

completing the proof. \square

For each $j \in [J]$, fix arbitrarily a complex $y_j \in \mathcal{C}_j \setminus \mathcal{Y}$. Property (I2) in the definition of intermediates ensures that $\mathcal{C}_j \setminus \mathcal{Y}$ is always nonempty. For each $i \in [p]$, let $j_i \in [J]$ be the index uniquely defined by the property that $Y_i \in \mathcal{C}_{j_i}$. Define

$$\begin{aligned} a: (\Gamma^*)^\perp \times \mathbb{R}^\ell &\longrightarrow \mathbb{R}^p \\ (\omega^*, x) &\longmapsto ((\omega^*, x, 0) \cdot y_{j_1}, \dots, (\omega^*, x, 0) \cdot y_{j_p}). \end{aligned}$$

Note that, by Lemma 7, a is independent of the chosen representatives $y_j \in \mathcal{C}_j \setminus \mathcal{Y}, j \in [J]$.

Lemma 8 Suppose $G^* = (\mathcal{S}^*, \mathcal{C}^*, \mathcal{R}^*)$ is the reduction of $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ by the removal of a set of intermediates \mathcal{Y} . Then

$$\Gamma^\perp = \{(\omega^*, x, a(\omega^*, x)) \mid (\omega^*, x) \in (\Gamma^*)^\perp \times \mathbb{R}^\ell\}.$$

Proof (I) We first show the inclusion \supseteq . To this end, fix arbitrarily $\omega^* \in (\Gamma^*)^\perp$, and $x \in \mathbb{R}^\ell$. Denote

$$\omega := (\omega^*, x, a(\omega^*, x)).$$

Fix arbitrarily $y \rightarrow y' \in \mathcal{R}$. We want to show that

$$\omega \cdot (y' - y) = 0.$$

There are four possibilities.

(1) If $\widehat{y} \rightarrow \widehat{y}' \in \mathcal{R}^*$, then

$$\omega \cdot (y' - y) = \omega^* \cdot (\widehat{y}' - \widehat{y}) = 0.$$

(2) If $y \rightarrow y' = y \rightarrow Y_i$ for some $Y_i \in \mathcal{Y}$, and some $y \in \mathcal{C} \setminus \mathcal{Y}$, then

$$\omega \cdot (y' - y) = \omega \cdot (Y_i - y) = (\omega^*, x, 0) \cdot y_{j_i} - (\omega^*, x, 0) \cdot y = 0,$$

by Lemma 7, since y and y_{j_i} belong to the same connected component of $(\mathcal{C}, \mathcal{R})$.

(3) If $y \rightarrow y' = Y_i \rightarrow y'$ for some $Y_i \in \mathcal{Y}$, and $y \in \mathcal{C} \setminus \mathcal{Y}$, then the argument is the same as in (2).

(4) If $y \rightarrow y' = Y_i \rightarrow Y_k$ for some $Y_i, Y_k \in \mathcal{Y}$, then $j_i = j_k$, and so

$$\omega \cdot (y' - y) = \omega \cdot (Y_k - Y_i) = (\omega^*, x, 0) \cdot y_{j_i} - (\omega^*, x, 0) \cdot y_{j_k} = 0.$$

This establishes the inclusion \supseteq . In particular,

$$\dim((\Gamma^*)^\perp \times \mathbb{R}^\ell) = \dim(\Gamma^*)^\perp + \ell \leq \dim \Gamma^\perp.$$

(II) To finish the proof, it is now enough to show that

$$\dim \Gamma^\perp \leq \dim(\Gamma^*)^\perp + \ell. \tag{16}$$

We claim that

$$\dim \Gamma \geq \dim \Gamma^* + p. \tag{17}$$

On the one hand, for each reaction $\widehat{y} \rightarrow \widehat{y}' \in \mathcal{R}^*$, there exists a reaction path in G connecting y to y' , so y and y' are in the same connected component of $(\mathcal{C}, \mathcal{R})$. It follows by Lemma 5 that $y' - y \in \Gamma$. On the other hand, for each intermediate $Y_i \in \mathcal{Y}$, there exists a $y^{(i)} \in \mathcal{C} \setminus \mathcal{Y}$ and a reaction path in G connecting Y_i to $y^{(i)}$. Again by Lemma 5, we have $Y_i - y^{(i)} \in \Gamma$. Furthermore, $Y_1 - y^{(1)}, \dots, Y_p - y^{(p)}, (y' - y)$ are linearly independent for each $\widehat{y} \rightarrow \widehat{y}' \in \mathcal{R}^*$. This gives us (17).

Combining $\dim \Gamma + \dim \Gamma^\perp = \dim \mathbb{R}^{n+\ell+p}$ and (17), we get

$$\dim \Gamma^\perp = n + \ell + p - \dim \Gamma \leq n - \dim \Gamma^* + \ell = \dim(\Gamma^*)^\perp + \ell.$$

This establishes (16), completing the proof. □

Proof of Theorem 1(iv). (\Rightarrow) If $\omega = (\omega^*, x, a(\omega^*, x))$ is a strictly positive conservation law of G , then ω^* is a strictly positive conservation law of G^* .

(\Leftarrow) If ω^* is a strictly positive conservation law of G^* , then choose any $x \in \mathbb{R}_{>0}^\ell$. It holds that $(\omega^*, x, 0) \cdot y_{j_i} > 0$ for all $i \in [p]$ since $y_{j_i} \neq 0$ and the support of y_{j_i} is included in the support of (ω^*, x) . Then $\omega = (\omega^*, x, a(\omega^*, x))$ is a strictly positive conservation law of G by Lemma 8. \square

Siphons

Lemma 9 *Suppose $G^* = (S^*, C^*, R^*)$ is the reduction of $G = (S, C, R)$ by the removal of a set of intermediates \mathcal{Y} . If Σ is a siphon of G , then*

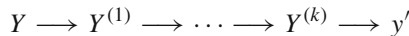
$$\Sigma^* := \Sigma \cap S^*$$

is either the empty set, or a siphon of G^ . Furthermore, if Σ^* is empty, then $\Sigma \cap \mathcal{X}$ is nonempty.*

Proof First suppose $\Sigma^* \neq \emptyset$. Pick any $S' \in \Sigma^*$, and let $\widehat{y} \rightarrow \widehat{y}' \in R^*$ be any reaction having S' as one of its products. Then there exists a reaction path in G connecting y and y' . Since Σ is a siphon of G , some species S constituting y belongs to Σ . Since $\widehat{y} \in C^*$, we must have $S \in \Sigma^*$. Thus, Σ^* is a siphon of G^* .

Now suppose $\Sigma^* = \emptyset$. Since $\Sigma \neq \emptyset$ and $S = S^* \cup \mathcal{X} \cup \mathcal{Y}$, we must have $\Sigma \cap \mathcal{X} \neq \emptyset$ or $\Sigma \cap \mathcal{Y} \neq \emptyset$. If $\Sigma \cap \mathcal{X} \neq \emptyset$, then we have nothing left to prove. So, assume $\Sigma \cap \mathcal{Y} \neq \emptyset$, and fix arbitrarily a $Y \in \Sigma \cap \mathcal{Y}$. By (I2), there exist $y \in C \setminus \mathcal{Y}$ and a reaction path in G connecting y to Y , and we conclude that one of the species in y belongs to Σ . Since y is supported in $S^* \cup \mathcal{X}$, and since $\Sigma \cap S^* = \emptyset$ by hypothesis, we conclude that $\Sigma \cap \mathcal{X} \neq \emptyset$. \square

Given a subset $\Sigma \subseteq S$, we define $M(\Sigma)$ to be the subset of intermediates $Y \in \mathcal{Y}$ that appear in a reaction path



for some y' whose support intersects Σ , and some $Y^{(1)}, \dots, Y^{(k)} \in \mathcal{Y}$. Note that if Σ is a siphon of G , then $M(\Sigma) \subseteq \Sigma$ by Remark 1.

