The unraveling of balanced complexes in metabolic networks

Damoun Langary^{1,2}, Anika Küken², Zoran Nikoloski^{1,2}

¹Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany

²Bioinformatics, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

Abstract

Balanced complexes in biochemical networks are at core of several theoretical and computational approaches that make statements about the properties of the steady states supported by the network. Recent computational approaches have employed balanced complexes to reduce metabolic networks, while ensuring preservation of particular steady-state properties; however, the underlying factors leading to the formation of balanced complexes have not been studied, yet. Here, we present a number of factorizations providing insights in mechanisms that lead to the origins of the corresponding balanced complexes. The proposed factorizations enable us to categorize balanced complexes into four distinct classes, each with specific origins and characteristics. They also provide the means to efficiently determine if a balanced complex in large-scale networks belongs to a particular class from the categorization. The results are obtained under very general conditions and irrespective of the network kinetics, rendering them broadly applicable across variety of network models. Application of the categorization shows that all classes of balanced complexes are present in large-scale metabolic models across all kingdoms of life, therefore paving the way to study their relevance with respect to different properties of steady states supported by these networks.

Keywords: balanced complexes, steady state analysis, constraint-based modeling, chemical reaction network theory, metabolic networks

Highlights

- Balanced complexes are ubiquitous in real-world networks, and facilitate insights in steady state flux phenotypes in large-scale metabolic networks.
- Novel factorizations are proposed that explain the formation of balanced complexes and enable their categorization.
- The results also provide a computationally-efficient tool for the identification of balanced complexes in large-scale networks.
- Examination of metabolic network models across all kingdoms of life shows that all categories naturally arise in this type of networks.

1 Introduction

The last two decades have witnessed the generation of large-scale metabolic networks along with the development of computational approaches within the constraint-based modeling framework that facilitate insights in the genotype – phenotype map (Bordbar, et al., 2014). Metabolic networks include the entirety of biochemical reactions through which multiple species (e.g. metabolites, metabolite-enzyme complexes) are taken up from the environment

and/or are transformed into the building blocks of biological systems. Since the first stoichiometric models of *E. coli* (Varma & Palsson, 1994; Pramanik & Keasling, 1997), more refined metabolic models have been generated not only by increasing the gene and reaction coverage, but also by considering interactions of multiple cell types, tissues, organs in an organism (Martins Conde, et al., 2016) as well as interactions between organisms in communities (Fang, et al., 2020).

The success of the constraint-based modeling framework can in part be ascribed to: (i) invoking the simplifying assumption of steady state for the concentration of species and (ii) taking a flux-centered view that simplifies the metabolic constraints to a system of linear equations, in terms of the reaction rates (i.e. fluxes). The system of linear equation can be readily derived solely based on the stoichiometry of the analyzed network. Together with the capacity to include natural constraints on fluxes, these simplifications allow prediction of metabolic phenotypes with an assumed objective by employing classical convex optimization techniques (e.g. as done in flux balanced analysis (Orth, et al., 2010)).

The same quest for linearity is followed in chemical reaction network theory (CRNT) (Gunawardena, 2003). Like constraint-based modeling, CRNT also invokes the steady-state constraints; however, CRNT takes a concentration-centered view and often makes the simplifying assumption that the modeled reactions follow the classical mass action kinetics (Feinberg, 2019). As a result, one obtains a system of polynomial equations in terms of the species' concentrations which can be analyzed at different levels of abstraction. For instance, linearity in CRNT is achieved due to the consideration of complexes, corresponding to the substrate / product side of each modelled reaction. This allows the rewriting of the systems of polynomial equation into linear equations in terms of the monomials, arising from mass action kinetics, parameterized by the rate constants of the assumed mass action kinetics.

As a consequence, CRNT has relied on linear algebraic tools to address very general questions about whether or not a dynamical property arises given any or some choices for the value of parameters (Feinberg, 1987; 1988). For instance, one of the seminal results in CRNT deals with so-called complex balanced networks in which every complex is balanced, in the sense that the sums of reaction rates using the complex as a substrate or product are the same at every steady state supported by the network. Complex balanced networks have been shown not to exhibit exotic dynamic behavior, including capacity for multistationarity (Feinberg, 1987) or presence of species showing robust concentrations (Shinar & Feinberg, 2011). Determining if a network is complex balanced can be done by calculating the deficiency of the network, relying solely on the structure of a given network (Feinberg, 1987; Gunawardena, 2003).

However, balanced complexes are not only present in complex balanced networks. For instance, under the steady-state assumption, a complex including a species that does not occur in any other complex of a given network is balanced (Küken, et al., 2021; Feliu & Wiuf, 2013). Therefore, it is natural to ask the following three questions: Can balanced complexes be efficiently computed in networks of arbitrary kinetics? What are the effects that realistic flux constraints have on presence of balanced complexes? What are the network mechanisms that lead to the presence of balanced complexes with and without invoking flux constraints? Answers to these questions go in the direction of reconciling the flux- and concentration-

centered views, typical for the constraint-based and CRNT approaches. They also can help in shedding light on recent developments based on constraint-based linear programming approaches to identify balanced complexes in networks with arbitrary as well as mass action kinetics (Küken, et al., 2021). Since it has also been shown that balanced complexes can be used for effective reduction of large-scale metabolic models, answers to the abovementioned questions can help in addressing this problem that has gained some renewed interest (Singh & Lercher, 2020).

Here, we consider networks with arbitrary kinetics and provide a categorization of balanced complexes. The categorization delineates the contribution of network structure and flux constraints in categorizing a balanced complex into one of the four classes. We also formulate feasibility problems whose solutions can be used to categorize a balanced complex in a respective class. Our results also provide conditions that preclude the presence of particular class of balanced complexes. In addition, by analyzing high-quality metabolic networks across kingdoms of life we show that all classes of balanced complexes are prevalent and, therefore, relevant for properties of the supported steady states.

2 Material and methods

2.1 Preliminaries on Chemical Reaction Network Theory (CRNT)

In the classical CRNT literature (Aris & Mah, 1963; Horn & Jackson, 1972; Feinberg, 1979; 1980; 1987; 1988), a chemical reaction network (CRN) is usually defined by the 3-tuple $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$, where $\mathcal{S}$ is a set of m *species/metabolites* and $\mathcal{C}$ is a set of n *complexes*, whose elements $C \in \mathcal{C}$ can be regarded as multisets of species, $C \in \mathbb{N}^{\mathcal{S}}$. The set $\mathcal{R}$ is formally defined as a relation between network complexes, $\mathcal{R} \subset \mathcal{C} \times \mathcal{C}$; each element of $\mathcal{R}$ symbolizes the conversion of one complex to another and is referred to as a *reaction* of the network.

Throughout this manuscript, the standard basis for $\mathbb{R}^n$ will be denoted $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$, where

$$\mathbf{e}_j = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]^T,$$

is an n -vector with a unit value at the j^{th} entry and zero entries elsewhere. Assuming some arbitrary ordering on the sets of species $(S_1, \dots, S_m)$, complexes $(C_1, \dots, C_n)$, and reactions $(R_1, \dots, R_r)$, any complex $C_j \in \mathcal{C}$ can be alternatively represented by a vector $\mathbf{e}_j \in \mathbb{R}^n$ indexing its position in the ordered set, and at the same time associated with a unique vector $\mathbf{y}_j \in \mathbb{R}^m$, representing its species content. This defines the following mapping

$$\mathbf{e}_j \mapsto \mathbf{y}_j \in \mathbb{R}^m, \quad 1 \leq j \leq n. \quad (1)$$

As a result, any given CRN is associated with a matrix defined as a compilation of vectors $\mathbf{Y} = [\mathbf{y}_1 \ \dots \ \mathbf{y}_n]$, referred to as the *stoichiometric map*. In a similar fashion, one can associate any reaction $R_k : C_{k_1} \rightarrow C_{k_2} \in \mathcal{R}$ with vectors $\mathbf{a}_k = \mathbf{e}_{k_2} - \mathbf{e}_{k_1} \in \mathbb{R}^n$ and $\mathbf{n}_k = \mathbf{y}_{k_2} - \mathbf{y}_{k_1} \in \mathbb{R}^m$. Assembling these vectors in the same order gives the *incidence matrix* $\mathbf{A} = [\mathbf{a}_1 \ \dots \ \mathbf{a}_r]$ and the *stoichiometry matrix* $\mathbf{N} = [\mathbf{n}_1 \ \dots \ \mathbf{n}_r]$, respectively. It follows that $\mathbf{N} = \mathbf{Y} \mathbf{A}$.

The dynamics of the CRN is formulated by the following equation

$$\frac{d}{dt} \mathbf{x} = \mathbf{N} \mathbf{v} = \mathbf{Y} \mathbf{A} \mathbf{v}, \quad (2)$$

where $\mathbf{x} \in \mathbb{R}_+^m$ is the vector of species concentrations and $\mathbf{v} = \mathbf{v}(\mathbf{x}) \in \mathbb{R}^r$ is the vector of kinetic rates, referred to as *fluxes*, which is generally a nonlinear function of $\mathbf{x}$. The *flux distribution* $\mathbf{v}$ is often bounded by an upper bound $\mathbf{v}_u$ and a lower bound $\mathbf{v}_l$ as follows

$$\mathbf{v}_l \leq \mathbf{v} \leq \mathbf{v}_u. \quad (3)$$

By convention, $\mathbf{v}_u$ has strictly positive entries for all reactions. The set of *irreversible reactions* $\mathcal{R}^{\text{irr}} \subseteq \mathcal{R}$ can be determined by the nonnegative elements of $\mathbf{v}_l$, that is

$$\mathcal{R}^{\text{irr}} = \{R \in \mathcal{R} \mid v_{l,R} \geq 0\},$$

where $v_{l,R}$ is the designated lower bound on flux through reaction R .

We say the system is operating in a *canonical flux regime*, if $v_{l,R} = 0$, $\forall R \in \mathcal{R}^{\text{irr}}$. The set of *reversible reactions* is simply defined as the complement of $\mathcal{R}^{\text{irr}}$ in $\mathcal{R}$, i.e. $\mathcal{R}^{\text{rev}} = \mathcal{R} \setminus \mathcal{R}^{\text{irr}}$. The system is said to be operating in an unbounded flux regime, if $v_{u,R} = \infty$, $\forall R \in \mathcal{R}$; $v_{l,R} = -\infty$, $\forall R \in \mathcal{R}^{\text{rev}}$. One can always choose an ordering of the reactions in which elements of $\mathcal{R}^{\text{rev}}$ precede those of $\mathcal{R}^{\text{irr}}$. In this case, the incidence matrix $\mathbf{A}$ can be partitioned into matrix blocks $\mathbf{A} = [\mathbf{A}^{\text{rev}} \ \mathbf{A}^{\text{irr}}]$.

For the system to be at steady state, the flux vector $\mathbf{v}$ must lie in the nullspace of matrix $\mathbf{N}$, that is

$$\mathbf{N} \mathbf{v} = \mathbf{Y} \mathbf{A} \mathbf{v} = \mathbf{0}. \quad (4)$$

The intercept of the inequality constraints in Eq. (3) and equality constraint in Eq. (4) defines a convex set of feasible vectors $\mathbf{v}$, the *steady state flux set*, denoted $\mathcal{F}(G) = \{\mathbf{v} \in \mathbb{R}^r \mid \mathbf{Y} \mathbf{A} \mathbf{v} = \mathbf{0}, \mathbf{v}_l \leq \mathbf{v} \leq \mathbf{v}_u\}$. In a canonical and unbounded flux regime, the feasible set forms a polyhedral cone, referred to as the *steady state flux cone*.

