*Article*

# An Automated Model Reduction Method for Biochemical Reaction Networks

**Manvel Gasparyan [1,2,\*]** [ID]**, Arnout Van Messem [1,3]** [ID] **and Shodhan Rao [1,2]** [ID]

[1]  Center for Biotech Data Science, Ghent University Global Campus, Incheon 21985, Korea; arnout.vanmessem@ugent.be (A.V.M.); shodhan.rao@ugent.be (S.R.)

[2]  Department of Data Analysis and Mathematical Modeling, Ghent University, 9000 Ghent, Belgium

[3]  Department of Applied Mathematics, Computer Science and Statistics, Ghent University, 9000 Ghent, Belgium

**\***  Correspondence: manvel.gasparyan@ugent.be

check for updates

**Abstract:** We propose a new approach to the model reduction of biochemical reaction networks governed by various types of enzyme kinetics rate laws with non-autocatalytic reactions, each of which can be reversible or irreversible. This method extends the approach for model reduction previously proposed by Rao et al. which proceeds by the step-wise reduction in the number of complexes by Kron reduction of the weighted Laplacian corresponding to the complex graph of the network. The main idea in the current manuscript is based on rewriting the mathematical model of a reaction network as a model of a network consisting of linkage classes that contain more than one reaction. It is done by joining certain distinct linkage classes into a single linkage class by using the conservation laws of the network. We show that this adjustment improves the extent of applicability of the method proposed by Rao et al. We automate the entire reduction procedure using Matlab. We test our automated model reduction to two real-life reaction networks, namely, a model of neural stem cell regulation and a model of hedgehog signaling pathway. We apply our reduction approach to meaningfully reduce the number of complexes in the complex graph corresponding to these networks. When the number of species' concentrations in the model of neural stem cell regulation is reduced by 33.33%, the difference between the dynamics of the original model and the reduced model, quantified by an error integral, is only 4.85%. Likewise, when the number of species' concentrations is reduced by 33.33% in the model of hedgehog signaling pathway, the difference between the dynamics of the original model and the reduced model is only 6.59%.
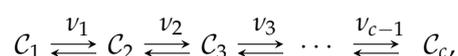
**Keywords:** graph theory; biochemistry; chemical reaction network theory; systems biology; mathematical modeling; weighted Laplacian matrix; conservation laws; complex balancing; Schur complement; Kron reduction

## 1. Introduction

A mathematical model of a chemical reaction network (CRN) can be represented by a system of ordinary differential equations (ODEs) that describe the dynamics of the concentrations of the constituent species participating in the network. Generally, kinetic models of most CRNs involve an immensely large number of variables, thereby making the models difficult to analyze due to the huge computational effort required for the analysis. On top of that, as in general not all of the species' concentrations can be measured in an experimental set-up, identifying the parameters contained in the mathematical model is not feasible and demands massive experimental datasets. Thus, there exists a necessity to simplify the mathematical models of CRNs to reduced variants that preserve the key properties of the original models but contain fewer ordinary differential equations

and fewer unknown variables. Numerous model reduction methods for CRNs are described in the literature. For a comprehensive survey of popular model reduction techniques, see, for example, in [1,2]. The quasi-steady-state approximation (QSSA) (see, for example, in ([3] Chapter 2) and ([4] Chapter 4)) is one of the most commonly used tools for model reduction of CRNs governed by enzyme kinetics. Even though QSSA can be used for a large variety of CRNs, in many cases it demands high computational efforts. The method is not simple to apply, as it requires a priori experimental (biological) knowledge.