Lemma 10 *Suppose $G^* = (S^*, C^*, R^*)$ is the reduction of $G = (S, C, R)$ by the removal of a set of intermediates \mathcal{Y} . If Σ^* is a siphon of G^* , then*

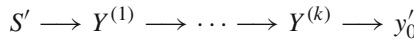
$$\Sigma := \Sigma^* \cup M(\Sigma^*)$$

is a siphon of G . Furthermore, any siphon of G containing Σ^ must also contain $M(\Sigma^*)$.*

Proof Pick any $S' \in \Sigma$, and let $y \rightarrow y' \in R$ be any reaction having S' as one of its products.

Suppose first $S' \in \Sigma^*$. If $\widehat{y} \rightarrow \widehat{y}' \in R^*$, then Σ^* contains some reactant of $y \rightarrow y'$, and so does Σ . If $\widehat{y} \rightarrow \widehat{y}' \notin R^*$, then $y \rightarrow y' = Y \rightarrow y'$ for some $Y \in \mathcal{Y}$. By construction, $Y \in M(\Sigma^*) \subseteq \Sigma$.

Now suppose $S' \notin \Sigma^*$. Then $S' \in M(\Sigma^*)$, meaning that $y' = S' \in \mathcal{Y}$, and that there exists a reaction path



in G such that the support of y'_0 intersects Σ^* , and $Y^{(1)}, \dots, Y^{(k)} \in \mathcal{Y}$. If $y \in \mathcal{Y}$, then it follows that $y \in M(\Sigma^*)$, and so $y \rightarrow y'$ has a reactant in Σ . If $y \notin \mathcal{Y}$, we have $\widehat{y} \rightarrow \widehat{y}'_0 \in \mathcal{R}^*$, and so one of the species constituting \widehat{y} belongs to Σ^* . We conclude that one of the reactants of $y \rightarrow y'$ belongs to Σ . This completes the proof that Σ is a siphon of G .

It follows straight from the construction of $M(\Sigma^*)$ and Remark 1, that any siphon of G containing Σ^* must also contain all the intermediates in $M(\Sigma^*)$. \square

Drainable and self-replicable siphons

Lemma 11 *Suppose $G^* = (S^*, C^*, \mathcal{R}^*)$ is the reduction of $G = (S, C, \mathcal{R})$ by the removal of a set of intermediates \mathcal{Y} . If $\Sigma \subseteq S$ is a drainable or self-replicable siphon of G , then $\Sigma^* := \Sigma \cap S^*$ is nonempty.*

Proof Suppose on contrary that $\Sigma^* = \emptyset$. Then $\Sigma \cap \mathcal{X}$ is nonempty by Lemma 9. Consider the P-semiflow

$$\omega := (0, x, a(0, x)) > 0, \quad \text{where } x := \sum_{i: X_i \in \Sigma \cap \mathcal{X}} e_i.$$

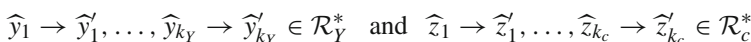
Let $i \in [p]$ such that $a_i(0, x) = (0, x, 0) \cdot y_{j_i} > 0$. Then $\text{supp}(y_{j_i}) \cap (\Sigma \cap \mathcal{X}) \neq \emptyset$. By Remark 5, y_{j_i} is the only non-intermediate complex in the connected component where Y_i belongs to. Therefore there must exist a reaction path connecting Y_i to y_{j_i} and hence $Y_i \in M(\Sigma \cap \mathcal{X})$. It follows that $\text{supp } \omega \subseteq (\Sigma \cap \mathcal{X}) \cup M(\Sigma \cap \mathcal{X}) \subseteq \Sigma$. So, Σ contains the support of a P-semiflow, in other words, Σ is noncritical. We conclude by Proposition 1(i) that Σ can be neither drainable nor self-replicable, contradicting the hypotheses. \square

We are now ready to prove Theorem 1(i).

Proof of Theorem 1(i). In virtue of Lemma 1, it suffices to show that the result holds for the removal of a single intermediate Y . The general result then follows by induction on the size of the set of intermediates.

(\Leftarrow) Suppose $\Sigma^* \subseteq S^*$ is a drainable siphon of G^* . By Lemma 10, $\Sigma := \Sigma^* \cup M(\Sigma^*)$ is a siphon of G . We will show that it is drainable.

By construction, we have $M(\Sigma^*) = \emptyset$ or $M(\Sigma^*) = \{Y\}$. If $M(\Sigma^*) = \emptyset$, then $\Sigma = \Sigma^*$, and any reaction that contains a species in Σ^* in the product belongs to \mathcal{R}_c^* . Thus we have nothing left to show. So, we may assume $M(\Sigma^*) = \{Y\}$. Since Σ^* is drainable, there exist reactions



such that

$$\theta_i := \sum_{j=1}^{k_Y} (\widehat{y}'_j - \widehat{y}_j)_i + \sum_{j=1}^{k_c} (\widehat{z}'_j - \widehat{z}_j)_i < 0, \quad \forall i \in [n] : S_i^* \in \Sigma^*.$$

Let T be a large enough positive integer such that $(\widehat{y}'_1)_i + T\theta_i < 0$ for every $i \in [n]$ such that $S_i^* \in \Sigma^*$. We have

$$y_1 \rightarrow Y, Y \rightarrow y'_1, \dots, y_{k_Y} \rightarrow Y, Y \rightarrow y'_{k_Y}, z_1 \rightarrow z'_1, \dots, z_{k_c} \rightarrow z'_{k_c} \in \mathcal{R}.$$

Let

$$\alpha := (y'_1 - Y) + \sum_{j=1}^{k_Y} T(Y - y_j) + \sum_{j=1}^{k_Y} T(y'_j - Y) + \sum_{j=1}^{k_c} T(z'_j - z_j)$$

We have

$$\alpha_i = \begin{cases} (\widehat{y}'_1)_i + T\theta_i < 0 & \text{if } i \in [n] : S_i^* \in \Sigma^* \\ [lex] - 1 < 0 & \text{if } i = n + \ell + 1. \end{cases}$$

The last case corresponds to the coordinate of Y . This shows Σ is drainable.

(\Rightarrow) Suppose $\Sigma \subseteq \mathcal{S}$ is a drainable siphon of G . By Lemmas 9 and 11, $\Sigma^* := \Sigma \cap \mathcal{S}^*$ is a siphon of G^* . We will show that it is drainable.

Since Σ is drainable, there exist reactions

$$y_1 \rightarrow Y, \dots, y_{k_Y} \rightarrow Y, Y \rightarrow y'_1, \dots, Y \rightarrow y'_{k_d}, z_1 \rightarrow z'_1, \dots, z_{k_c} \rightarrow z'_{k_c} \in \mathcal{R} \quad (18)$$

such that $y_1, \dots, y_{k_Y}, y'_1, \dots, y'_{k_d}, z_1, z'_1, \dots, z_{k_c}, z'_{k_c} \in \mathcal{C} \setminus \mathcal{Y}$, and

$$\alpha_i := \sum_{j=1}^{k_Y} (Y - y_j)_i + \sum_{j=1}^{k_d} (y'_j - Y)_i + \sum_{j=1}^{k_c} (z'_j - z_j)_i < 0,$$

for all $i \in [n]$ such that $S_i^* \in \Sigma^* = \Sigma \cap \mathcal{S}^*$, and also for $i = n + \ell + 1$, if $Y \in \Sigma$. In the latter case, since $\alpha_{n+\ell+1} = k_Y - k_d$, it must hold that $k_Y < k_d$.