For any $R \in \mathcal{R}$, we say R is a *blocked reaction* in G , if for all the feasible set distributions $\mathbf{v} \in \mathcal{F}(G)$, there is zero flux through R , that is, $v_R = 0$. This is a commonly occurring phenomenon in metabolic networks, especially in scenarios when flux bounds and/or optimization of particular objectives are imposed. Similarly, we say a reaction $R \in \mathcal{R}$ is *fixated* at some flux value f , if for all the flux distributions $\mathbf{v} \in \mathcal{F}(G)$, the flux through R is unchanged, namely, $v_R = f$, $\forall \mathbf{v} \in \mathcal{F}(G)$. Clearly, any blocked reaction is one fixated at zero.

2.2 Balanced complexes

An important class of steady state fluxes are the so-called *complex-balanced steady states*, which were introduced by Horn and Jackson (1972) (Horn, 1972) as a generalization of detailed balanced steady states. At a complex-balanced steady state, the flux distribution satisfies

$$\mathbf{A} \mathbf{v} = \mathbf{0}. \quad (5)$$

Eq. (5) constrains $\mathbf{v}$ to lie in $\ker \mathbf{A}$, which is generally a subset of $\ker \mathbf{N}$, and, hence, is a stronger condition compared to Eq. (4). From a conceptual viewpoint, complex-balanced steady states are those in which the algebraic sum of the total flux entering and leaving each complex equals zero. In more technical terms, the flux vector is composed of a number of cyclic generators

(Johnston, 2014), i.e. nonstoichiometric elementary flux modes (Conradi, et al., 2007). Of particular interest are systems for which all steady states are complex-balanced, that is, Eq. (4) can be seamlessly replaced by Eq. (5). This class of systems -referred to as *complex-balanced systems*- are well-studied in the classical body of work on CRNT, and very strong results have been derived for them, most notably the well-known Deficiency Zero Theorem (Feinberg, 1987).

Unfortunately, in most cases those strong results cannot be applied to the majority of real-world metabolic networks: Due to the more complex nature of such systems, often not all but only some complexes are balanced at steady state. However, recent studies have shown that, even in absence of full complex-balancing, existence of balanced complexes can be exploited to obtain significant network reductions and throw some light on the steady states (Feliu & Wiuf, 2013; van der Schaft, et al., 2015; Küken, et al., 2021). This gives rise to the question regarding the underlying mechanisms that lead to the formation of complex balancing in metabolic networks.

Let us next give a formal definition of a balanced complex. For a given matrix $\mathbf{X}$, let $\mathbf{X}^{:i}$ and $\mathbf{X}^{j:}$ denote the i^{th} column and j^{th} row of $\mathbf{X}$, respectively. A complex $C_j \in \mathcal{C}$ is defined to be a *balanced complex* (BC) for network G , if $\mathbf{A}^{j:} \mathbf{v} \equiv 0$ for all $\mathbf{v} \in \mathcal{F}(G)$. Given how incidence matrix $\mathbf{A}$ is constructed, this definition complies with the notion that the algebraic sum of fluxes entering and leaving C_j must be always zero at any steady state.

2.3 Primal-dual formulations

To investigate whether a complex $C_j \in \mathcal{C}$ is a BC, one may form and solve the following two linear programs (LP)

$$\begin{aligned} & \underset{\mathbf{v} \in \mathbb{R}^r}{\text{minimize}} && \mathbf{e}_j^T \mathbf{A} \mathbf{v} \\ & \text{subject to} && \mathbf{Y} \mathbf{A} \mathbf{v} = \mathbf{0} \quad , \\ & && \mathbf{v}_l \leq \mathbf{v} \leq \mathbf{v}_u \end{aligned} \quad (6)$$

$$\begin{aligned} & \underset{\mathbf{v} \in \mathbb{R}^r}{\text{maximize}} && \mathbf{e}_j^T \mathbf{A} \mathbf{v} \\ & \text{subject to} && \mathbf{Y} \mathbf{A} \mathbf{v} = \mathbf{0} \quad . \\ & && \mathbf{v}_l \leq \mathbf{v} \leq \mathbf{v}_u \end{aligned} \quad (7)$$

The definition implies that $C_j \in \mathcal{C}$ is a BC if and only if both above LPs have zero optimal values. The optimization problems (6) and (7) have a Lagrange dual of the following form [derivation in Section S1.2 of the Supplementary Information]

$$\begin{aligned} & \underset{\lambda_u, \lambda_l \in \mathbb{R}^r, \zeta \in \mathbb{R}^m}{\text{maximize}} && \lambda_l^T \mathbf{v}_l - \lambda_u^T \mathbf{v}_u \\ & \text{subject to} && \mathbf{A}^T \mathbf{e}_j - \mathbf{A}^T \mathbf{Y}^T \zeta + \lambda_u - \lambda_l = \mathbf{0} \quad . \\ & && \lambda_u, \lambda_l \geq \mathbf{0} \end{aligned} \quad (8)$$

$$\begin{aligned} & \underset{\lambda_u, \lambda_l \in \mathbb{R}^r, \zeta \in \mathbb{R}^m}{\text{maximize}} && \lambda_l^T \mathbf{v}_l - \lambda_u^T \mathbf{v}_u \\ & \text{subject to} && -\mathbf{A}^T \mathbf{e}_j + \mathbf{A}^T \mathbf{Y}^T \zeta + \lambda_u - \lambda_l = \mathbf{0} \quad . \\ & && \lambda_u, \lambda_l \geq \mathbf{0} \end{aligned} \quad (9)$$

Clearly, given the linear constraints, strong duality holds and the duality gap is zero (Boyd & Vandenberghe, 2004); therefore, a complex $C_j \in \mathcal{C}$ is a BC if and only if both Problems (8) and (9) have optimal values equal to zero.

3 Results and discussion

3.1 Generalization of trivial BCs

Taking into consideration the stoichiometry of a chemical reaction network, it is straightforward to see how some balanced complexes may emerge. A complex $C_j \in \mathcal{C}$ is *trivially-balanced* (a *trivial BC*), if there exist some species $S_i \in \mathcal{S}$ that appears only in C_j and nowhere else in the network (Küken, et al., 2021). Given the structure of the stoichiometric map $\mathbf{Y}$, it follows that $\mathbf{Y}^i = \mathbf{e}_j^T$. Hence, the steady state equation $\mathbf{Y} \mathbf{A} \mathbf{v} = \mathbf{0}$ immediately yields $\mathbf{e}_j^T \mathbf{A} \mathbf{v} \equiv 0$, $\forall \mathbf{v} \in \mathcal{F}(G)$.

One can readily generalize the idea behind the formation of a trivial BC to potentially arrive at a larger class of balanced complexes in the network. Intuitively, the vector $\mathbf{e}_j$ does not have to appear explicitly as a row in $\mathbf{Y}$; all it takes is for $\mathbf{e}_j$ to be a linear combination of rows in $\mathbf{Y}$, that is

$$\exists \boldsymbol{\zeta} \in \mathbb{R}^m : \mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta} ; \quad (10)$$

it then follows that $\mathbf{e}_j^T \mathbf{A} \mathbf{v} = \boldsymbol{\zeta}^T \mathbf{Y} \mathbf{A} \mathbf{v} = 0$, $\forall \mathbf{v} \in \mathcal{F}(G)$. We remark that Eq. (10) can potentially explain a higher number of balanced complexes in the network, and it fully contains all the trivial BCs.

One may take another step and further generalize this notion to potentially identify an even larger class of BCs. Note that all BCs predicted by Eq. (10) have purely stoichiometric grounds, derived from stoichiometric map $\mathbf{Y}$. However, the topological structure of a network, pinpointed by matrix $\mathbf{A}$, can also play a role in the formation of BCs. Let us next consider a simple example, in which some species $S_i \in \mathcal{S}$ makes a unimolecular appearance in complexes C_{j_1} and C_{j_2} , but is absent elsewhere in the network. Assume additionally that C_{j_1}, C_{j_2} form a linkage class with another complex, namely C_{j_3} . The situation is shown in Fig. 1.

Fig. 1. The topology contributing to the formation of a BC.

A small linkage class of a network comprising three complexes is shown. The species S_i is assumed to be present in complexes C_{j_1} and C_{j_2} , but nowhere else in the network.

On the one hand, $\mathbf{Y}^{i:} = (\mathbf{e}_{j_1} + \mathbf{e}_{j_2})^T$ and the steady state condition for species S_i imply that $(\mathbf{A}^{j_1:} + \mathbf{A}^{j_2:}) \mathbf{v} = 0$, $\forall \mathbf{v} \in \mathcal{F}(G)$; on the other hand, the closedness of the linkage class requires that $\mathbf{A}^{j_1:} + \mathbf{A}^{j_2:} + \mathbf{A}^{j_3:} = \mathbf{0}^T$. Thus, it follows that $\mathbf{A}^{j_3:} \mathbf{v} = 0$, $\forall \mathbf{v} \in \mathcal{F}(G)$, that is, C_{j_3} is a BC. Note that the balancing of C_{j_3} has nothing to do with the species in it, but with its connection to other complexes.

Formalizing this generalization relies on integrating the linkage structure of the network, which is closely interlinked with the left nullspace of the incidence matrix $\mathbf{A}$, into Eq. (10). Let G be a closed network composed of ℓ linkage classes $L_1, \dots, L_\ell$. The left nullspace of $\mathbf{A}$, i.e. $\ker \mathbf{A}^T$ has a basis of the form $\{\mathbf{u}_1, \dots, \mathbf{u}_\ell\}$, where $\mathbf{u}_l = \sum_{i: C_i \in L_l} \mathbf{e}_i$, $1 \leq l \leq \ell$. Next, define the matrix $\mathbf{U} \in \mathbb{R}^{n \times \ell}$ as $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_\ell]$. Now, let a complex $\mathbf{e}_j$ have the following form

$$\exists \boldsymbol{\zeta} \in \mathbb{R}^m, \boldsymbol{\xi} \in \mathbb{R}^\ell : \mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta} + \mathbf{U} \boldsymbol{\xi} ; \quad (11)$$

it follows that $\mathbf{e}_j^T \mathbf{A} \mathbf{v} = \boldsymbol{\zeta}^T \mathbf{Y} \mathbf{A} \mathbf{v} + \boldsymbol{\xi}^T \mathbf{U} \mathbf{A} \mathbf{v} = 0$, $\forall \mathbf{v} \in \mathcal{F}(G)$. Eq. (11) shall be viewed as a generalization of Eq. (10), which can potentially account for a higher number of BCs in the network.

3.2 Factorizations of balanced complexes

Eq. (11), like Eq. (10), provides a sufficient condition: If there exist vectors $\boldsymbol{\zeta} \in \mathbb{R}^m, \boldsymbol{\xi} \in \mathbb{R}^\ell$ such that Eq. (11) holds for some $\mathbf{e}_j$, then complex $C_j \in \mathcal{C}$ is a BC. This poses the following question: Does Eq. (11) account for all BCs in the network? We seek to determine whether all BCs are created as a result of an interplay of stoichiometry and the linkage structure or there exist other factors contributing to the formation of BCs. We are also interested in establishing the conditions under which Eq. (11) would become a necessary and sufficient condition for BCs.

The dual formulation in Eqs. (8) and (9) helps address these questions.