A recent model reduction method proposed in [5] reduces the mathematical model of a CRN by eliminating complexes from its corresponding complex graph. In chemical reaction network theory, complexes are defined as the left- (substrate) and right-hand (product) sides of the reactions. Furthermore, the complexes of a CRN can be naturally associated with the vertices of a complex graph with edges corresponding to the reactions of the network. The method described in [5] selects the complexes to be eliminated in such a way that the behavior of the remaining relevant species in the reduced CRN is close to their original behavior. It is based on the Kron reduction method [6], and reduces the weighted Laplacian matrix associated with the complex graph. The Kron reduction method is a popular method of model reduction for electrical networks and proceeds by the computation of Schur complements [7] of the weighted Laplacian matrix. As an example, let us consider the following CRN consisting of a single linkage class,
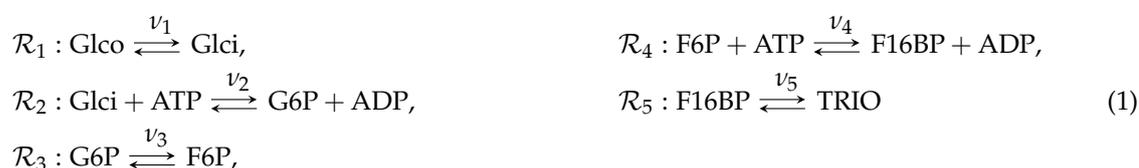
$$\mathcal{C}_1 \xrightleftharpoons{\nu_1} \mathcal{C}_2 \xrightleftharpoons{\nu_2} \mathcal{C}_3 \xrightleftharpoons{\nu_3} \cdots \xrightleftharpoons{\nu_{c-1}} \mathcal{C}_c,$$

where $\mathcal{C}_j$, $j = 1, \cdots, c$, are distinct complexes and $\nu_j$, $j = 1, \cdots, c-1$, are the overall reaction rates in the forward direction. We recall from the work in [5] that a *linkage class* is a connected component of the complex graph corresponding to the CRN. After eliminating the complex $\mathcal{C}_2$ using the method proposed in [5], we obtain the following reduced network.

$$\mathcal{C}_1 \xrightleftharpoons{\bar{\nu}} \mathcal{C}_3 \xrightleftharpoons{\nu_3} \cdots \xrightleftharpoons{\nu_{c-1}} \mathcal{C}_c.$$
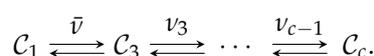
Here, the principle of *complex balancing* is applied to $\mathcal{C}_2$ to replace the two reversible reactions $\mathcal{C}_1 \xrightleftharpoons{\nu_1} \mathcal{C}_2$ and $\mathcal{C}_2 \xrightleftharpoons{\nu_2} \mathcal{C}_3$ of the original network by the new reversible reaction $\mathcal{C}_1 \xrightleftharpoons{\bar{\nu}} \mathcal{C}_3$. We say that $\mathcal{C}_2$ is complex balanced if its inflow is equal to its outflow. In other words, the concentrations of the species participating in $\mathcal{C}_2$ are adjusted in such a way that $\nu_1 = \nu_2$, which leads to the deletion of $\mathcal{C}_2$ from the complex graph. Moreover, the reaction rates $\nu_j$, $j = 3, \cdots, c-1$, of the other reactions remain unchanged. Note that the goal of considering this example is to illustrate how the reduction procedure in [5] is carried out. It describes how the aforementioned method is used to delete a complex (in this case the complex $\mathcal{C}_2$) from the corresponding complex graph. Thus, without loss of generality, we assume that the deletion of $\mathcal{C}_2$ does not cause major changes in the dynamics of the considered chemical reaction network.