If $k_d > k_Y$, then

$$\alpha_i = \sum_{j=1}^{k_Y} (y'_j - y_j)_i + \sum_{j=k_c+1}^{k_d} (y'_j)_i + \sum_{j=1}^{k_c} (z'_j - z_j)_i \geq \sum_{j=1}^{k_Y} (\widehat{y}'_j - \widehat{y}_j)_i + \sum_{j=1}^{k_c} (\widehat{z}'_j - \widehat{z}_j)_i,$$

for all i such that $S_i^* \in \Sigma^*$. Thus Σ^* is drainable using the reactions $\widehat{y}_1 \rightarrow \widehat{y}'_1, \dots, \widehat{y}_{k_Y} \rightarrow \widehat{y}'_{k_Y} \in \mathcal{R}_Y^*$ and $\widehat{z}_1 \rightarrow \widehat{z}'_1, \dots, \widehat{z}_{k_c} \rightarrow \widehat{z}'_{k_c} \in \mathcal{R}_c^*$.

If $k_Y \geq k_d$, then necessarily $Y \notin \Sigma$. Then, for all reactions in (18) of the form $Y \rightarrow y'$, the support of y' does not contain any species in Σ^* . In particular this holds for y'_1, \dots, y'_{k_d} . Choose any $y' \in \mathcal{C} \setminus \mathcal{Y}$ such that $Y \rightarrow y' \in \mathcal{R}$. We have

$$\alpha_i = \sum_{j=1}^{k_Y} -(y_j)_i + \sum_{j=1}^{k_Y} (y'_j)_i + \sum_{j=1}^{k_c} (z'_j - z_j)_i = \sum_{j=1}^{k_Y} (\hat{y}'_j - \hat{y}_j)_i + \sum_{j=1}^{k_c} (\hat{z}'_j - \hat{z}_j)_i,$$

for all i such that $S_i^* \in \Sigma^*$, since $(y'_j)_i = 0$. Thus Σ^* is drainable using the reactions $\hat{y}_1 \rightarrow \hat{y}'_1, \dots, \hat{y}_{k_Y} \rightarrow \hat{y}'_{k_Y} \in \mathcal{R}_Y^*$ and $\hat{z}_1 \rightarrow \hat{z}'_1, \dots, \hat{z}_{k_c} \rightarrow \hat{z}'_{k_c} \in \mathcal{R}_c^*$.

The proof for self-replicable siphons is entirely analogous, with the appropriate inequalities reversed, and the roles played by reactions creating or consuming Y swapped. \square

Consistency

Lemma 12 *Suppose $G^* = (S^*, \mathcal{C}^*, \mathcal{R}^*)$ is the reduction of a reaction network $G = (S, \mathcal{C}, \mathcal{R})$ by the removal of a set of intermediates $\{Y\}$ containing a single intermediate Y . Then G^* is consistent if, and only if network G is consistent.*

Proof (\Rightarrow) Suppose G^* is consistent. This is equivalent to say that

$$\sum_{y \rightarrow y' \in \mathcal{R}^*} v_{y \rightarrow y'} (y' - y) = 0$$

for some $v_{y \rightarrow y'} > 0$, $y \rightarrow y' \in \mathcal{R}^*$. Let $\mathcal{R}_Y^* \subseteq \mathcal{R}^*$ be the subset of reactions $y \rightarrow y' \in \mathcal{R}^*$ such that $y \rightarrow Y, Y \rightarrow y' \in \mathcal{R}$, and let $\mathcal{R}_c^* \subseteq \mathcal{R}^*$ be the subset of all reactions $y \rightarrow y' \in \mathcal{R}^*$ such that $y \rightarrow y' \in \mathcal{R}$. Note that

$$\mathcal{R}^* = \mathcal{R}_Y^* \cup \mathcal{R}_c^*,$$

and that the union need not be disjoint. Let $\mathcal{C}_{\leftrightarrow} \subseteq \mathcal{C}$ be the subset of complexes $y \in \mathcal{C}$ such that $y \rightarrow Y, Y \rightarrow y \in \mathcal{R}$. Then

$$\begin{aligned} 0 &= \left(\sum_{y \rightarrow y' \in \mathcal{R}_Y^* \setminus \mathcal{R}_c^*} + \sum_{y \rightarrow y' \in \mathcal{R}_Y^* \cap \mathcal{R}_c^*} + \sum_{y \rightarrow y' \in \mathcal{R}_c^* \setminus \mathcal{R}_Y^*} \right) v_{y \rightarrow y'} (y' - y) \\ &\quad + \sum_{y \in \mathcal{C}_Y} (Y - y) + \sum_{y' \in \mathcal{C}_Y} (y' - Y) \\ &= \left(\frac{1}{2} \sum_{y \rightarrow y' \in \mathcal{R}_Y^* \cap \mathcal{R}_c^*} + \sum_{y \rightarrow y' \in \mathcal{R}_c^* \setminus \mathcal{R}_Y^*} \right) v_{y \rightarrow y'} (y' - y) \\ &\quad + \left(\sum_{y \rightarrow y' \in \mathcal{R}_Y^* \setminus \mathcal{R}_c^*} v_{y \rightarrow y'} + \frac{1}{2} \sum_{y \rightarrow y' \in \mathcal{R}_Y^* \cap \mathcal{R}_c^*} v_{y \rightarrow y'} + \sum_{y \in \mathcal{C}_Y} \right) (Y - y) \end{aligned}$$

$$\begin{aligned}
 & + \left(\sum_{y \rightarrow y' \in \mathcal{R}_Y^* \setminus \mathcal{R}_c^*} v_{y \rightarrow y'} + \frac{1}{2} \sum_{y \rightarrow y' \in \mathcal{R}_Y^* \cap \mathcal{R}_c^*} v_{y \rightarrow y'} + \sum_{y' \in \mathcal{C}_Y} \right) (y' - Y) \\
 = & \sum_{y \rightarrow y' \in \mathcal{R}} w_{y \rightarrow y'} (y' - y),
 \end{aligned}$$

where

$$w_{y \rightarrow y'} = v_{y \rightarrow y'}, \quad \text{if } y \rightarrow y' \in \mathcal{R}_c^* \setminus \mathcal{R}_Y^*, \tag{19}$$

$$w_{y \rightarrow y'} = \frac{v_{y \rightarrow y'}}{2}, \quad \text{if } y \rightarrow y' \in \mathcal{R}_Y^* \cap \mathcal{R}_c^*, \tag{20}$$

$$w_{y \rightarrow Y} = \begin{cases} \left(\sum_{y \rightarrow y' \in \mathcal{R}_Y^* \setminus \mathcal{R}_c^*} + \frac{1}{2} \sum_{y \rightarrow y' \in \mathcal{R}_Y^* \cap \mathcal{R}_c^*} \right) v_{y \rightarrow y'} + 1, & \text{if } y \in \mathcal{C}_Y \\ \left(\sum_{y \rightarrow y' \in \mathcal{R}_Y^* \setminus \mathcal{R}_c^*} + \frac{1}{2} \sum_{y \rightarrow y' \in \mathcal{R}_Y^* \cap \mathcal{R}_c^*} \right) v_{y \rightarrow y'}, & \text{if } y \notin \mathcal{C}_Y, \end{cases} \tag{21}$$