Theorem 1. A complex $C_j \in \mathcal{C}$ is a BC, if and only if there exist variables $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in \mathbb{R}^m$, $\lambda_{u1}, \lambda_{l1}, \lambda_{u2}, \lambda_{l2} \in \mathbb{R}^r$, such that

$$\begin{cases} \mathbf{A}^T \mathbf{e}_j = \mathbf{A}^T \mathbf{Y}^T \boldsymbol{\zeta}_1 + \lambda_{l1} - \lambda_{u1} \\ \mathbf{v}_l^T \lambda_{l1} - \mathbf{v}_u^T \lambda_{u1} = 0 \\ \mathbf{A}^T \mathbf{e}_j = \mathbf{A}^T \mathbf{Y}^T \boldsymbol{\zeta}_2 + \lambda_{u2} - \lambda_{l2} \\ \mathbf{v}_l^T \lambda_{l2} - \mathbf{v}_u^T \lambda_{u2} = 0 \\ \lambda_{u1}, \lambda_{l1}, \lambda_{u2}, \lambda_{l2} \geq 0 \end{cases} \quad (12)$$

Eq. (12) implicitly relates the balanced complex not only to the stoichiometry but also to the upper- and lower bounds on reaction fluxes. It can actually be viewed as a generalization of Eq. (11), which brings into play the flux bound constraints as a third factor in the formation of BCs. The following result presents an intermediary formulation that elucidates the connection between Eqs. (11) and (12):

Theorem 2. Suppose the network is operating under a canonical flux regime. There exist variables $\zeta_1, \zeta_2 \in \mathbb{R}^m$, $\xi_1, \xi_2 \in \mathbb{R}^\ell$, $\theta_1, \theta_2 \in \mathbb{R}^n$, such that

$$\begin{cases} \mathbf{e}_j = \mathbf{Y}^T \zeta_1 + \mathbf{U} \xi_1 + \theta_1 \\ \mathbf{e}_j = \mathbf{Y}^T \zeta_2 + \mathbf{U} \xi_2 - \theta_2 \\ \mathbf{A}^{\text{rev}T} \theta_1 = \mathbf{0} \\ \mathbf{A}^{\text{irr}T} \theta_1 \geq \mathbf{0} \\ \mathbf{A}^{\text{rev}T} \theta_2 = \mathbf{0} \\ \mathbf{A}^{\text{irr}T} \theta_2 \geq \mathbf{0} \end{cases}, \quad (13)$$

if and only if the complex $C_j \in \mathcal{C}$ is a BC. By convention, $\mathbf{A}^{\text{irr}}$ denotes the columns of $\mathbf{A}$ corresponding to zero lower bounds in $\mathbf{v}_l$, and $\mathbf{A}^{\text{rev}}$ is the complementary block in $\mathbf{A}$.

It is straightforward to see how Eq. (13) is a generalization of Eq. (11). In particular, any feasible solution of Eq. (11) for parameter values ζ, ξ corresponds to a solution of Eq. (13) with feasible parameter values $\zeta_1 = \zeta_2 = \zeta$, $\xi_1 = \xi_2 = \xi$, $\theta_1 = \theta_2 = \mathbf{0}$. Note that irreversibility patterns play a key role in distinguishing the solution sets of Eqs. (13) and (11). In addition, it is not difficult to show how Eq. (12) can be viewed as a generalization of Eq. (13). One only needs to define $\lambda_{lt} = \mathbf{A}^T \theta_t$, $\lambda_{ut} = \mathbf{0}$; $t = 1, 2$ [refer to Section S1.4 of the Supplementary Information for mathematical details].

3.3 Categorization of BCs

In Eqs. (10), (11), (13) and (12), we presented four different formulations that can explain the emergence of a BC potentially as a combined effect of distinct factors, namely stoichiometry, linkage structure, irreversibility patterns and flux bounds, respectively. Each “factorization” can be viewed as a feasibility problem; let us next define the *solution set* of each factorization as the set of all complexes $C_j \in \mathcal{C}$, where vector $\mathbf{e}_j$ satisfies the feasibility problem for some parameter values. It is worth noting that the solution set of each factorization is a subset of the next one, in the abovementioned order of factorizations. Moreover, the solution set of Eq. (12) includes every existing BC in the network. This property paves the way to classify each BC of the network into one of the following four categories:

- (I) A complex $C_j \in \mathcal{C}$ is called a *strictly stoichiometric* BC, if it has a factorization of the form Eq. (10), i.e. $\exists \zeta \in \mathbb{R}^m : \mathbf{e}_j = \mathbf{Y}^T \zeta$. This relation is referred to as a *strictly stoichiometric factorization* for the BC represented by vector $\mathbf{e}_j$.
- (II) A complex $C_j \in \mathcal{C}$ is called a *stoichiometric* BC, if it has a factorization of the form Eq. (11), i.e. $\exists \zeta \in \mathbb{R}^m, \xi \in \mathbb{R}^\ell : \mathbf{e}_j = \mathbf{Y}^T \zeta + \mathbf{U} \xi$. This relation is referred to as a *stoichiometric factorization* for the BC represented by vector $\mathbf{e}_j$.
- (III) A complex $C_j \in \mathcal{C}$ is called a *type-I nonstoichiometric* BC, if it has a factorization of the form Eq. (13), but no stoichiometric factorization. Eq. (13) is called a *type-I nonstoichiometric factorization* for the BC represented by vector $\mathbf{e}_j$.

- (IV) A complex $C \in \mathcal{C}$ is called a *type-II nonstoichiometric BC*, if it has a factorization of the form Eq. (12), but no factorization of the form Eq. (13). Eq. (12) is called a *type-II nonstoichiometric factorization* for the BC represented by vector $\mathbf{e}_j$.

It follows directly from these definitions that any strictly stoichiometric BC/factorization is also a stoichiometric BC/factorization. Moreover, a balanced complex is labeled with the general term *nonstoichiometric BC*, if it has no stoichiometric factorization, regardless of whether it has an explicit factorization of the form Eq. (13) or an implicit factorization of the form Eq. (12).

3.4 Characteristics of distinct BC categories

Recent studies show that balanced complexes can play an important role in the reduction of metabolic networks and simplification of steady state analysis (Feliu & Wiuf, 2013; Küken, et al., 2021). The categorization of BCs into distinct classes not only helps identify the origin of the balancing in each case, but also enables one to study their properties. On the one hand, this equips us with an insight into how certain modification of a network may affect existing BCs. On the other hand, it may demonstrate if and how modeling inaccuracies may lead to a detection failure for certain existing BCs.

The strictly stoichiometric BCs constitute the most trivial class, whose formation relies merely on the network stoichiometry. This simplicity gives rise to an intrinsic form of robustness: As long as network complexes are unaltered, the balancing property of strictly stoichiometric BCs remains preserved, regardless of any potential changes in the network, e.g. even if the reactions in these networks are altered (referred to as different enzymatic regimes). In this case, the formation of strictly stoichiometric BCs roots solely in the steady state property for the species.

For the more general class of stoichiometric BCs, their formation not only relies on the network stoichiometry, but also partly on a topological feature, namely the linkage structure. They are affected neither by the irreversibility patterns in the network nor by the constraints on reaction rates. However, they may be altered under different enzymatic/catalytic regimes. Their balancing property is rooted not only in the steady state property but also in the conversion patterns (i.e. reaction structure) of the network.

The formation of nonstoichiometric BCs is a more sophisticated and at the same time more interesting phenomenon. They often emerge as a combined effect of several network features, including the stoichiometry and the linkage structure, but also irreversibility patterns as well as upper- and lower bounds on fluxes. In other words, these BCs carry information in the form of an equality constraint on reaction fluxes, which is not encoded in the steady state equation $\mathbf{Y}\mathbf{A}\mathbf{v} = \mathbf{0}$.

In a canonical flux regime, i.e. when irreversible reactions have a zero lower bound on reaction rates, all nonstoichiometric BCs are in the type-I category. As their factorization suggests, their emergence relies specifically on the irreversibility patterns and not on the exact values of the flux bounds. This makes the detection of these BCs insensitive to potential modeling errors in the form of inaccuracies in flux bounds, even when they appear in a non-canonical flux regime. However, BCs in this category have the following undesirable property:

Proposition 3. Let a network G contain a type-I nonstoichiometric BC. There exist at least two blocked irreversible reactions in the network.

It follows that once all blocked reactions are removed, no type-I nonstoichiometric BCs may exist in the network. In fact, the removal of such reactions changes the linkage structure of the network such that all type-I nonstoichiometric BCs in the original network appear as stoichiometric BCs in the reduced network.

Type-II nonstoichiometric BCs do not automatically require a number of network reactions to be blocked in the steady state flux set. By contrast, they require a number of reactions to be fixated at a nonzero upper- or lower bound, as the following statement demonstrates.

Proposition 4. Let G be a network all blocked reactions of which have been removed. Suppose G contains a nonstoichiometric BC. There exist at least three reactions in G , which are fixated at a corresponding nonzero lower- or upper bound, for all $\mathbf{v} \in \mathcal{F}(G)$.

More specifically, it can be shown that the existence of a type-II nonstoichiometric BC implies at least one irreversible reaction is fixated at a positive lower bound. It is not surprising that such BCs only emerge in non-canonical flux regimes. However, an undesirable feature with type-II nonstoichiometric BCs is that their factorization relies explicitly on the values of upper- and lower bounds, which may make their detection sensitive to modeling assumptions.

3.5 Toy examples

Distinct classes of balanced complexes identified in this study are practically observed in all sorts of small and large networks. However, here we present small networks specifically designed for illustrative purposes. The fact that the chemistry in these toy networks may seem unrealistic or may not reflect real-world mechanisms should not be viewed critically; they were constructed with the idea of providing simple illustrations involving very few species.

With that in mind, let us next consider the toy network on Fig. 2.

Fig. 2. Stoichiometric BCs in a toy network.

The conversion diagram depicts a base network with $m = 5$ species, $n = 7$ complexes, and $r = 6$ base reactions, two of which are reversible. The network has four BCs, which will remain balanced even if one adds the dashed reaction (highlighted in green). However, once one adds the dotted reaction (highlighted in red), only two of these complexes remain balanced.

Let us first only consider the base network in Fig. 2, which excludes the dotted and dashed reactions. It is trivial to show that $(A + B)$ and $(2B)$ have strictly stoichiometric factorizations. The two complexes $(2C)$ and $(C + D)$ are stoichiometric BCs but not strictly stoichiometric ones. All four complexes will remain balanced, if one changes the irreversibility of reactions arbitrarily or if the dashed reaction is added to the network. However, as soon as one adds the dotted reaction to the network, the linkage structure will be altered. As a result, the strictly

stoichiometric BCs ($A + B$) and ($2B$) will still be balanced, but the stoichiometric BCs ($2C$) and ($C + D$) will not remain balanced.

Nonstoichiometric BCs are more complex phenomena that are expected to emerge more frequently in larger networks, and, thus, are more challenging to illustrate in toy networks. To demonstrate an example, let us consider the conversion diagram in Fig. 3.

Fig. 3. A nonstoichiometric BC in a toy network.

The conversion diagram depicts a network with $m = 6$ species, $n = 9$ complexes, and $r = 8$ reactions. The network has one balanced complex ($2A$) which is a type-I nonstoichiometric BC. As predicted by Proposition 3, the two irreversible reactions in this network (highlighted in brown) are both blocked at steady state.

The network in Fig. 3 contains no stoichiometric BCs. Let us sort the species by alphabetical order and sort the complexes as follows: $(C_1, \dots, C_9) = (A + B, A + C, C + D, C + E, C + F, B + D, 2E, 2F, 2A)$. Given the linkage structure of this network, let us define the basis matrix $\mathbf{U}$ as follows

$$\mathbf{U} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}^T.$$

It can be shown that the complex indexed by vector $\mathbf{e}_9$ has a type-I nonstoichiometric factorization with the following parameters

$$\begin{aligned} \zeta_1 &= [1 \ 0 \ 0 \ 0 \ 0 \ 0]^T, & \zeta_2 &= [1 \ 0 \ 0 \ 1 \ 1 \ 1]^T; \\ \xi_1 &= [-1 \ 0]^T, & \xi_2 &= [-1 \ -1]^T; \\ \theta_1 &= [0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0]^T, & \theta_2 &= [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0]^T; \end{aligned}$$

It follows that c_9 is a type-I nonstoichiometric BC for the network. It can also be shown that the two irreversible reactions (shown in brown) are blocked at steady state, which is consistent with the prediction of Proposition 3. Note that removing the two blocked reactions would alter the linkage structure of the network, as a result of which the complex $C_9 = 2A$ would become a stoichiometric BC for the reduced network.