This method is effective in eliminating intermediate complexes from linkage classes that have more than one reaction. We note that an intermediate complex is a complex participating in more than one reaction, with each reversible reaction considered as a single reaction. For instance, with reference to the network given above, $\mathcal{C}_2$ is an intermediate complex, but $\mathcal{C}_1$ is not. However, as mentioned in [5,8], when every linkage class has just two complexes, the method fails in producing a meaningful reduction as it just deletes reactions from the network. For example, consider the following part of a kinetic model of yeast glycolysis,

$$\mathcal{R}_1 : \text{Glco} \xrightleftharpoons{\nu_1} \text{Glci}, \qquad\qquad\qquad \mathcal{R}_4 : \text{F6P} + \text{ATP} \xrightleftharpoons{\nu_4} \text{F16BP} + \text{ADP},$$

$$\mathcal{R}_2 : \text{Glci} + \text{ATP} \xrightleftharpoons{\nu_2} \text{G6P} + \text{ADP}, \qquad \mathcal{R}_5 : \text{F16BP} \xrightleftharpoons{\nu_5} \text{TRIO} \qquad\qquad (1)$$

$$\mathcal{R}_3 : \text{G6P} \xrightleftharpoons{\nu_3} \text{F6P},$$

where $v_j$ is the overall rate of the $j$th reaction in the forward direction. Note that a reaction rate is a function of species' concentrations of the network and it depends on the governing law of the corresponding reaction. In this case, elimination of any complex by the approach proposed in [5] results in deleting the corresponding reaction. For instance, if we delete the complex G6P + ADP from the network, the method will result in deleting the entire second reaction $\mathcal{R}_2$ from the network. The reduced CRN will then have the reaction $\mathcal{R}_1$ occurring independently of the remaining reactions. Therefore, the reduced network will possess a behavior that is completely different from the original one. It is easy to see that the complex graph corresponding to (1) is not connected. As this complex graph does not have any linkage class that contains more than one reaction, the method proposed in [5] does not result in a meaningful reduction of the corresponding model. In this case, there is a need to find a reasonable method of rewriting the complex graph of the given network as a graph with linkage classes that have more than one reaction in order to ensure that the application of the method proposed in [5] results in a meaningful reduction in the number of complexes.

We propose to eliminate the species ADP from the network in such a way that we can replace the reactions $\mathcal{R}_2$ and $\mathcal{R}_4$ with the reversible reactions Glci + ATP $\rightleftharpoons$ G6P and F6P + ATP $\rightleftharpoons$ F16BP, respectively. This replacement can be done, for example, by expressing the concentration of ADP in terms of the concentrations of ATP. This can be done by using the two ODEs that represent the change in concentration over time for ADP and ATP. These ODEs can be represented in the following form,

$$\frac{d[\text{ADP}]}{dt} = v_2 + v_4, \qquad\qquad \frac{d[\text{ATP}]}{dt} = -v_2 - v_4,$$

where $[X]$ denotes the molar concentration of the species $X$. From these equations, we easily derive the following conservation law, which express the concentrations of ADP in terms of the concentration of ATP,

$$[\text{ADP}] \;=\; -[\text{ATP}] + C, \tag{2}$$

where $C$ is a non-negative constant (known as a conserved quantity) depending on the initial concentrations of ATP and ADP. The conservation law (2) can now be used to eliminate the concentration of the species ADP from the mathematical model of the network. This elimination allows us to rewrite the mathematical model as a model of a network containing two linkage classes that consist of more than one reaction.

$$\text{Glco} \rightleftharpoons \text{Glci}$$
$$\text{Glci} + \text{ATP} \rightleftharpoons \text{G6P} \rightleftharpoons \text{F6P} \tag{3}$$
$$\text{F6P} + \text{ATP} \rightleftharpoons \text{F16BP} \rightleftharpoons \text{TRIO}$$

Note that even though the species ADP is not explicitly participating anymore in the CRN, its concentration still appears in the corresponding mathematical model in terms of the concentration of ATP. The method in [5] may now be used on the network (3) to reduce its mathematical model in a meaningful way by deleting the species (complexes) G6P and F16BP from the corresponding complex graph. Even though the elimination of the concentration of ADP enables the application of the reduction method [5], the reduced model is valid only for trajectories for which the total concentration of the pool of ADP and ATP is fixed.