and, similarly,

$$w_{Y \rightarrow y'} = \begin{cases} \left(\sum_{y \rightarrow y' \in \mathcal{R}_Y^* \setminus \mathcal{R}_c^*} + \frac{1}{2} \sum_{y \rightarrow y' \in \mathcal{R}_Y^* \cap \mathcal{R}_c^*} \right) v_{y \rightarrow y'} + 1, & \text{if } y' \in \mathcal{C}_Y \\ \left(\sum_{y \rightarrow y' \in \mathcal{R}_Y^* \setminus \mathcal{R}_c^*} + \frac{1}{2} \sum_{y \rightarrow y' \in \mathcal{R}_Y^* \cap \mathcal{R}_c^*} \right) v_{y \rightarrow y'}, & \text{if } y' \notin \mathcal{C}_Y. \end{cases} \tag{22}$$

Since G^* is obtained from G by the removal of a single intermediate Y , every reaction in \mathcal{R} is of the form $y \rightarrow y'$, $y \rightarrow Y$ or $Y \rightarrow y'$ for some $y \rightarrow y' \in \mathcal{R}^*$, or of the form $y \rightarrow Y$ or $Y \rightarrow y$ for some $y \in \mathcal{C}_Y$. Thus, (19)–(22) above yield $w_{y \rightarrow y'} > 0$ for every $y \rightarrow y' \in \mathcal{R}$, and we conclude that G is consistent.

(\Leftarrow) Now suppose G is consistent, so that there exist $w_{y \rightarrow y'} > 0$, for all $y \rightarrow y' \in \mathcal{R}$, such that

$$\sum_{y \rightarrow y' \in \mathcal{R}} w_{y \rightarrow y'} (y' - y) = 0. \tag{23}$$

We partition the set \mathcal{R} of reactions of G as the (disjoint) union

$$\mathcal{R} = \mathcal{R}_c^* \cup \mathcal{R}_{\rightarrow Y} \cup \mathcal{R}_{Y \rightarrow},$$

where \mathcal{R}_c^* is defined as in the first part of the proof, $\mathcal{R}_{\rightarrow Y}$ is the subset of \mathcal{R} of reactions having Y as a product, and $\mathcal{R}_{Y \rightarrow}$ is the subset of \mathcal{R} of reactions having Y as

a reactant. Observe that Y is linearly independent with each complex in \mathcal{C}^* . Combining all coefficients of Y in (23), we obtain

$$\sum_{y \rightarrow Y \in \mathcal{R}_{\rightarrow Y}} w_{y \rightarrow Y} - \sum_{Y \rightarrow y' \in \mathcal{R}_{Y \rightarrow}} w_{Y \rightarrow y'} = 0,$$

and so

$$\sum_{Y \rightarrow y' \in \mathcal{R}_{Y \rightarrow}} w_{Y \rightarrow y'} y' - \sum_{y \rightarrow Y \in \mathcal{R}_{\rightarrow Y}} w_{y \rightarrow Y} y + \sum_{y \rightarrow y' \in \mathcal{R}_c^*} w_{y \rightarrow y'} (y' - y) = 0. \quad (24)$$

Set

$$V := \sum_{y \rightarrow Y \in \mathcal{R}_{\rightarrow Y}} w_{y \rightarrow Y} = \sum_{Y \rightarrow y' \in \mathcal{R}_{Y \rightarrow}} w_{Y \rightarrow y'}.$$

We have

$$\sum_{Y \rightarrow y' \in \mathcal{R}_{Y \rightarrow}} w_{Y \rightarrow y'} y' = \sum_{y \rightarrow Y \in \mathcal{R}_{\rightarrow Y}} \sum_{Y \rightarrow y' \in \mathcal{R}_{Y \rightarrow}} \frac{w_{y \rightarrow Y} w_{Y \rightarrow y'}}{V} y'$$

and

$$\sum_{y \rightarrow Y \in \mathcal{R}_{\rightarrow Y}} w_{y \rightarrow Y} y = \sum_{y \rightarrow Y \in \mathcal{R}_{\rightarrow Y}} \sum_{Y \rightarrow y' \in \mathcal{R}_{Y \rightarrow}} \frac{w_{y \rightarrow Y} w_{Y \rightarrow y'}}{V} y.$$

Plugging these last two identities into (24), we may rewrite it as

$$\sum_{y \rightarrow y' \in \mathcal{R}^*} v_{y \rightarrow y'} (y' - y) = 0,$$

where

$$v_{y \rightarrow y'} := \begin{cases} w_{y \rightarrow y'} & \text{if } y \rightarrow y' \in \mathcal{R}_c^* \setminus \mathcal{R}_{\rightarrow}^*, \\ w_{y \rightarrow y'} + \frac{w_{y \rightarrow Y} w_{Y \rightarrow y'}}{V} & \text{if } y \rightarrow y' \in \mathcal{R}_c^* \cap \mathcal{R}_{\rightarrow}^*, \\ \frac{w_{y \rightarrow Y} w_{Y \rightarrow y'}}{V} & \text{if } \mathcal{R}_c^* \setminus \mathcal{R}_{\rightarrow}^*. \end{cases}$$

In particular, $v_{y \rightarrow y'} > 0$ for every $y \rightarrow y' \in \mathcal{R}^*$, showing that G^* is consistent. \square

Proof of Theorem 1(iii). Let $G_p := G$ and, for $i = p, \dots, 1$, let G_{i-1} be the reaction network obtained from G_i by the removal of the set of intermediates $\{Y_i\}$. By Lemma 1, $G_0 = G^*$. Iterating Lemma 12, we conclude that G^* is consistent if, and only if G is consistent. \square

5.2 Catalysts

Suppose $G^* = (S^*, C^*, \mathcal{R}^*)$ is the reduction of $G = (S, C, \mathcal{R})$ by the removal of a set of catalysts \mathcal{E} . Let $G_{\mathcal{E}} = (\mathcal{S}_{\mathcal{E}}, \mathcal{C}_{\mathcal{E}}, \mathcal{R}_{\mathcal{E}})$ be the subnetwork of G implied by \mathcal{E} , and write

$$S^* = \{S_1^*, \dots, S_n^*\}, \quad \mathcal{S}_{\mathcal{E}} = \{E_1^a, \dots, E_{q_a}^a\}, \quad \text{and} \quad \mathcal{E} \setminus \mathcal{S}_{\mathcal{E}} = \{E_1^u, \dots, E_{q_u}^u\}.$$

Thus

$$S = \{S_1^*, \dots, S_n^*, E_1^a, \dots, E_{q_a}^a, E_1^u, \dots, E_{q_u}^u\}.$$

These are the orderings we shall assume on the species whenever working with the stoichiometric matrices or stoichiometric subspaces of G , G^* or $G_{\mathcal{E}}$.