It is worth noting that changing the irreversibility patterns elsewhere in the network does not affect the balancing property; however, C_9 would not be balanced anymore if one modifies either of the two reactions in brown. It is precisely the irreversibility of these two reactions in this particular constellation that makes C_9 a balanced complex.

3.6 Presence of the different categories of BCs in metabolic models

To show the prevalence of the proposed categorization of BCs in real-world networks, we categorize balanced complexes identified in twelve genome-scale metabolic networks of organisms from all kingdoms of life (Küken, et al., 2021). Küken et al. removed blocked

reactions from the networks prior to the identification of BCs, which results in the absence of type-I nonstoichiometric BCs. Our results indicate that all remaining categories of BCs, namely: strictly stoichiometric, stoichiometric, and type-II nonstoichiometric, are found across all analyzed genome-scale networks. We find that stoichiometric BCs form the largest group of BCs with 59% to 94% of all BCs identified, except for models of *A. thaliana*, *M. musculus*, *N. pharaonis* and *P. putida* where stoichiometric BCs could not be found or comprise below 1.5% of all BCs (Fig. 4). In contrast, the class of strictly stoichiometric BCs is present in all twelve analyzed models, with 6% to 94% (on average 26%) of BCs falling into this category. Type-II nonstoichiometric BCs could be identified in the metabolic models of *A. thaliana* (6%), *M. musculus* (66%), *N. pharaonis* (93%) and *P. putida* (90%). Interestingly, models including type-II nonstoichiometric BCs, were found to include less than 1.5% stoichiometric BCs. Our analysis of metabolic networks from organisms of kingdoms across life shows the prevalence of all proposed factors (stoichiometry, linkage structure and flux bounds) for the formation of BCs in biological networks.

Fig. 4. Classification of balanced complexes into strictly stoichiometric, stoichiometric and type-II nonstoichiometric.

Balanced complexes from twelve genome-scale models of organisms across kingdoms of life are classified and the percentage of model BCs that fall into the respective class is presented. Type-I nonstoichiometric BCs are not present in the analyzed models since blocked reactions were removed from the networks prior to analyses. For model references, see Supplementary Information.

4 Concluding remarks

The nonlinearity and dimensionality of steady state equations makes it extremely difficult to study the properties of metabolic networks. While closed-form solutions seem evasive in most cases, recent approaches exploit balanced complexes as a means to reduce metabolic networks and facilitate steady state analysis. However, the underlying mechanisms leading to the formation of balanced complexes have not been studied so far. This work investigates this question and identifies stoichiometry, linkage structure, irreversibility patterns and flux bounds as the main driving factors behind this phenomenon.

The analysis enables us to classify balanced complexes into distinct stoichiometric and nonstoichiometric categories based on their origins and to study their properties. Formulation of the problem as a linear feasibility program provides a computationally efficient tool that is applicable to even large-scale metabolic networks. It is a fascinating aspect of these results that they are not obtained under any specific kinetic assumptions, which suggest they enjoy a ubiquity and are unperturbed by kinetic approximations. Indeed, examination of several well-established metabolic models demonstrates that all these categories are present in distinct metabolic networks across diverse kingdoms of life.

Supplementary Information

The full derivation of the dual problems and all mathematical formulations, as well as detailed proofs for the theorems and propositions in the text are provided in [Supplementary Information \[link\]](#). In addition, we provide information on the analyzed genome-scale metabolic networks and computational procedure of BC classification.

References

- Aris, R. & Mah, R. H. S., 1963. Independence of Chemical Reactions. *Industrial & Engineering Chemistry Fundamentals*, May, 2(2), pp. 90-94.
- Bordbar, A., Monk, J. M., King, Z. A. & Palsson, B. O., 2014. Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics*, 1, Volume 15, p. 107–120.
- Boyd, S. & Vandenberghe, L., 2004. Duality. In: *Convex Optimization*. s.l.:Cambridge University Press, p. 215–288.
- Conradi, C., Flockerzi, D., Raisch, J. & Stelling, J., 2007. Subnetwork analysis reveals dynamic features of complex (bio)chemical networks. *Proc Natl Acad Sci USA*, 12, Volume 104, p. 19175.
- Fang, X., Lloyd, C. J. & Palsson, B. O., 2020. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology*, 9, Volume 18, p. 731–743.
- Feinberg, M., 1979. *Lectures On Chemical Reaction Networks*. [Online] Available at: <https://crnt.osu.edu/LecturesOnReactionNetworks> [Accessed 21 March 2021].
- Feinberg, M., 1980. Chemical Oscillations, Multiple Equilibria, and Reaction Network Structure. In: W. E. STEWART, R. A. Y. W. HARMON & C. C. CONLEY, eds. *Dynamics and Modelling of Reactive Systems*. s.l.:Academic Press, pp. 59-130.

- Feinberg, M., 1987. Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency zero and deficiency one theorems. *Chemical Engineering Science*, Volume 42, pp. 2229-2268.
- Feinberg, M., 1988. Chemical reaction network structure and the stability of complex isothermal reactors—II. Multiple steady states for networks of deficiency one. *Chemical Engineering Science*, Volume 43, pp. 1-25.
- Feinberg, M., 2019. Foundations of Chemical Reaction Network Theory.
- Feliu, E. & Wiuf, C., 2013. Simplifying biochemical models with intermediate species. *Journal of The Royal Society Interface*, 10, Volume 10, p. 20130484.
- Gunawardena, J., 2003. Chemical reaction network theory for in-silico biologists. 6.
- Horn, F., 1972. Necessary and sufficient conditions for complex balancing in chemical kinetics. *Archive for Rational Mechanics and Analysis*, Volume 49, p. 172–186.
- Horn, F. & Jackson, R., 1972. General mass action kinetics. *Archive for Rational Mechanics and Analysis*, Volume 47, p. 81–116.
- Johnston, M. D., 2014. Translated chemical reaction networks.. *Bulletin of mathematical biology*, 5, 76(5), pp. 1081-116.
- Küken, A., Wendering, P., Langary, D. & Nikoloski, Z., 2021. A structural property for reduction of biochemical networks. *bioRxiv*, 1.p. 2021.03.17.435785.
- Martins Conde, P. d. R., Sauter, T. & Pfau, T., 2016. Constraint Based Modeling Going Multicellular. *Frontiers in Molecular Biosciences*, 2, Volume 3, p. 3.
- Orth, J. D., Thiele, I. & Palsson, B. Ø., 2010. What is flux balance analysis?. *Nature biotechnology*, 3, Volume 28, p. 245–248.
- Pramanik, J. & Keasling, J. D., 1997. Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements.. *Biotechnology and bioengineering*, 11, 56(4), pp. 398-421.
- Shinar, G. & Feinberg, M., 2011. Design principles for robust biochemical reaction networks: what works, what cannot work, and what might almost work. *Mathematical biosciences*, 5, Volume 231, p. 39–48.
- Singh, D. & Lercher, M. J., 2020. Network reduction methods for genome-scale metabolic models. *Cellular and Molecular Life Sciences*, 11, Volume 77, p. 481–488.
- van der Schaft, A., Rao, S. & Jayawardhana, B., 2015. Complex and detailed balancing of chemical reaction networks revisited. *Journal of Mathematical Chemistry*, 4, Volume 53, p. 1445–1458.
- Varma, A. & Palsson, B. O., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110.. *Applied and environmental microbiology*, 10, 60(10), pp. 3724-31.

Supplementary Information to

The unraveling of balanced complexes in metabolic networks

Damoun Langary^{1,2}, Anika Küken², Zoran Nikoloski^{1,2}

¹Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany

²Bioinformatics, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

5 S1 Mathematical Analysis of Balanced Complexes

5.1 S1.1 Definitions and motivation

Let $G = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a chemical reaction network with stoichiometric map $\mathbf{Y}$ and complex-to-reaction incidence matrix $\mathbf{A}$. The steady state flux set for G is defined as follows

$$\mathcal{F}(G) = \{\mathbf{v} \mid \mathbf{Y}\mathbf{A}\mathbf{v} = \mathbf{0}, \mathbf{v}_l \leq \mathbf{v} \leq \mathbf{v}_u\}$$

where $\mathbf{v}_l$ and $\mathbf{v}_u$ denote lower- and upper bounds for flux through reactions, respectively.

A complex $C_j \in \mathcal{C}$ is referred to as a balanced complex (BC), if for all flux vectors in $\mathcal{F}(G)$ the total flux entering this complex is equal to the total flux leaving this complex. This is equivalent to

$$\mathbf{e}_j^T \mathbf{A} \mathbf{v} = 0, \quad \forall \mathbf{v} \in \mathcal{F}(G),$$

where $\mathbf{e}_j$ is the vector with a unit value for the j^{th} entry and zero values elsewhere.

For any $R \in \mathcal{R}$, we say R is a blocked reaction in G , if for all the feasible set distributions $\mathbf{v} \in \mathcal{F}(G)$, there is zero flux through R , that is, $v_R = 0$. This is a commonly occurring phenomenon in metabolic networks, especially in scenarios when flux bounds and/or optimization of particular objectives are imposed. Similarly, we say a reaction $R \in \mathcal{R}$ is fixated at some flux value f , if for all flux distributions $\mathbf{v} \in \mathcal{F}(G)$, the flux through R is unchanged, namely, $v_R = f, \forall \mathbf{v} \in \mathcal{F}(G)$. Clearly, any blocked reaction is one fixated at zero.

Looking at the stoichiometry of a chemical reaction network, it is easy to see how balanced complexes may emerge. In the most trivial case, any complex comprising a species that appears nowhere else in the network must be balanced at any steady state, thus a BC. Note that this case amounts to finding a single row of the stoichiometric map $\mathbf{Y}$ —as a binary indexing vector for a complex— that defines a balanced complex.

One can readily generalize this idea to come up with more balanced complexes for the network. Suppose a vector $\mathbf{e}_j$ can be expressed as a linear combination of rows of the stoichiometric map $\mathbf{Y}$, that is

$$\exists \boldsymbol{\zeta} : \mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta} ; \quad (1)$$

it immediately follows that

$$\mathbf{e}_j^T \mathbf{A} \mathbf{v} = \boldsymbol{\zeta}^T \mathbf{Y} \mathbf{A} \mathbf{v} = \mathbf{0} , \quad \forall \mathbf{v} \in \mathcal{F}(G) .$$

One can further expand the idea by taking into account the left nullspace of the incidence matrix $\mathbf{A}$. Let G be a closed system; the left nullspace of $\mathbf{A}$ has a simplified structure: If the CRN has a connected graph, then the nullspace is the one-dimensional subspace defined as span of basis vector, $\langle \mathbf{1} \rangle$. This yields

$$\exists \boldsymbol{\zeta}, \xi : \quad \mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta} + \xi \mathbf{1};$$

If the CRN is not connected, then the nullspace has basis vectors of the same structure, but for separate linkage classes. Let us assume G consists of ℓ linkage classes, denoted by $L_1, L_2, \dots, L_\ell$. With each linkage class l , for $l = 1, \dots, \ell$, one can associate a vector $\mathbf{u}^{(l)} \in \mathbb{R}^c$ defined as follows

$$u_j^{(l)} = \begin{cases} 0 & c_j \notin L_l \\ 1 & c_j \in L_l \end{cases} .$$

The columns of the matrix $\mathbf{U} = [\mathbf{u}^{(1)} \ \mathbf{u}^{(2)} \ \dots \ \mathbf{u}^{(\ell)}] \in \mathbb{R}^{|\mathcal{C}| \times \ell}$ form a basis for the left nullspace of $\mathbf{A}$. Now suppose a vector $\mathbf{e}_j$ can be expressed as follows

$$\exists \boldsymbol{\zeta}, \boldsymbol{\xi} : \quad \mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta} + \mathbf{U} \boldsymbol{\xi} ; \quad (2)$$

it is easy to see that

$$\mathbf{e}_j^T \mathbf{A} \mathbf{v} = \mathbf{0} , \quad \forall \mathbf{v} \in \mathcal{F}(G) .$$

Therefore, $\mathbf{e}_j$ represents a balancing relation for G .