Most CRNs in real-life have reaction network structures similar to the original yeast glycolysis network (1), with certain linkage classes consisting of only one reaction. New tools for model reduction are required for these types of networks. In this paper, we generalize the approach of the complex graph rewriting procedure briefly illustrated above. This procedure, in conjunction with the method proposed in [5], results in a meaningful model reduction.

The main idea of our model reduction method is to enable the application of the model reduction method proposed in [5] by joining certain linkage classes of a CRN into a single linkage class using

linearly independent conservation laws. An important result related to conservation laws is Noether's theorem [9], which states that there is a one-to-one correspondence between a conservation law and a differentiable symmetry of nature. For instance, an example of a conservation law in the field of classical physics is the law of conservation of energy in Hamiltonian systems, which corresponds to time translation symmetry.

We develop an automation process for our model reduction method in order to make it straightforward to apply for a wide range of CRNs governed by any kind of enzyme kinetics. In the automation procedure we assume that the CRN admits a single steady state and is asymptotically stable around it. This assumption is necessary for determining the settling time of the network over which we compare the original model and the reduced model. We created a MATLAB library, which contains all the files corresponding to each step of our step-wise reduction procedure. It uses the input files containing the information of a given CRN provided by the user to reduce the corresponding mathematical model in a fully automated fashion.

Finally, we exhibit the application of our new technique of model reduction for real-life examples of CRNs drawn from the Biomodels database [10]. For each of the cases, we explain how the corresponding mathematical model can be meaningfully reduced.

## 2. Preliminaries

In this section, we give a compact description of the mathematical techniques that are necessary for our automated model identification method. We commence by introducing the notations that are used throughout the manuscript and continue by explaining a framework for deriving a system of ODEs that describes the dynamics of the constituent species of a given CRN. Subsequently, we recall the Kron reduction method proposed in [5].

$\mathcal{S} \rightleftharpoons \mathcal{P}$ denotes a reaction from the substrate $\mathcal{S}$ to the product $\mathcal{P}$, where the dotted line indicates that it can be reversible or irreversible. In this case, when referring to its reaction rate we refer to the outgoing overall reaction rate in the forward direction. The transpose and the null space of an $m \times n$ matrix $A = [A_{ij}]$, $j = 1, \cdots, n$, $i = 1, \cdots, m$, are denoted by $A^\top$ and $\ker(A)$, respectively. The number of elements of a set $I$ is denoted by $n(I)$. For any vector $v \in \mathbb{R}^n$, denote by $\mathrm{diag}(v)$ the $n \times n$ diagonal matrix whose diagonal elements are the corresponding elements of the vector $v$. Define the function $\mathrm{Exp} : \mathbb{R}^m \to \mathbb{R}^m_+$ as $[x_i]_{i=1}^m \mapsto [\exp(x_i)]_{i=1}^m$ and the function $\mathrm{Ln} : \mathbb{R}^m_+ \to \mathbb{R}^m$ as $[x_i]_{i=1}^m \mapsto [\ln(x_i)]_{i=1}^m$. The Schur complement of the block matrix $D \in R^{q \times q}$ of the matrix $L \in \mathbb{R}^{p \times p}$ is denoted by $\mathcal{L}_D \in \mathbb{R}^{(p-q) \times (p-q)}$.

### 2.1. Mathematical Models of Reaction Networks

Consider a CRN that involves $m$ distinct chemical species, $c$ complexes, and has $r$ unidirectional reactions. We present a framework for describing the dynamics of reaction networks using their complex graphs [5,11]. The $c$ complexes of the network are represented by an $m \times c$ complex composition matrix $Z$, whose columns express the complexes in terms of their species. Any CRN is defined by a $c \times r$ *incidence* matrix $B$, where

$$B_{ij} = \begin{cases} -1, & \text{if the } i\text{th complex is the substrate of the } j\text{th reaction,} \\ 1, & \text{if the } i\text{th complex is the product of the } j\text{th reaction,} \\ 0, & \text{otherwise.} \end{cases}$$