Conservation laws

Lemma 13 *Suppose $G^* = (S^*, C^*, \mathcal{R}^*)$ is the reduction of $G = (S, C, \mathcal{R})$ by the removal of a set of catalysts \mathcal{E} . Then*

$$\Gamma^{\perp} = (\Gamma^*)^{\perp} \times \Gamma_{\mathcal{E}}^{\perp} \times \mathbb{R}^{q_u} \subseteq \mathbb{R}^{n+q_a+q_u}.$$

Proof Write

$$\mathcal{R}^* = \{R_1^*, \dots, R_m^*\},$$

and set $\mathcal{R}_S := \mathcal{R} \setminus \mathcal{R}_{\mathcal{E}}$. For each $j \in [m]$, let $R_1^{(j)}, \dots, R_{k_j}^{(j)} \in \mathcal{R}_S$ be the reactions of G from which R_j^* is obtained by removing the catalysts from both reactant and product in the construction of G^* . Write

$$\mathcal{R}_{\mathcal{E}} = \{R_1^{\mathcal{E}}, \dots, R_{m_{\mathcal{E}}}^{\mathcal{E}}\}.$$

Thus

$$\mathcal{R} = \{R_1^{(1)}, \dots, R_{k_1}^{(1)}, \dots, R_1^{(m)}, \dots, R_{k_m}^{(m)}, R_1^{\mathcal{E}}, \dots, R_{m_{\mathcal{E}}}^{\mathcal{E}}\}.$$

With these orderings on \mathcal{R} , \mathcal{R}^* and $\mathcal{R}_{\mathcal{E}}$, we may express the stoichiometric matrix N of G as

$$N = \begin{bmatrix} N' & 0 \\ 0 & N_{\mathcal{E}} \\ 0 & 0 \end{bmatrix}, \tag{25}$$

where N' has n rows, $k_1 + \dots + k_m$ columns, and has the property that the columns corresponding to $R_1^{(j)}, \dots, R_{k_j}^{(j)}$ are equal to the j^{th} column of N^* , for $j = 1, \dots, m$, where $N_{\mathcal{E}}$ is the stoichiometric matrix of $G_{\mathcal{E}}$, and where the bottom q_u rows are zero.

Given $\omega^* \in \mathbb{R}^n$, we have $(\omega^*)^T N^* = 0$ if, and only if $(\omega^*)^T N' = 0$. Hence, given

$$\omega = (\omega^*, \omega_{\mathcal{E}}, x) \in \mathbb{R}^{n+q_a+q_u},$$

we have $\omega^T N = 0$ if, and only if $(\omega^*)^T N^* = 0$, and $\omega_{\mathcal{E}}^T N_{\mathcal{E}} = 0$. This proves the lemma. \square

Proof of Theorem 2(iv). (\Rightarrow) If $\omega = (\omega^*, \omega_{\mathcal{E}}, x)$ is a strictly positive conservation law of G , then ω^* is a strictly positive conservation law of G^* by Lemma 13.

(\Leftarrow) If ω^* is a strictly positive conservation law of G^* , then choose any $x \in \mathbb{R}_{>0}^{q_u}$ and a strictly positive vector $\omega_{\mathcal{E}} \in \Gamma_{\mathcal{E}}^{\perp}$, which exists since $G_{\mathcal{E}}$ is conservative. Then $\omega = (\omega^*, \omega_{\mathcal{E}}, x)$ is a strictly positive conservation law of G by Lemma 13. \square

Siphons

Lemma 14 *Suppose $G^* = (\mathcal{S}^*, \mathcal{C}^*, \mathcal{R}^*)$ is the reduction of $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ by the removal of a set of catalysts \mathcal{E} . Let Σ be a minimal siphon of G . Then one of the three possibilities below is true.*

- (i) $\Sigma \subseteq \mathcal{S}^*$, and it is a minimal siphon of G^* .
- (ii) $\Sigma \subseteq \mathcal{S}_{\mathcal{E}}$, and it is a minimal siphon of $G_{\mathcal{E}}$.
- (iii) $\Sigma = \{E\}$ for some $E \in \mathcal{E} \setminus \mathcal{S}_{\mathcal{E}}$.

Proof Suppose $\Sigma \cap (\mathcal{E} \setminus \mathcal{S}_{\mathcal{E}}) \neq \emptyset$. Pick any $E \in (\mathcal{E} \setminus \mathcal{S}_{\mathcal{E}})$. Then E appears as a reactant in every reaction in which it also appears as a product. We conclude that $\{E\}$ is a siphon, which must then be minimal. It follows that (iii) holds.

Now suppose $\Sigma \cap (\mathcal{E} \setminus \mathcal{S}_{\mathcal{E}}) = \emptyset$. We have two possibilities.

If $\Sigma_{\mathcal{E}} := \Sigma \cap \mathcal{S}_{\mathcal{E}} \neq \emptyset$, then it is a siphon of G . Indeed, pick any $S' \in \Sigma_{\mathcal{E}}$, and let $y \rightarrow y' \in \mathcal{R}$ be any reaction having S' as one of its products. Since Σ is a siphon of G , $y \rightarrow y'$ must have one of its reactants S in Σ . If $y \rightarrow y' \notin \mathcal{R}_{\mathcal{E}}$, then S' is also a reactant in $y \rightarrow y'$, and we may assume without loss of generality that $S = S'$. If $y \rightarrow y' \in \mathcal{R}_{\mathcal{E}}$, then we have $y = S$ and $S \in \mathcal{S}_{\mathcal{E}}$. In either case, $y \rightarrow y'$ has a reactant S in $\Sigma_{\mathcal{E}}$. This shows $\Sigma_{\mathcal{E}}$ is a siphon of $G_{\mathcal{E}}$. By the minimality assumption, we must have $\Sigma = \Sigma_{\mathcal{E}} \subseteq \mathcal{S}_{\mathcal{E}}$. Since every siphon of $G_{\mathcal{E}}$ is also a siphon of G , we conclude that $\Sigma_{\mathcal{E}}$ must be a minimal siphon of $G_{\mathcal{E}}$.

If $\Sigma \cap \mathcal{E} = \emptyset$, then $\Sigma \subseteq \mathcal{S}^*$. It follows from the construction of G^* that Σ is a minimal siphon of G^* . \square

Drainable and self-replicable siphons

Proof of Theorem 2(i). (\Leftarrow) In virtue of (C1), any drainable (respectively, self-replicable) siphon of G^* is also a drainable (respectively, self-replicable) siphon of G .

(\Rightarrow) Suppose Σ is a drainable or self-replicable siphon of G . It is evident from Definition 5 that any siphon of G contained in Σ is also drainable or self-replicable. Therefore, we may assume without loss of generality that Σ is minimal. By Lemma 14, either $\Sigma \subseteq \mathcal{S}^*$, or $\Sigma \subseteq \mathcal{S}_{\mathcal{E}}$, or $\Sigma = \{E\}$ for some $E \in \mathcal{E} \setminus \mathcal{S}_{\mathcal{E}}$.

If $\Sigma = \{E\}$ for some $E \in \mathcal{E} \setminus \mathcal{S}_{\mathcal{E}}$, then the row of N corresponding to E is identically zero, and so the vector of $\mathbb{R}_{\geq 0}^{q+p}$ having its entry corresponding to E equal to 1 and all other entries equal to zero is a P-semiflow supported in Σ . In particular, Σ is not critical, therefore neither drainable nor self-replicable by Proposition 1(i).

If $\Sigma \subseteq \mathcal{S}_{\mathcal{E}}$, then it follows from (C2) that Σ cannot be drainable or self-replicable either.