We refer to Eq. (2) as a stoichiometric factorization for the indexing vector $\mathbf{e}_j$. Any BC that can be parameterized in the form of Eq. (2) is referred to as a stoichiometric BC.

Here come the key questions we want to address next: Are these the only forms of balanced complexes that may emerge in a network? In other words, are all balanced complexes in a network simply a result of the stoichiometric- and linkage structure? Could other constraints on a network, such as lower and upper bounds on flux levels, also play a role in creating some other balanced complexes?

5.2 S1.2 The primal-dual formulation

To answer these questions, we first formulate the balancing problem as a pair of optimization problems as follows

$$\begin{aligned} & \text{minimize} && \mathbf{e}_j^T \mathbf{A} \mathbf{v} \\ & \text{subject to} && \mathbf{Y} \mathbf{A} \mathbf{v} = \mathbf{0} \\ & && \mathbf{v}_l \leq \mathbf{v} \leq \mathbf{v}_u \end{aligned} \quad (P1) ,$$

$$\begin{aligned} & \text{maximize} && \mathbf{e}_j^T \mathbf{A} \mathbf{v} \\ & \text{subject to} && \mathbf{Y} \mathbf{A} \mathbf{v} = \mathbf{0} \\ & && \mathbf{v}_l \leq \mathbf{v} \leq \mathbf{v}_u \end{aligned} .$$

For the problem formulation to conform to standard conventions in convex optimization, we simply replace the 2nd optimization problem by an equivalent minimization problem as follows.

$$\begin{aligned} & \text{minimize} && -\mathbf{e}_j^T \mathbf{A} \mathbf{v} \\ & \text{subject to} && \mathbf{Y} \mathbf{A} \mathbf{v} = \mathbf{0} \\ & && \mathbf{v}_l \leq \mathbf{v} \leq \mathbf{v}_u \end{aligned} \quad (P2) .$$

A vector $\mathbf{e}_j$ represents a balanced complex, if and only if both optimization problems $P1$ & $P2$ have an optimal value of zero.

$$p_1^* = p_2^* = 0 .$$

Any BC has to satisfy this criterion.

To further analyze which vectors $\mathbf{e}_j$ may satisfy this criterion, we will transform these optimization problems into Lagrange dual forms. It is well-established that strong duality will hold for these convex problems with linear constraints [1], hence the optimal duality gap is zero. Therefore, instead of directly solving $P1$ and $P2$, one can try to solve the dual problems.

To derive the dual problem for $P1$, let us first define the Lagrangian as follows

$$\mathcal{L}(\mathbf{v}, \boldsymbol{\zeta}, \boldsymbol{\lambda}_u, \boldsymbol{\lambda}_l) = \mathbf{e}_j^T \mathbf{A} \mathbf{v} - \boldsymbol{\zeta}^T (\mathbf{Y} \mathbf{A} \mathbf{v}) + \boldsymbol{\lambda}_u^T (\mathbf{v} - \mathbf{v}_u) + \boldsymbol{\lambda}_l^T (\mathbf{v}_l - \mathbf{v}) , \quad \boldsymbol{\lambda}_u, \boldsymbol{\lambda}_l \geq \mathbf{0} .$$

The Lagrange dual function is then defined as follows

$$\tilde{g}(\boldsymbol{\zeta}, \boldsymbol{\lambda}) = \inf_{\mathbf{v}} \mathcal{L}(\mathbf{v}, \boldsymbol{\zeta}, \boldsymbol{\lambda}) , \quad \boldsymbol{\lambda} \stackrel{\text{def}}{=} (\boldsymbol{\lambda}_u, \boldsymbol{\lambda}_l) .$$

Since the Lagrangian is an affine function of $\mathbf{v}$, it is unbounded from below, hence its infimum value is in general $-\infty$, unless the slope of the affine function is zero, that is

$$\nabla_{\mathbf{v}} \mathcal{L} = \mathbf{A}^T \mathbf{e}_j - \mathbf{A}^T \mathbf{Y}^T \boldsymbol{\zeta} + \boldsymbol{\lambda}_u - \boldsymbol{\lambda}_l = \mathbf{0} ,$$

in which case, the Lagrange dual function would be simplified to the following function

$$g(\boldsymbol{\lambda}) = \boldsymbol{\lambda}_l^T \mathbf{v}_l - \boldsymbol{\lambda}_u^T \mathbf{v}_u$$

Hence, the dual problem is

$$\begin{aligned} & \text{maximize}_{\boldsymbol{\lambda}, \boldsymbol{\zeta}} && \boldsymbol{\lambda}_l^T \mathbf{v}_l - \boldsymbol{\lambda}_u^T \mathbf{v}_u \\ & \text{subject to} && \mathbf{A}^T \mathbf{e}_j - \mathbf{A}^T \mathbf{Y}^T \boldsymbol{\zeta} + \boldsymbol{\lambda}_u - \boldsymbol{\lambda}_l = \mathbf{0} \\ & && \boldsymbol{\lambda}_u, \boldsymbol{\lambda}_l \geq \mathbf{0} \end{aligned} \quad (D1) .$$

Similarly, one can process the optimization problem $P2$ to arrive at the following dual formulation

$$\begin{aligned} & \underset{\lambda, \zeta}{\text{maximize}} && \lambda_l^T \mathbf{v}_l - \lambda_u^T \mathbf{v}_u \\ & \text{subject to} && -\mathbf{A}^T \mathbf{e}_j + \mathbf{A}^T \mathbf{Y}^T \boldsymbol{\zeta} + \lambda_u - \lambda_l = \mathbf{0} \quad (D2) \ . \\ & && \lambda_u, \lambda_l \geq \mathbf{0} \end{aligned}$$

For $\mathbf{e}_j$ to represent a BC, both problems $D1$ and $D2$ must have zero optimal values, i.e.

$$d_1^* = d_2^* = 0 \ .$$

It is worth noting that, while the primal problems $P1$ and $P2$ have a slightly different objective but share the same constraints, the dual problems share the same objective function while having a slightly different feasible set.

5.3 S1.3 Analyzing the dual problems(s) in a canonical flux regime

5.3.1 S1.3.1 Analyzing the dual objective

The objective function for the dual problems can be written as

$$\tilde{g}(\boldsymbol{\lambda}) = \lambda_l^T \mathbf{v}_l - \lambda_u^T \mathbf{v}_u = \sum_{R \in \mathcal{R}} (\lambda_{l,R} v_{l,R} - \lambda_{u,R} v_{u,R}) \ . \quad (3)$$

To simplify this objective function, let us assume the CRN operates in a *canonical flux regime*, that is the bounds on reaction fluxes are standardized as follows

$$\begin{aligned} & \mathbf{v}_u > \mathbf{0} \ ; \\ & v_{l,R} < 0, \quad \forall R \in \mathcal{R}^{\text{rev}} \ ; \\ & v_{l,R} = 0, \quad \forall R \in \mathcal{R}^{\text{irr}} \ ; \end{aligned}$$

where $v_{l,R}$ denotes the R^{th} entry of vector $\mathbf{v}_l$, corresponding to reaction $R \in \mathcal{R}$. In a canonical regime, the irreversible reactions are exactly those corresponding to zero entries of $\mathbf{v}_l$, while the rest of entries in $\mathbf{v}_l$ must be negative and correspond to reversible reactions.

In a canonical regime, it is easy to verify that all terms on the right hand side of Eq. (3) are nonpositive, given the dual feasibility conditions $\lambda_u, \lambda_l \geq \mathbf{0}$. For the sum of all those terms to equal zero, every single term in the RHS must be zero. At any optimal point, we must have

$$\begin{aligned} & \lambda_{u,R}, v_{u,R} = 0 \ , \quad \forall R \in \mathcal{R} \ ; \\ & \lambda_{l,R}, v_{l,R} = 0 \ , \quad \forall R \in \mathcal{R} \ . \end{aligned}$$

This, in turn, yields

$$\begin{aligned} & \lambda_{u,R} = 0, \quad \forall R \in \mathcal{R} \ ; \\ & \lambda_{l,R} = 0, \quad \forall R \in \mathcal{R}^{\text{rev}} \ . \end{aligned}$$

Therefore, for the complex balancing to hold, out of all entries in λ_u, λ_l , only those associated with lower bounds of irreversible reactions can take arbitrary values (though they must be nonnegative due to dual feasibility). The rest of entries in λ_u, λ_l must all be zero.

$$\begin{aligned}\lambda_u &= \mathbf{0}, \\ \lambda_l &= \lambda_l^{\text{irr}}.\end{aligned}$$

Note that λ_u, λ_l having the above form is a necessary condition for the dual problem(s) to admit feasible solution(s) with an objective value of zero. Moreover, if such a solution is found, then it is clearly an optimal point.

It is also worth noting that under a canonical flux regime, the upper bounds on flux levels (as well as negative lower bounds on flux through reversible reactions) have no impact on the formation of balanced complexes. Hence, as far as the emergence of balanced complexes is concerned, it makes sense to just ignore those upper bounds on fluxes and simply study the resulting steady state flux cone.

Moreover, for some $\mathbf{e}_j$ to represent a BC, the dual variables must also satisfy the other constraints in dual problems $D1$ and $D2$.

$$\mathbf{A}^T \mathbf{e}_j = \mathbf{A}^T \mathbf{Y}^T \boldsymbol{\zeta}_1 + \lambda_{l1}^{\text{irr}}, \quad (4-1)$$

$$\mathbf{A}^T \mathbf{e}_j = \mathbf{A}^T \mathbf{Y}^T \boldsymbol{\zeta}_2 - \lambda_{l2}^{\text{irr}}. \quad (4-2)$$

Both the above equations must hold, for the dual problems $D1$ and $D2$ to both have zero optimal values, that is, $d_1^* = d_2^* = 0$.

Therefore, Any vector $\mathbf{e}_j$ which satisfies the above two equations for arbitrary parameter values $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2$ and $\lambda_{l1}^{\text{irr}}, \lambda_{l2}^{\text{irr}} \geq 0$, provides a zero optimal solution for the primal optimization problem and hence, represents a balanced complex for the given CRN. In other words, this can be seen as a necessary and sufficient condition, under a canonical flux regime.

5.3.2 S1.3.2 Relation to stoichiometric factorization

Let us further study the two equalities in Eq. (4). First of all, both vectors $\lambda_{l1}^{\text{irr}}, \lambda_{l2}^{\text{irr}}$ must lie in the range of $\mathbf{A}^T$, that is, the row space of $\mathbf{A}$. Hence, we can introduce a change of variables

$$\exists \boldsymbol{\theta}_k : \lambda_{lt}^{\text{irr}} = \mathbf{A}^T \boldsymbol{\theta}_t, \quad t = 1, 2.$$

Given the discussion in Section S1.3.1, each new variables $\boldsymbol{\theta}_t$ must satisfy the following constraints,

$$\begin{cases} (\mathbf{A}^{\cdot R})^T \boldsymbol{\theta}_t = 0, & \forall R \in \mathcal{R}^{\text{rev}} \\ (\mathbf{A}^{\cdot R})^T \boldsymbol{\theta}_t \geq 0, & \forall R \in \mathcal{R}^{\text{irr}} \end{cases}, \quad t = 1, 2. \quad (5)$$

Therefore, one can replace the parameters $\lambda_{l1}^{\text{irr}}, \lambda_{l2}^{\text{irr}}$ in Eq. (4) with their equivalents in terms of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, while integrating extra constraints [of dual feasibility] given in Eq. (5).