It can be shown that the $m \times r$ stoichiometric matrix $S$ of the network is $S = ZB$. We denote by $\nu(x) \in \mathbb{R}^r_+$ the vector of reaction rates of the given CRN. The basic structure underlying the dynamics of the species' concentration vector $x \in \mathbb{R}^m_+$ is given by the balance law [12]:

$$\frac{dx}{dt} = S\nu(x). \tag{4}$$

Let $\xi \in \ker(S^\top)$. Using (4) we easily see that this vector satisfies the condition $\xi^\top \frac{dx}{dt} = 0$. Consequently, $\xi^\top x$ is a constant, and thus

$$\xi^\top x = \xi^\top x_0, \tag{5}$$

where $x_0 \in \mathbb{R}_+^m$ is the vector of initial species' concentrations, i.e., $x_0 = x|_{t=0}$. Equation (5) is called a conservation law of the CRN. The constant term $\xi^\top x_0$ appearing in the conservation law (5) is called a conserved quantity. Clearly, the number of linearly independent conservation laws is equal to the dimension of $\ker(S^\top)$. In other words, irrespective of the governing laws, we can obtain a maximal set of linearly independent conservation laws by computing $\ker(S^\top)$.

Let $Z_{\mathcal{S}_i}$ be the column of $Z$ corresponding to the substrate $\mathcal{S}_i$ of the $i$th reaction. Assuming that every reaction of the CRN is governed by an *enzyme kinetic rate law*, the reaction rate of the $i$th unidirectional reaction between the substrate $\mathcal{S}_i$ and product $\mathcal{P}_i$, as explained in [5], can be expressed as

$$v_i(x) = d_i(x)k_i \exp(Z_{\mathcal{S}_i}^\top \mathrm{Ln}(x)), \tag{6}$$

where, for every $i = 1, \cdots, r$, $d_i : \mathbb{R}_+^m \to \mathbb{R}_+$ is a rational function and $k_i$ denotes the rate constant of the $i$th reaction. Note that if the governing law of the $i$th reaction is *mass action kinetics*, then $d_i(x) \equiv 1$. Consider the $r \times r$ diagonal matrices $D(x) = \mathrm{diag}(d_1, \cdots, d_r)$ and $K = \mathrm{diag}(k_1, \cdots, k_r)$. Define the $c \times r$ *outgoing* matrix $\Delta = [\Delta_{ij}]$ as follows,

$$\Delta_{ij} = \begin{cases} 0, & \text{if } B_{ij} = 1, \\ B_{ij}, & \text{otherwise.} \end{cases}$$

It can be shown that the $c \times c$ weighted Laplacian matrix associated with the complex graph of the network is $L(x) = BKD(x)\Delta^\top$. As explained in [5], the balance laws (4) can be expressed in terms of the Laplacian matrix as

$$\frac{dx}{dt} = -ZL(x)C(x), \tag{7}$$

where $C(x) = \mathrm{Exp}\left(Z^\top \mathrm{Ln}(x)\right)$. Some well-known enzyme kinetic rate laws for which the balance laws can be written in matrix multiplication form (7) are the reversible Michaelis–Menten kinetics [13], Monod–Wyman–Changeaux kinetics [14], Hill kinetics [15,16], Ping Pong Bi Bi kinetics [17], and Koshland–Nemety–Filmer kinetics [18]. In order to clearly illustrate the modeling procedure, we demonstrate it for the following reaction network with four unidirectional reactions,