So, it must be the case that $\Sigma \subseteq \mathcal{S}^*$. Consider a reaction $y \rightarrow y'$ in \mathcal{R}^* and the reaction $\tilde{y} \rightarrow \tilde{y}'$ in \mathcal{R} giving rise to it. Then the i -th coordinate of the vectors $y' - y$ and $\tilde{y}' - \tilde{y}$ agree for all $i \in [q]$. Using this observation, we conclude that $\Sigma^* := \Sigma \cap \mathcal{S}^* = \Sigma$ is a drainable or self-replicable siphon of G^* . \square

Consistency

Proof of Theorem 2(iii). We write the stoichiometric matrix N of G as in the proof of Lemma 13.

First suppose that G is consistent, and let $v \gg 0$ be such that $Nv = 0$. Thus, $N'v' = 0$, where $v' := (v_1, \dots, v_{k_1+\dots+k_m}) \gg 0$. Defining $v^* \in \mathbb{R}^m$ by setting

$$v_j^* := v_{k_1+\dots+k_{j-1}+1} + \dots + v_{k_1+\dots+k_{j-1}+k_j},$$

we then get $v^* \gg 0$ and $N^*v^* = 0$, showing that G^* is consistent.

Now suppose G^* is consistent and $G_{\mathcal{E}}$ is conservative. Let $v^* \gg 0$ be any vector such that $N^*v^* = 0$. Set

$$v'_j := \frac{1}{k_j}(v_j, \dots, v_j) \in \mathbb{R}^{k_j}, \quad j = 1, \dots, m,$$

and then set

$$v' := (v'_1, \dots, v'_m) \in \mathbb{R}^{k_1+\dots+k_m}.$$

Then $N'v' = 0$. Since $G_{\mathcal{E}}$ does not have drainable siphons and is conservative, it follows from Proposition 2 that $G_{\mathcal{E}}$ is consistent. Let $v_{\mathcal{E}} \gg 0$ be such that $N_{\mathcal{E}}v_{\mathcal{E}} = 0$. Setting $v := (v', v_{\mathcal{E}})$, we have $v \gg 0$, and $Nv = 0$, proving that G is consistent. \square

5.3 Uniqueness of the primitive reduction

To prove Theorem 3, we will use induction on the number of species. We start with a few observations and auxiliary results.

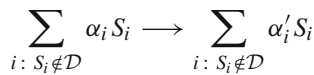
In this subsection we will use the following notation. Given a reaction network $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ and a set $\mathcal{A} \subseteq \mathcal{S}$ of intermediates or catalysts of G , we will denote by $G^*_{\mathcal{A}} = (\mathcal{S}^*_{\mathcal{A}}, \mathcal{C}^*_{\mathcal{A}}, \mathcal{R}^*_{\mathcal{A}})$ the reaction network obtained from G by the removal of \mathcal{A} (as a set of intermediates or catalysts, whichever happens to be the case). Given another set $\mathcal{B} \subseteq \mathcal{S}$ of intermediates (respectively, catalysts) of G , note that $\mathcal{B} \setminus \mathcal{A}$ is either empty, or else also a set of intermediates (respectively, catalysts) of $G^*_{\mathcal{A}}$. We then denote by

$G_{AB}^* = (S_{AB}^*, C_{AB}^*, \mathcal{R}_{AB}^*)$ the reaction network obtained from $G_{\mathcal{A}}^*$ by the removal of $B \setminus \mathcal{A}$.

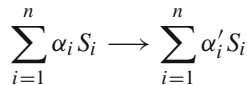
Lemma 15 *Given a reaction network $G = (S, \mathcal{C}, \mathcal{R})$, suppose $\mathcal{A}, \mathcal{B} \subseteq S$ are two sets of intermediates or two sets of catalysts of G . Let $\mathcal{D} := \mathcal{A} \cup \mathcal{B}$. Then $G_{\mathcal{D}}^* = G_{\mathcal{AB}}^* = G_{\mathcal{BA}}^*$.*

Proof If \mathcal{A} and \mathcal{B} are both sets of intermediates, then the result follows from Lemma 1. Removing first the intermediates in \mathcal{A} one at a time, then removing the intermediates in $B \setminus \mathcal{A}$ yields $G_{\mathcal{AB}}^*$. The analogue procedure starting with the intermediates in \mathcal{B} yields $G_{\mathcal{BA}}^*$. One then concludes by the same lemma that $G_{\mathcal{AB}}^* = G_{\mathcal{BA}}^* = G_{\mathcal{D}}^*$.

Now suppose \mathcal{A} and \mathcal{B} are both sets of catalysts. Then both $\mathcal{R}_{\mathcal{AB}}^*$ and $\mathcal{R}_{\mathcal{BA}}^*$ consist of the reactions



such that



belongs to \mathcal{R} , and $\alpha_{i_0} > 0$ or $\alpha'_{i_0} > 0$ for some $i_0 \in [n]$ such that $S_{i_0} \notin \mathcal{D}$. This shows $\mathcal{R}_{\mathcal{AB}}^* = \mathcal{R}_{\mathcal{BA}}^* = \mathcal{R}_{\mathcal{D}}^*$, establishing the result. \square

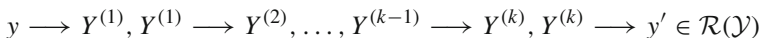
Finally, the removal of a set of catalysts also commutes with the removal of a set of intermediates, in the following sense.

Lemma 16 *Let $G = (S, \mathcal{C}, \mathcal{R})$ be a reaction network, $\mathcal{Y} \subseteq S$ be a set of intermediates, and $\mathcal{E} \subseteq S$ be a set of catalysts. Then $G_{\mathcal{Y}\mathcal{E}}^* = G_{\mathcal{E}\mathcal{Y}}^*$.*

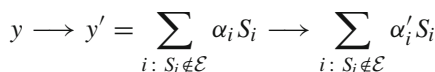
Proof Let $\mathcal{R}(\mathcal{Y})$ be the subset of reactions $c \rightarrow c' \in \mathcal{R}$ having some intermediate in \mathcal{Y} as a reactant or product. It follows directly from property (I2) of intermediates that $\mathcal{R}(\mathcal{Y})$ is the subset of reactions $c \rightarrow c' \in \mathcal{R}$ which appear in some reaction path



such that $y, y' \in \mathcal{C} \setminus \mathcal{Y}$ and $Y^{(1)}, \dots, Y^{(k)} \in \mathcal{Y}$. Let $\mathcal{R}(\mathcal{E})$ be the subset of reactions $c \rightarrow c' \in \mathcal{R}$ having some catalyst in \mathcal{E} as both reactant and product. Observe that $\mathcal{R}(\mathcal{Y}) \cap \mathcal{R}(\mathcal{E}) = \emptyset$. Thus, both $\mathcal{R}_{\mathcal{Y}\mathcal{E}}^*$ and $\mathcal{R}_{\mathcal{E}\mathcal{Y}}^*$ consist of the set of reactions $y \rightarrow y'$ such that $y \rightarrow y' \in \mathcal{R}(\mathcal{R}(\mathcal{Y}) \cup \mathcal{R}(\mathcal{E}) \cup \mathcal{R}_{\mathcal{E}})$, or



for some $y, y' \in \mathcal{C} \setminus \mathcal{Y}$ and $Y^{(1)}, \dots, Y^{(k)} \in \mathcal{Y}$, or



for some

$$\sum_{i=1}^n \alpha_i S_i \longrightarrow \sum_{i=1}^n \alpha'_i S_i$$

belonging to $\mathcal{R}(\mathcal{E})$. □

Proof of Theorem 3. We use induction on the number of species. A reaction network with zero species (the empty network) is already primitive, so, in this case, the result holds vacuously.