Using the new parameter θ_1 , Eq. (4-1) can be written as

$$\mathbf{A}^T \mathbf{e}_j = \mathbf{A}^T \mathbf{Y}^T \boldsymbol{\zeta}_1 + \mathbf{A}^T \boldsymbol{\theta}_1 \quad (6-1),$$

in which, $\boldsymbol{\zeta}_1$ and $\boldsymbol{\theta}_1$ are independent parameters that are only constrained to satisfy Eq. (5).

The general solution for $\mathbf{e}_j$ in Eq. (6-1) will be of the form

$$\mathbf{e}_j = \langle \text{any particular solution} \rangle + \text{Ker}(\mathbf{A}^T).$$

Assuming the CRN is closed, $\text{Ker}(\mathbf{A}^T)$ has the well-known structure discussed in Section S1.1. Therefore, from Eq. (4), one obtains the following general expressions for $\mathbf{e}_j$

$$\begin{cases} \mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta}_1 + \mathbf{U} \boldsymbol{\xi}_1 + \boldsymbol{\theta}_1 & (7-1) \\ \mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta}_2 + \mathbf{U} \boldsymbol{\xi}_2 - \boldsymbol{\theta}_2 & (7-2) \end{cases},$$

for independent parameters $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, subject to the constraints expressed in Eq. (5).

A special case for $\mathbf{e}_j$ may be obtained by setting $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = 0$. Note that this choice of parameter values also satisfies the constraints in Eq. (5). With $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = 0$, Eq. (7) reduces to two identical equations of the form

$$\mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta}_1 + \mathbf{U} \boldsymbol{\xi}_1,$$

which is exactly the stoichiometric factorization presented in Eq. (2). Therefore, the factorization in Eq. (7), obtained by solving the dual optimization problems, can be viewed as a generalization of the stoichiometric factorization.

In a canonical flux regime, a vector $\mathbf{e}_j$ is said to represent a nonstoichiometric BC, if it has a factorization corresponding to Eq. (7), but no stoichiometric factorization [corresponding to Eq. (2)]. For reasons that will become clear soon, any such a BC is referred to as a *type-I nonstoichiometric BC*.

Now, if one partitions the incidence matrix $\mathbf{A}$ into reversible and irreversible blocks as in $\mathbf{A} = [\mathbf{A}^{\text{rev}} \mathbf{A}^{\text{irr}}]$, then the constraints in Eq. (5) can be rewritten in the following simpler block form

$$\begin{cases} \mathbf{A}^{\text{rev}T} \boldsymbol{\theta}_k = 0 \\ \mathbf{A}^{\text{irr}T} \boldsymbol{\theta}_k \geq 0 \end{cases}, \quad k = 1, 2. \quad (8)$$

5.3.3 S1.3.3 The factorization of BCs in a canonical flux regime

Assuming the system operates in a canonical flux regime, the discussion above can serve as a proof for the following statement.

Theorem S1. Suppose the network is operating in a canonical flux regime. There exist variables $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in \mathbb{R}^{|\mathcal{S}|}$, $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in \mathbb{R}^{\ell}$, $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}^{|\mathcal{C}|}$, such that

$$\begin{cases} \mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta}_1 + \mathbf{U} \boldsymbol{\xi}_1 + \boldsymbol{\theta}_1 \\ \mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta}_2 + \mathbf{U} \boldsymbol{\xi}_2 - \boldsymbol{\theta}_2 \\ \mathbf{A}^{\text{rev}T} \boldsymbol{\theta}_1 = \mathbf{0} \\ \mathbf{A}^{\text{irr}T} \boldsymbol{\theta}_1 \geq \mathbf{0} \\ \mathbf{A}^{\text{rev}T} \boldsymbol{\theta}_2 = \mathbf{0} \\ \mathbf{A}^{\text{irr}T} \boldsymbol{\theta}_2 \geq \mathbf{0} \end{cases}, \quad (9)$$

if and only if the complex $C_j \in \mathcal{C}$ is a BC.

Due to the linearity of the above factorizations and constraints, Eq. (9) can serve as a linear feasibility problem, with variables $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, to check whether any given complex $C_j \in \mathcal{C}$ is a BC of the network. In view of the highly efficient computational tools available for LP problems, this provides an alternative approach for finding balanced complexes. What comes on top of it as a bonus is the fact that it allows one to systematically identify the underlying factors contributing to the formation of any BC.

Let us further reflect on these constraints. Out of all variables in the factorization, only $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ are constrained by the following constraints, and $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ are not. Hence, one may transform the constraints in Eq. (9) into the following -more compact- form

$$\begin{cases} \mathbf{A}^{\text{rev}T}(\mathbf{e}_j - \mathbf{Y}^T \boldsymbol{\zeta}_1 - \mathbf{U} \boldsymbol{\xi}_1) = 0 \\ \mathbf{A}^{\text{irr}T}(\mathbf{e}_j - \mathbf{Y}^T \boldsymbol{\zeta}_1 - \mathbf{U} \boldsymbol{\xi}_1) \geq 0 \\ \mathbf{A}^{\text{rev}T}(\mathbf{Y}^T \boldsymbol{\zeta}_2 + \mathbf{U} \boldsymbol{\xi}_2 - \mathbf{e}_j) = 0 \\ \mathbf{A}^{\text{irr}T}(\mathbf{Y}^T \boldsymbol{\zeta}_2 + \mathbf{U} \boldsymbol{\xi}_2 - \mathbf{e}_j) \geq 0 \end{cases}, \quad (10)$$

which can replace Eq. (9) in the LP framework. This may be more practical in terms of implementation, but not necessarily in terms of clarification, because it fails to provide an explicit factorization of $\mathbf{e}_j$, unlike Eq. (9).

5.3.4 S1.3.4 Formation of type-I nonstoichiometric BCs

The following statement sheds light on the properties on type-I nonstoichiometric BCs and how they are formed.

Proposition S2. Let a network G contain a type-I nonstoichiometric BC. There exist at least two blocked irreversible reactions in the network.

Proof: Let $\mathbf{e}_j$ represent the type-I nonstoichiometric BC. It immediately follows from the definition that we must have

$$\boldsymbol{\theta}_1 \neq \mathbf{0}, \quad \boldsymbol{\theta}_2 \neq \mathbf{0}, \quad (11-1)$$

as well as,

$$\boldsymbol{\lambda}_{lt}^{\text{irr}} = \mathbf{A}^T \boldsymbol{\theta}_t \neq \mathbf{0}, \quad t = 1, 2. \quad (11-2)$$

The balancing relation itself implies that

$$\mathbf{e}_j^T \mathbf{A} \mathbf{v} = \mathbf{0} , \quad \forall \mathbf{v} \in \mathcal{F}(G) .$$

Replacing $\mathbf{e}_j$ by its factorization equations yields

$$\boldsymbol{\theta}_t^T \mathbf{A} \mathbf{v} = \mathbf{0} , \quad t = 1,2, \quad \forall \mathbf{v} \in \mathcal{F}(G) ,$$

that is,

$$\begin{aligned} \boldsymbol{\lambda}_{l1}^{\text{irr}T} \mathbf{v} &= \mathbf{0} \\ \boldsymbol{\lambda}_{l2}^{\text{irr}T} \mathbf{v} &= \mathbf{0} \end{aligned} , \quad \forall \mathbf{v} \in \mathcal{F}(G) . \quad (12)$$

Note that

$$\boldsymbol{\lambda}_{lt}^{\text{irr}T} \mathbf{v} = \sum_{R \in \mathcal{R}^{\text{irr}}} \lambda_{lt,R} v_R , \quad t = 1,2 .$$

Having the abovementioned structure of the dual variables $\boldsymbol{\lambda}_{lt}^{\text{irr}}$, $t = 1,2$, and knowing that $v_R \geq 0$, $\forall R \in \mathcal{R}^{\text{irr}}$, it follows from Eq. (12) that

$$\lambda_{lt,R} v_R = 0 , \quad t = 1,2, \quad \forall R \in \mathcal{R}^{\text{irr}} .$$

Taking into consideration that Eq. (11-2) implies $\boldsymbol{\lambda}_{lt}^{\text{irr}}$ has at least one positive entry for both $t = 1,2$, it follows that for $\mathbf{e}_j$ to represent a nonstoichiometric BC, at least one irreversible reaction must be blocked.

Moreover, the two vectors $\boldsymbol{\lambda}_{lt}^{\text{irr}}$, $t = 1,2$ cannot be collinear; otherwise one can cancel out those terms in Eq. (4), which yields a stoichiometric factorization for the BC, which would be a contradiction. Therefore, there exist at least two blocked reactions in the network. ■

Corollary S3. Let G be a network operating under a canonical flux regime, all blocked reactions of which have been removed. Then all balanced complexes of G have stoichiometric factorizations.

5.4 S1.4 Analyzing the dual problem(s) for general flux bounds

5.4.1 S1.4.1 Derivation of an equivalent LP

Next, we consider the network under general -not necessarily canonical- flux bounds, in which some irreversible reactions may also take positive lower bounds on flux levels. Irrespective of the flux bounds, for a vector $\mathbf{e}_j$ to represent a BC, the dual optimization problems $D1$ and $D2$ must take zero optimal values.

$$g(\boldsymbol{\lambda}) = \boldsymbol{\lambda}_{lt}^T \mathbf{v}_l - \boldsymbol{\lambda}_{ut}^T \mathbf{v}_u = 0 , \quad t = 1,2 .$$

Here is what contrast this case from the canonical flux regime: Some terms in the dual objective may now take positive values; hence we are not in a position to further simplify the objective function by asserting that every term on the RHS has to be zero, as we did in Section S1.3.1.

A complex represented by vector $\mathbf{e}_j$ is balanced if and only if both dual problems $D1$ and $D2$ have zero optimal values. Now, we represent the equivalent of Theorem S1 for the more general case:

Theorem S4. A complex $C_j \in \mathcal{C}$ is a BC, if and only if there exist variables $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in \mathbb{R}^{|\mathcal{S}|}$, $\boldsymbol{\lambda}_{u1}, \boldsymbol{\lambda}_{l1}, \boldsymbol{\lambda}_{u2}, \boldsymbol{\lambda}_{l2} \in \mathbb{R}^{|\mathcal{R}|}$ such that

$$\begin{cases} \mathbf{A}^T \mathbf{e}_j - \mathbf{A}^T \mathbf{Y}^T \boldsymbol{\zeta}_1 + \boldsymbol{\lambda}_{u1} - \boldsymbol{\lambda}_{l1} = \mathbf{0} \\ \mathbf{v}_l^T \boldsymbol{\lambda}_{l1} - \mathbf{v}_u^T \boldsymbol{\lambda}_{u1} = 0 \\ -\mathbf{A}^T \mathbf{e}_j + \mathbf{A}^T \mathbf{Y}^T \boldsymbol{\zeta}_2 + \boldsymbol{\lambda}_{u2} - \boldsymbol{\lambda}_{l2} = \mathbf{0} \\ \mathbf{v}_l^T \boldsymbol{\lambda}_{l2} - \mathbf{v}_u^T \boldsymbol{\lambda}_{u2} = 0 \\ \boldsymbol{\lambda}_{u1}, \boldsymbol{\lambda}_{l1} \geq \mathbf{0} \\ \boldsymbol{\lambda}_{u2}, \boldsymbol{\lambda}_{l2} \geq \mathbf{0} \end{cases} \quad . \quad (13)$$

Proof: let us momentarily take a break from directly addressing optimality conditions for $D1$ and $D2$, and instead try to seek feasible variables that set the corresponding dual objectives to zero. Hence, we now seek variables $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in \mathbb{R}^{|\mathcal{S}|}$, $\boldsymbol{\lambda}_{u1}, \boldsymbol{\lambda}_{l1}, \boldsymbol{\lambda}_{u2}, \boldsymbol{\lambda}_{l2} \in \mathbb{R}^{|\mathcal{R}|}$ that satisfy Eq. (13).

Note that $(\boldsymbol{\zeta}_1, \boldsymbol{\lambda}_{u1}, \boldsymbol{\lambda}_{l1})$ gives a feasible point for $D1$, while $(\boldsymbol{\zeta}_2, \boldsymbol{\lambda}_{u2}, \boldsymbol{\lambda}_{l2})$ gives a feasible point for $D2$. Therefore, if one denotes the optimal (maximal) values of $D1$ and $D2$ by d_1^*, d_2^* , respectively, the feasibility of these points implies that

$$d_1^* \geq 0, \quad d_2^* \geq 0. \quad (14)$$

On the other hand, p_1^* is the minimum of some function $(\mathbf{e}_j^T \mathbf{A} \mathbf{v})$ on the steady state flux set, while $-p_2^*$ is the maximum of the same function on the same set. Hence, $p_1^* \leq -p_2^*$. Therefore,

$$p_1^* + p_2^* \leq 0.$$

As a result, the general properties of the duality [1] ensure that

$$d_1^* + d_2^* \leq p_1^* + p_2^* \leq 0. \quad (15)$$

Taking into account Eqs. (13) and (15), it follows that

$$d_1^* = d_2^* = 0.$$

Therefore, even though we did not explicitly enforce optimality in Eq. (13), by the virtue of Eq. (5), any set of variables that satisfies Eq. (13) automatically provides optimality, hence it translates to solutions for the dual optimization problems $D1$ and $D2$, with an achieved optimal value of zero.

On the other hand, if vector $\mathbf{e}_j$ represents a BC, it has to lead to zero optimal values for the dual problem(s), hence there exist variables $\zeta_1, \zeta_2 \in \mathbb{R}^{|S|}$, $\lambda_{u1}, \lambda_{l1}, \lambda_{u2}, \lambda_{l2} \in \mathbb{R}^{|R|}$ that satisfy Eq. (13). ■

As a result, in the general –not necessarily canonical- case, Eq. (13) replaces the duality problems, and may serve as a linear feasibility problem to check whether a complex $\mathbf{e}_j$ is a BC.

5.4.2 S1.4.2 Nonstoichiometric balancing with general boundary constraints

Given the two equalities containing $\mathbf{e}_j$ in Eq. (13), it follows that the terms $\lambda_{u1} - \lambda_{l1}$ and $\lambda_{u2} - \lambda_{l2}$ must lie in the rowspan of $\mathbf{A}$; therefore, similar to Eq. (7), one can introduce new variables θ_1 and θ_2 , as follows

$$\begin{aligned}\lambda_{l1} - \lambda_{u1} &= \mathbf{A}^T \theta_1, \\ \lambda_{l2} - \lambda_{u2} &= \mathbf{A}^T \theta_2.\end{aligned}$$

Using new variables θ_1 and θ_2 , one would arrive at the same factorization structure as in Eq. (7) for the balanced complex $\mathbf{e}_j$, which could pave the way for labeling of $\mathbf{e}_j$ as either a stoichiometric or nonstoichiometric balancing relation. However, note that the other constraints in Eq. (9) do not hold generally, under non-canonical flux regimes. In fact, the solution set of Eq. (9) is a subset of the solution set of Eq. (13), for which those additional constraints also hold.

Instead of using new variables θ_1 and θ_2 to identify nonstoichiometric BC, one can resort to the following equivalent definition: A vector $\mathbf{e}_j$ represents a *nonstoichiometric BC*, if it has an implicit factorization corresponding to Eq. (13), but no factorization of the form Eq. (2).

As we have seen with Proposition S2 and Corollary S3, blocked irreversible reactions play a key role in the formation of (type-I) nonstoichiometric BCs in a canonical flux regime. It follows that they can play exactly the same role in networks operating under general flux bounds. However, one might ask the question: Can other factors also lead to the formation of BCs under general –not necessarily canonical- flux bounds?

To answer this question, let us suppose all blocked reactions have been removed from the network. It follows from Corollary S3 that no type-I nonstoichiometric BC may exist in this network. We aim to investigate whether possibly different forms of nonstoichiometric BCs may emerge in this network. Let us define $\mathcal{R}^{\text{pos}} \subseteq \mathcal{R}^{\text{irr}}$ as the set of irreversible reactions with positive lower bounds on flux levels. Let us now assume there exists a nonstoichiometric BC in this network, represented by vector $\mathbf{e}_j$. From the definition of a nonstoichiometric BC, it follows that

$$\begin{aligned}\lambda_{l1} - \lambda_{u1} &\neq 0, \\ \lambda_{l2} - \lambda_{u2} &\neq 0.\end{aligned}$$

Moreover, it can be shown that the above two vectors cannot be collinear. Otherwise the equality

$$(\lambda_{l1} - \lambda_{u1}) + (\lambda_{l2} - \lambda_{u2}) = \mathbf{A}^T \mathbf{Y}^T (\zeta_2 - \zeta_1),$$

would mean that both $(\lambda_{lt} - \lambda_{ut})$, $t = 1, 2$ can be replaced by a scalar factor of $\mathbf{A}^T \mathbf{Y}^T (\zeta_2 - \zeta_1)$, hence $\mathbf{e}_j$ would have a stoichiometric factorization, which is a contradiction. Therefore, the two vectors $(\lambda_{lt} - \lambda_{ut})$, $t = 1, 2$ are linearly independent. The other constraints in Eq. (13),

$$\begin{aligned} \lambda_{lt}^T \mathbf{v}_l - \lambda_{ut}^T \mathbf{v}_u &= 0, \quad t = 1, 2, \\ \lambda_{ut}, \lambda_{lt} &\geq 0, \end{aligned}$$

alongside $\mathbf{v}_u > \mathbf{0}$ imply that $\lambda_{ut}^T \mathbf{v}_u \geq 0$, $\lambda_{lt}^T \mathbf{v}_l \geq 0$, $t = 1, 2$; and we must also have

$$\begin{aligned} \lambda_{l1} &\neq 0, \\ \lambda_{l2} &\neq 0. \end{aligned}$$

In addition, it is easy to show [by contradiction]

$$\exists R_1, R_2 \in \mathcal{R}^{\text{pos}} : \lambda_{l1, R_1} > 0, \quad \lambda_{l2, R_2} > 0, \quad (16)^1$$

in which, $\mathcal{R}^{\text{pos}}$ denotes the set of (irreversible) reactions with a positive flux lower bound.

From the definition of the balanced complex, we know that $\mathbf{e}_j^T \mathbf{A} \mathbf{v} = \mathbf{0}$, $\forall \mathbf{v} \in \mathcal{F}(G)$.

Replacing $\mathbf{e}_j^T \mathbf{A} = (\mathbf{A}^T \mathbf{e}_j)^T$ by its equivalent from Eq. (13) then yields

$$(\lambda_{ut} - \lambda_{lt})^T \mathbf{v} = \mathbf{0}, \quad \forall \mathbf{v} \in \mathcal{F}(G), \quad t = 1, 2. \quad (17)$$

Therefore, existence of a nonstoichiometric balancing relation in the network implies the existence of at least two independent linear couplings of reaction fluxes.

In addition, via combining Eq. (17) with Eqs. (13-2) and (13-4), one obtains

$$\lambda_{ut}^T (\mathbf{v} - \mathbf{v}_u) + \lambda_{lt}^T (\mathbf{v}_l - \mathbf{v}) = \mathbf{0}, \quad \forall \mathbf{v} \in \mathcal{F}(G), \quad t = 1, 2.$$

However, we know that

$$\lambda_u \geq \mathbf{0}, \quad \mathbf{v} - \mathbf{v}_u \leq \mathbf{0}, \quad \lambda_l \geq \mathbf{0}, \quad \mathbf{v}_l - \mathbf{v} \geq \mathbf{0},$$

Therefore, all reactions for which at least one corresponding entry of $\lambda_{u1}, \lambda_{u2}, \lambda_{l1}, \lambda_{l2}$ is nonzero, must be fixated at their respective lower/upper bounds for all values of $\mathbf{v}$ in the steady state flux set. In particular, Eq. (16) ensures that for at least one irreversible reaction with a positive lower bound is fixated at the minimum for all $\mathbf{v} \in \mathcal{F}(G)$.

Proposition S5. Let G be a network all blocked reactions of which have been removed. Suppose G contains a nonstoichiometric BC. There exist at least one irreversible reaction with a positive lower bound $R \in \mathcal{R}^{\text{pos}}$, such that $v_R = v_{l,R}$, $\forall \mathbf{v} \in \mathcal{F}(G)$.

¹ Otherwise, it can be shown that some irreversible reactions must be blocked.

The following statement follows immediately from Proposition S5.

Proposition S6. Let G be a network all blocked reactions of which have been removed. Suppose G contains a nonstoichiometric BC. There exist at least three reactions in G , which are fixated at a corresponding [nonzero] lower- or upper bound, for all $\mathbf{v} \in \mathcal{F}(G)$.

Proof: From Eq. (16), we know that $\exists R_1 \in \mathcal{R}^{\text{pos}} : \lambda_{11,R_1} > 0$, and the reaction R_1 is fixated at the corresponding lower bound. From the BC constraints in Eq. (17), we have $\lambda_{11}^T \mathbf{v}_1 - \lambda_{u1}^T \mathbf{v}_u = \lambda_{11,R_1} v_{1,R_1} + \sum_{\rho \neq R_1} \lambda_{11,\rho} v_{1,\rho} - \sum_{\rho} \lambda_{u1,\rho} v_{u,\rho} = 0$. Since $\lambda_{11,R_1} v_{1,R_1} > 0$, it follows that

$$\exists R'_1 \neq R_1 : \lambda_{11,R'_1} > 0, v_{1,R'_1} < 0 \quad \text{or} \quad \exists R'_1 \neq R_1 : \lambda_{u1,R'_1} > 0, v_{u,R'_1} > 0.$$

Therefore, there exist at least two distinct reactions $R_1 \in \mathcal{R}^{\text{pos}}, R'_1 \in \mathcal{R}$ fixated at their corresponding flux bounds.

With the same logic, the other constraint $\lambda_{12}^T \mathbf{v}_1 - \lambda_{u2}^T \mathbf{v}_u = 0$ implies that there exist at least two distinct reactions $R_2 \in \mathcal{R}^{\text{pos}}, R'_2 \in \mathcal{R}$ fixated at their corresponding flux bounds. Now, if there exist any other reaction $R_3 \notin \{R_1, R'_1, R_2, R'_2\}$ for which the strict inequality $\lambda_{1t,R_3} + \lambda_{ut,R_3} > 0$ holds for some $t \in \{1,2\}$, then we have found the three fixated reactions and the proof is complete.

If that is not the case, then we cannot have both $R_1 = R_2$ and $R'_1 = R'_2$; otherwise the two vectors $(\lambda_{11} - \lambda_{u1}), (\lambda_{12} - \lambda_{u2})$ would become collinear, which is a contradiction. Therefore, there exist at least three reactions fixated at nonzero upper- or lower bounds. ■

5.4.3 S1.4.3 Type-I vs. type-II nonstoichiometric BCs

Let us remind that in networks operating under a canonical flux regime, any BC encoded by vector $\mathbf{e}_j$ must satisfy the factorization scheme presented in Eq. (7). We showed that for a nonstoichiometric BC to exist, at least two irreversible reactions must be blocked at steady state. It can also be shown that those blocked reactions cannot lie within any strong linkage class of the network, but each must be a “bridge” reaction that connects distinct strong linkage classes within a linkage class². We refer to this class of nonstoichiometric BCs as *type-I*.