Now suppose the result holds for reaction networks with up to $n \geq 0$ species, and let $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network with $|\mathcal{S}| = n + 1$ species. If G is already primitive, then it is automatically its unique primitive reduction, in which case we have nothing left to prove. So, we may assume G is not primitive.

Let $\mathcal{A}, \mathcal{B} \subseteq \mathcal{S}$ be sets of intermediates or catalysts of G such that $\mathcal{A} \neq \mathcal{B}$. By the induction hypothesis, $G_{\mathcal{A}}^*$ and $G_{\mathcal{B}}^*$ have unique primitive reductions, respectively, $G_{\mathcal{A}}^{**}$ and $G_{\mathcal{B}}^{**}$. We want to show that $G_{\mathcal{A}}^{**} = G_{\mathcal{B}}^{**}$.

Let $G_{\mathcal{A}\mathcal{B}}^{**}$ (respectively, $G_{\mathcal{B}\mathcal{A}}^{**}$) be the primitive reduction of $G_{\mathcal{A}\mathcal{B}}^*$ (respectively, $G_{\mathcal{B}\mathcal{A}}^*$). Note that $G_{\mathcal{A}\mathcal{B}}^{**} = G_{\mathcal{A}}^{**}$ and $G_{\mathcal{B}\mathcal{A}}^{**} = G_{\mathcal{B}}^{**}$. By Lemmas 15 and 16, $G_{\mathcal{A}\mathcal{B}}^* = G_{\mathcal{B}\mathcal{A}}^*$, and hence $G_{\mathcal{A}}^{**} = G_{\mathcal{B}}^{**}$. □

Acknowledgments Elisenda Feliu, Michael Marcondes de Freitas and Carsten Wiuf acknowledge funding from the Danish Research Council of Independent Research. We would also like to thank the reviewers for their helpful comments and suggestions, which have greatly improved this work.

Appendix: Technical results

Proof of Proposition 2

Proposition 2(i). We prove that if G is persistent, then it is bounded-persistent. Take any $s_0 \gg 0$. If $\omega(s_0) = \emptyset$, then we have nothing to prove. So, suppose $\omega(s_0) \neq \emptyset$. Choose any $s \in \omega(s_0)$, and a sequence $(t_k)_{k \in \mathbb{N}}$ going to infinity in $\mathbb{R}_{\geq 0}$ such that

$$\lim_{k \rightarrow \infty} \sigma(t_k, s_0) = s.$$

Then

$$s_i = \liminf_{k \rightarrow \infty} \sigma_i(t_k, s_0) \geq \liminf_{t \rightarrow \infty} \sigma_i(t, s_0) > 0, \quad \forall i \in [n].$$

In particular, $s \notin \partial \mathbb{R}_{\geq 0}^n$. Thus, $\omega(s_0) \cap \partial \mathbb{R}_{\geq 0}^n = \emptyset$. □

Proposition 2(ii). The converse of Proposition 2(i) is not true. However, (4) holds for bounded trajectories of bounded-persistent networks—hence the terminology.

Since each stoichiometric compatibility class of a conservative network is compact (Horn and Jackson 1972, Appendix 1), every solution of (3) is bounded. The proof of Proposition 2(ii) then follows from the next lemma.

Lemma 17 *Suppose a solution $\sigma(\cdot, s_0): \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}^n$ of a bounded-persistent reaction network is bounded. Then*

$$\liminf_{t \rightarrow \infty} \sigma_i(t, s_0) > 0, \quad \forall i \in [n]. \tag{26}$$

Proof Suppose on contrary that

$$\liminf_{t \rightarrow \infty} \sigma_{i_0}(t, s_0) = 0$$

for some $i_0 \in [n]$. Then

$$\lim_{k \rightarrow \infty} \sigma_{i_0}(t_k, s_0) = 0$$

along some sequence $(t_k)_{k \in \mathbb{N}}$ going to infinity in $\mathbb{R}_{\geq 0}$. In virtue of boundedness, by passing into a subsequence, if necessary, we may assume without loss of generality that $(\sigma(t_k, s_0))_{k \in \mathbb{N}}$ converges, say,

$$\lim_{k \rightarrow \infty} \sigma(t_k, s_0) = s_\infty.$$

We have $s_\infty \in \omega(s_0)$ by definition. But since the i_0^{th} coordinate of s_∞ is zero, we conclude that $s_\infty \in \partial \mathbb{R}_{\geq 0}^n$ also. This contradicts the bounded-persistence hypothesis that $\omega(s_0) \cap \partial \mathbb{R}_{\geq 0}^n = \emptyset$. Thus, (26) must hold. □

Proposition 2(iii). See (Angeli et al. 2007, Theorem 1). □

Proposition 2(iv). The same argument as in Deshpande and Gopalkrishnan (2014), Theorem 6.2 works under our weaker assumptions on the reaction rates. □

Proposition 2(v). We define the zero coordinate set of a point $s \in \mathbb{R}_{\geq 0}^n$, with respect to some given reaction network G , as the set

$$Z(s) := \{S_i \in \mathcal{S} \mid s_i = 0\} = \mathcal{S} \setminus \text{supp } s.$$

Thus, a point $s \in \mathbb{R}_{\geq 0}^n$ is a boundary steady state if, and only if $Z(s) \neq \emptyset$.

Let s_0 be a boundary steady state of G . By Lemma 18 below and our hypothesis, the zero coordinate set $Z(s_0)$ of s_0 is a noncritical siphon. It follows by the equivalence between items 1 and 3 in Deshpande and Gopalkrishnan (2014), Theorem 3.7 that $(s_0 + S) \cap \mathbb{R}_{\geq 0}^n \neq \emptyset$.

The next lemma was proved in Shiu and Sturmfels (2010) for mass-action kinetics. The same argument holds under (r2), and we provide the details for the sake of completeness.

Lemma 18 *Let G be a reaction network. If s_0 is a boundary steady state, then $Z(s_0)$ is a siphon.*

Proof Pick any $S_i \in Z(s_0)$. Consider the set \mathcal{J}_i of indices $j \in [m]$ such that R_j is a reaction having S_i as one of its products, but not one of its reactants; that is,

$$\mathcal{J}_i := \{j \in [m] \mid \alpha'_{ij} > 0 \text{ and } \alpha_{ij} = 0\}.$$

If $\mathcal{J}_i \neq \emptyset$, we need to show that $Z(s_0)$ contains some species in the reactant of each R_j such that $j \in \mathcal{J}_i$. Since s_0 is a steady state, we have

$$\sum_{j=1}^m (\alpha'_{ij} - \alpha_{ij})r_j(s_0) = 0. \tag{27}$$

For each $j \notin \mathcal{J}_i$, we either have $\alpha_{ij} > 0$ (in which case $r_j(s_0) = 0$ by (r2) since $(s_0)_i = 0$) or $\alpha'_{ij} = \alpha_{ij} = 0$. Hence the sum in (27) can be simplified as

$$\sum_{j \in \mathcal{J}_i} \alpha'_{ij}r_j(s_0) = 0.$$

Since $\alpha'_{ij} > 0$ for every $j \in \mathcal{J}_i$ by construction, we conclude that $r_j(s_0) = 0$ for all $j \in \mathcal{J}_i$. It then follows from (r2) that $\alpha_{i(j)j} > 0$ for some $i(j) \in [n]$ such that $S_{i(j)} \in Z(s_0)$, that is, one of the reactants of R_j belongs to $Z(s_0)$ for each $j \in \mathcal{J}_i$. This completes the proof that $Z(s_0)$ is a siphon. \square

Drainable and self-replicable siphons

The next result shows that the concepts of drainable and self-replicable sets in Definition 5 are, respectively, equivalent to the concepts of drainable and self-replicable sets in Deshpande and Gopalkrishnan (2014), Definition 3.1 (called here *DG-drainable* and *DG-self-replicable*).

Given a reaction network $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$, we define a *G-reaction pathway* to be any sequence $y(0), y(1), \dots, y(k) \in \mathbb{R}_{\geq 0}^n$ such that

$$\begin{aligned} y(0) &= y_1 + w_1, \\ y(j) &= y'_j + w_j = y_{j+1} + w_{j+1}, \quad j = 1, \dots, k - 1, \\ y(k) &= y'_k + w_k, \end{aligned} \tag{28}$$

for some $y_1, y'_1, w_1, \dots, y_k, y'_k, w_k \in \mathbb{R}_{\geq 0}^n$ such that $y_1 \rightarrow y_1, \dots, y_k \rightarrow y'_k \in \mathcal{R}$. Note that

$$y(k) - y(0) = \sum_{j=1}^k (y(j) - y(j-1)) = \sum_{j=1}^k (y'_j + w_j - y_j - w_j) = \sum_{j=1}^k (y'_j - y_j). \tag{29}$$

A nonempty subset $\Sigma \subseteq \mathcal{S}$ is said to be *DG-drainable* (respectively, *DG-self-replicable*) if there exists a *G-reaction pathway* $y(0), y(1), \dots, y(k)$ such that $(y(k) - y(0))_i < 0$ (respectively, $(y(k) - y(0))_i > 0$), for every $i \in [n]$ such that $S_i \in \Sigma$.

Proposition 6 Let $G = (S, C, \mathcal{R})$ be a reaction network. A subset of S is drainable (respectively, self-replicable) if, and only if it is DG-drainable (respectively, DG-self-replicable).

Proof (\Leftarrow) Follows from (29) and Definition 5.

(\Rightarrow) Let $y_1 \rightarrow y'_1, \dots, y_k \rightarrow y'_k \in \mathcal{R}$ be any sequence of reactions. Define, iteratively,

$$w_1 := y_2 + \dots + y_k, \quad \text{and} \quad w_{j+1} := y'_j + w_j - y_{j+1}, \quad j = 1, \dots, k-1.$$

By construction, $w_j \in \mathbb{R}_{\geq 0}^n$ for all $j = 1, \dots, k$, and $y_{j+1} + w_{j+1} = y'_j + w_j$ for all $j = 1, \dots, k-1$. We can construct a G-reaction pathway $y(0), y(1), \dots, y(k) \in \mathbb{R}_{\geq 0}^n$ from $y_1, y'_1, w_1, \dots, y_k, y'_k, w_k$ using (28). The implication now follows again from (29) and Definition 5. \square

References

- Amann H (1990) Ordinary differential equations: an introduction to nonlinear analysis. De Gruyter studies in mathematics. de Gruyter, Berlin
- Anderson DF (2008) Global asymptotic stability for a class of nonlinear chemical equations. *SIAM J Appl Math* 68(5):1464–1476
- Anderson DF (2011) A proof of the global attractor conjecture in the single linkage class case. *SIAM J Appl Math* 71(4):1487–1508
- Angeli D, De Leenheer P, Sontag ED (2007) A Petri net approach to the study of persistence in chemical reaction networks. *Math Biosci* 210(2):598–618
- Angeli D, De Leenheer P, Sontag ED (2010) Graph-theoretic characterizations of monotonicity of chemical networks in reaction coordinates. *J Math Biol* 61(4):581–616
- Angeli D, De Leenheer P, Sontag ED (2011) Persistence results for chemical reaction networks with time-dependent kinetics and no global conservation laws. *SIAM J Appl Math* 71(1):128–146
- Cordone R, Ferrarini L, Piroddi L (2005) Enumeration algorithms for minimal siphons in petri nets based on place constraints. *IEEE Trans Syst Man Cybern Part A Syst Hum* 35(6):844–854
- Craciun G, Nazarov F, Pantea C (2013) Persistence and permanence of mass-action and power-law dynamical systems. *SIAM J Appl Math* 73(1):305–329
- Deshpande A, Gopalkrishnan M (2014) Autocatalysis in reaction networks. *Bull Math Biol* 76(10):2570–2595
- Feinberg M (1995) The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch Rational Mech Anal* 132(4):311–370
- Feinberg M (1980) Lectures on chemical reaction networks. Notes of lectures given at the Mathematics Research Centre, University of Wisconsin-Madison, in the autumn of 1979. <http://www.crnt.osu.edu/LecturesOnReactionNetworks>
- Feinberg M (1987) Chemical reaction network structure and the stability of complex isothermal reactors. I. the deficiency zero and deficiency one theorems. *Chem Eng Sci* 42(10):2229–2268
- Feliu E, Wiuf C (2013) Simplifying biochemical models with intermediate species. *J R Soc Interface* 10(87):20130484
- Gnacadjia G (2011) Reachability, persistence, and constructive chemical reaction networks (part I): reachability approach to the persistence of chemical reaction networks. *J Math Chem* 49(10):2117–2136
- Gnacadjia G (2011) Reachability, persistence, and constructive chemical reaction networks (part III): a mathematical formalism for binary enzymatic networks and application to persistence. *J Math Chem* 49(10):2158–2176
- Gopalkrishnan M, Miller E, Shiu A (2014) A geometric approach to the global attractor conjecture. *SIAM J Appl Dyn Syst* 13(2):758–797
- Gunawardena J (2003) Chemical reaction network theory for in-silico biologists. <http://vcp.med.harvard.edu/papers/crnt>

- Horn F, Jackson R (1972) General mass action kinetics. *Arch Rational Mech Anal* 47:81–116
- MacLean AL, Rosen Z, Byrne HM, Harrington HA (2015) Parameter-free methods distinguish Wnt pathway models and guide design of experiments. *PNAS* 112(9):2652–2657
- Murray JD (2002) *Mathematical biology*. Springer, Berlin
- Nguyen LK, Muñoz-García J, Maccario H, Ciechanover A, Kolch W, Kholodenko BN (2011) Switches, excitable responses and oscillations in the Ring1B/Bmi1 ubiquitination system. *PLoS Comput Biol* 7(12):e1002317
- Shiu A, Sturmfels B (2010) Siphons in chemical reaction networks. *Bull Math Biol* 72(6):1448–1463
- Siegel-Gaskins D, Franco E, Zhou T, Murray RM (2015) An analytical approach to bistable biological circuit discrimination using real algebraic geometry. *J Royal Soc Interface* 12(108). doi:[10.1098/rsif.2015.0288](https://doi.org/10.1098/rsif.2015.0288)
- Smith HL, Thieme HR (2011) *Dynamical systems and population persistence*, volume 118 of graduate studies in mathematics. American Mathematical Society, Providence
- Sontag ED (2001) Structure and stability of certain chemical networks and applications to the kinetic proofreading model of T-cell receptor signal transduction. *Inst Electr Electron Eng Trans Autom Control* 46(7):1028–1047
- Sontag ED (2001) Structure and stability of certain chemical networks and applications to the kinetic proofreading model of T-cell receptor signal transduction. *IEEE Trans Autom Control* 46(7):1028–1047
- Thomson M, Gunawardena J (2009) The rational parameterisation theorem for multisite post-translational modification systems. *J Theor Biol* 261(4):626–636