It is not surprising that type-I nonstoichiometric BCs can also emerge in non-canonical flux regimes. To seek type-I nonstoichiometric BCs under general flux bounds, one can basically

² We showed that $\lambda_i^{\text{irr}} \geq 0$ must be of the form $\lambda_i^{\text{irr}} = \mathbf{A}^T \boldsymbol{\theta}$. It follows that the algebraic sum of the entries in λ_i^{irr} corresponding to any reaction cycle (loop in the graph) must be zero. Taking into account also the fact that $\lambda_i^{\text{irr}} \geq \mathbf{0}$ and $\lambda_i^{\text{rev}} = \mathbf{0}$, it follows that every entry of λ_i^{irr} corresponding to a reaction inside a strong linkage class must be zero.

apply the same formulation as Eq. (9). There is, however, a minor caveat due to the potential existence of positive lower bounds on fluxes. Eq. (9) was introduced in the context of canonical flux bounds, where $\mathbf{A}^{\text{irr}}$ denoted exactly those columns of $\mathbf{A}$ corresponding to zero lower bounds in $\mathbf{v}_l$, and $\mathbf{A}^{\text{rev}}$ was its complementary block in $\mathbf{A}$. To stay consistent with that convention in more general case, the columns corresponding to positive lower bounds need to be removed from $\mathbf{A}^{\text{irr}}$, and instead transferred to $\mathbf{A}^{\text{rev}}$.³ Thereby, $\mathbf{A}^{\text{irr}}$ corresponds to $\mathcal{R}^{\text{irr}} \setminus \mathcal{R}^{\text{pos}}$ and $\mathbf{A}^{\text{rev}}$ corresponds to $\mathcal{R}^{\text{rev}} \cup \mathcal{R}^{\text{pos}}$. Whenever we speak of a factorization of the form Eq. (9) in the context of networks operating under general flux bounds, we also mean to incorporate this minor discretion.

The obtained factorization scheme demonstrates an interesting property of type-I nonstoichiometric BCs: the emergence of such BCs does not rely on the value of the flux bounds (specified by $\mathbf{v}_l$, $\mathbf{v}_u$). Consequently, BCs of this type originate from the network topology, and in particular, the existing irreversibility patterns. It follows that potential modeling errors in the form of inaccuracies in determining upper- and lower-bounds on fluxes has no impact on our ability to detect such BCs in a given network.

What is less desirable about them is the fact that they emerge in conjunction with blocked reactions in the network. Alternatively, suppose we first remove those blocked “bridge” reactions from the network. This would break down some linkage classes of the network into a number of smaller linkage classes, which means we would come up with a reduced network of slightly different linkage structure. This, in turn, expands the left nullspace of the incidence matrix $\mathbf{A}$, that is, the matrix $\mathbf{U}$ will contain a higher number of columns, i.e. basis vectors. It follows that the stoichiometric factorization,

$$\exists \boldsymbol{\zeta}, \boldsymbol{\xi} : \quad \mathbf{e}_j = \mathbf{Y}^T \boldsymbol{\zeta} + \mathbf{U} \boldsymbol{\xi} ,$$

may have a larger solution set in the case of the reduced network. This means, the reduced network obtained by the removal of blocked reactions may accommodate a higher number of stoichiometric BCs. In fact, one can show that all type-I nonstoichiometric BCs of the original network appear as stoichiometric BCs in the reduced network.

A somewhat different phenomenon is the case of nonstoichiometric BCs that may only arise under non-canonical flux regimes, that is, when at least one irreversible reaction has a positive lower bound. We hereby refer to such nonstoichiometric BCs as *type-II*. As was shown in Section S1.4.1, all BCs must satisfy the implicit factorization given in Eq. (13). Assuming all blocked reactions have been removed from some network G , any remaining nonstoichiometric BCs in G must be of type-II. Clearly, the existence of type-II nonstoichiometric BCs in G relies on having a nonempty set $\mathcal{R}^{\text{pos}}$.

Even though type-II nonstoichiometric BCs emerge in conjunction with the fixation of some reaction fluxes at upper- or lower bounds, as was shown in Section S1.4.2, they do not automatically rely on the network having some blocked reactions that effectively change its linkage structure. This is an interesting property of type-II nonstoichiometric BCs, because of which they can be seen as an unexpected nontrivial phenomenon in the steady state analysis,

³ This replaces the original partitioning $\mathbf{A} = [\mathbf{A}^{\text{rev}} \ \mathbf{A}^{\text{irr}}]$ into $\mathbf{A} = [\mathbf{A}^{\text{zr}} \ \mathbf{A}^{\text{nz}}]$, where $\mathbf{A}^{\text{zr}}$ denotes the columns corresponding to (irreversible) reactions with zero lower bounds.

by contrast to type-I nonstoichiometric BCs, which we associated with unsophisticated network reductions.

Looking back at Eq. (13), the potential existence of a type-II nonstoichiometric BC encoded by $\mathbf{e}_j$ directly relies on the existence of feasible dual variables $\lambda_{ut}, \lambda_{lt}$, $t = 1, 2$ that fit into the pattern $\lambda_{lt} - \lambda_{ut} \in \text{im}(\mathbf{A}^T)$, $t = 1, 2$. However, $\lambda_{lt}, \lambda_{ut}$, $t = 1, 2$ are not free variables, but are tied to each other via other equations in Eq. (13), that is

$$\begin{aligned} \mathbf{v}_l^T \lambda_{l1} - \mathbf{v}_u^T \lambda_{u1} &= 0, \\ \mathbf{v}_l^T \lambda_{l2} - \mathbf{v}_u^T \lambda_{u2} &= 0. \end{aligned}$$

This casts light on an undesirable property of type-II nonstoichiometric BCs: the values $\mathbf{v}_l, \mathbf{v}_u$ appear explicitly in their factorizations. It follows that potential modeling errors in the form of inaccuracies in determining upper- and lower-bounds on fluxes may have a negative impact and result in a failure to correctly detect type-II nonstoichiometric BCs: if the value of some upper/lower bounds in the model ($\mathbf{v}_l, \mathbf{v}_u$) are modified or miscalculated, this could change the values of $\lambda_{l1} - \lambda_{u1}$ and $\lambda_{l2} - \lambda_{u2}$, which could mean they would not necessarily satisfy the factorization for a given $\mathbf{e}_j$ anymore. This sensitivity of detection to the values of upper- and lower-bounds is an unwelcome feature with type-II nonstoichiometric BCs.

6 S2 Analysis of biological networks

We categorized balanced complexes identified by Küken et al. [2] in twelve genome-scale metabolic models from organisms of all kingdoms of life, namely *A. niger* [3], *A. thaliana* [4], *C. reinhardtii* [5], *E. coli* [6], *M. acetivorans* [7], *M. barkeri* [8], *M. musculus* [9], *M. tuberculosis* [10], *N. pharaonis* [11], *P. putida* [12], *T. maritima* [13] and *S. cerevisiae* [14].

Balanced complexes were identified in networks including no blocked reactions; therefore, these networks do not include type-I nonstoichiometric BCs. To check if a BC C_j is a strictly stoichiometric BC we solve the feasibility problem in Eq. (18), where $\mathbf{e}_j$ is the vector with a unit value for the j^{th} entry and zero values elsewhere and $\mathbf{Y}$ being the stoichiometric map. The linear program will be feasible if $\mathbf{e}_j$ is a linear combination of rows in $\mathbf{Y}$.

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \quad 0 \\ &\text{subject to} \quad \mathbf{Y}^T \mathbf{x} = \mathbf{e}_j \end{aligned} \quad (18) \quad .$$

Moreover, Küken et al. identified BCs under different sets of constraints, including the scenario where reaction reversibility is considered and canonical flux bounds are imposed and the scenario where the lower bound of the biomass reaction is set to at least 99% of its optimum obtained from flux balance analysis [15]. Comparing the sets of BCs found under the respective scenarios allows to identify type-II nonstoichiometric BCs, which are those complexes being balanced under imposed lower bound on biomass, but not under canonical flux bounds. The remaining BCs are then classified as stoichiometric BCs.

7 References

- [1] S. Boyd and L. Vandenberghe, "Duality," in *Convex Optimization*, Cambridge University Press, 2004, pp. 215-288.
- [2] A. Küken, P. Wendering, D. Langary and Z. Nikoloski, "A structural property for reduction of biochemical networks," *bioRxiv*, p. 2021.03.17.435785, 1 2021.
- [3] M. R. Andersen, M. L. Nielsen and J. Nielsen, "Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger*," *Molecular Systems Biology*, vol. 4, p. 178, 1 2008.
- [4] A. Arnold and Z. Nikoloski, "Bottom-up Metabolic Reconstruction of Arabidopsis and Its Application to Determining the Metabolic Costs of Enzyme Production," *Plant Physiology*, vol. 165, pp. 1380-1391, 5 2014.
- [5] S. Imam, S. Schäuble, J. Valenzuela, A. López García de Lomana, W. Carter, N. D. Price and N. S. Baliga, "A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses," *The Plant Journal*, vol. 84, pp. 1239-1256, 11 2015.
- [6] J. D. Orth, T. M. Conrad, J. Na, J. A. Lerman, H. Nam, A. M. Feist and B. Ø. Palsson, "A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011," *Molecular Systems Biology*, vol. 7, p. 535, 1 2011.
- [7] M. N. Benedict, M. C. Gonnerman, W. W. Metcalf and N. D. Price, "Genome-Scale Metabolic Reconstruction and Hypothesis Testing in the Methanogenic Archaeon *Methanosarcina acetivorans* C2A," *Journal of Bacteriology*, vol. 194, p. 855–865, 12 2012.
- [8] A. M. Feist, J. C. M. Scholten, B. Ø. Palsson, F. J. Brockman and T. Ideker, "Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*," *Molecular Systems Biology*, vol. 2, p. 2006.0004, 1 2006.
- [9] L.-E. Quek and L. K. Nielsen, "ON THE RECONSTRUCTION OF THE MUSCULUS GENOME-SCALE METABOLIC NETWORK MODEL," in *Genome Informatics 2008*, IMPERIAL COLLEGE PRESS, 2008, pp. 89-100.
- [10] X. Fang, A. Wallqvist and J. Reifman, "Development and analysis of an in vivo-compatible metabolic network of *Mycobacterium tuberculosis*," *BMC Systems Biology*, vol. 4, p. 160, 2010.
- [11] O. Gonzalez, T. Oberwinkler, L. Mansueto, F. Pfeiffer, E. Mendoza, R. Zimmer and D. Oesterhelt, "Characterization of Growth and Metabolism of the Haloalkaliphile *Natronomonas pharaonis*," *PLOS Computational Biology*, vol. 6, pp. 1-10, 6 2010.

- [12] J. Nogales, B. Ø. Palsson and I. Thiele, "A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory," *BMC Systems Biology*, vol. 2, p. 79, 2008.
- [13] Y. Zhang, I. Thiele, D. Weekes, Z. Li, L. Jaroszewski, K. Ginalska, A. M. Deacon, J. Wooley, S. A. Lesley, I. A. Wilson, B. Palsson, A. Osterman and A. Godzik, "Three-Dimensional Structural View of the Central Metabolic Network of *Thermotoga maritima*," *Science*, vol. 325, p. 1544–1549, 9 2009.
- [14] H. Lu, F. Li, B. J. Sánchez, Z. Zhu, G. Li, I. Domenzain, S. Marcišauskas, P. M. Anton, D. Lappa, C. Lieven, M. E. Beber, N. Sonnenschein, E. J. Kerkhoven and J. Nielsen, "A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism," *Nature Communications*, vol. 10, p. 3586, 8 2019.
- [15] J. D. Orth, I. Thiele and B. Ø. Palsson, "What is flux balance analysis?," *Nature biotechnology*, vol. 28, p. 245–248, 3 2010.