$$\begin{aligned} \varnothing &\longrightarrow X_1 \\ 2X_1 + X_2 &\rightleftharpoons 3X_3 + X_4 \\ X_4 &\longrightarrow \varnothing \end{aligned} \tag{8}$$
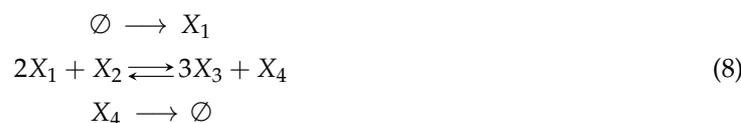
where $\varnothing$ is the zero complex representing the environment. It is treated as a regular complex, in which the number of moles of each species is zero as explained in [11]. More precisely, the first reaction is a creation of $X_1$ and the last reaction is a degradation of $X_4$. In this example, the substrate complexes of the four reactions are $\mathcal{S}_1 = \varnothing$, $\mathcal{S}_2 = 2X_1 + X_2$, $\mathcal{S}_3 = 3X_3 + X_4$, and $\mathcal{S}_4 = X_4$, respectively. Thus, $Z_{\mathcal{S}_1} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^\top$, $Z_{\mathcal{S}_2} = \begin{bmatrix} 2 & 1 & 0 & 0 \end{bmatrix}^\top$, $Z_{\mathcal{S}_3} = \begin{bmatrix} 0 & 0 & 3 & 1 \end{bmatrix}^\top$, and $Z_{\mathcal{S}_4} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\top$. The complex composition matrix and the incidence matrix corresponding to the network (8) are given by

$$Z = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Let $V_f$ and $V_r$ be the maximum reaction rates of the second and the third reactions of (8), respectively. The expressions for the reaction rates (6) depend on the governing laws of the reactions. One possibility is

$$v_1(x) = p_1,$$

$$v_2(x) = \frac{\frac{V_f}{K_1 K_2} x_1^2 x_2}{1 + \frac{x_1}{K_1} \frac{x_2}{K_2} + \frac{x_3}{K_3} \frac{x_4}{K_4}},$$

$$v_3(x) = \frac{\frac{V_r}{K_3 K_4} x_3^3 x_4}{1 + \frac{x_1}{K_1} \frac{x_2}{K_2} + \frac{x_3}{K_3} \frac{x_4}{K_4}},$$

$$v_4(x) = p_2 x_4,$$

where $K_i$ is the Michaelis constant of the species $X_i$ for $i = 1, \cdots, 4$, and $p_1$, $p_2$ are positive constants. The reaction rates of the network (8) can be represented in the form (6) by defining the rate constants $k_i$ as

$$k_1 = p_1, \qquad k_2 = \frac{V_f}{K_1 K_2}, \qquad k_3 = \frac{V_r}{K_3 K_4}, \qquad k_4 = p_2,$$

and the rational terms $d_i(x)$ in the expressions of reaction rates as

$$d_1(x) = 1,$$

$$d_2(x) = \frac{1}{1 + \frac{x_1}{K_1} \frac{x_2}{K_2} + \frac{x_3}{K_3} \frac{x_4}{K_4}},$$

$$d_3(x) = \frac{1}{1 + \frac{x_1}{K_1} \frac{x_2}{K_2} + \frac{x_3}{K_3} \frac{x_4}{K_4}},$$

$$d_4(x) = 1.$$

The $5 \times 5$ Laplacian matrix corresponding to the network (8) is

$$L(x) = \begin{bmatrix} k_1 d_1(x) & 0 & 0 & 0 & -k_4 d_4(x) \\ -k_1 d_1(x) & 0 & 0 & 0 & 0 \\ 0 & 0 & k_2 d_2(x) & -k_3 d_3(x) & 0 \\ 0 & 0 & -k_2 d_2(x) & k_3 d_3(x) & 0 \\ 0 & 0 & 0 & 0 & k_4 d_4(x) \end{bmatrix}.$$

Note that open CRNs are modeled in [5] by extended balance laws of the form $\dfrac{dx}{dt} = -ZL(x)C(x) + Zf(x)$, where the free term $f \in \mathbb{R}^c$ represents the inflows to and outflows from the complexes of the network. Here, as the zero complex is considered as a regular complex it is included in the Laplacian matrix in order to avoid the unnecessary free term. On top of that, unlike the reduction method proposed in [5], if the zero complex is an intermediate complex, we consider it as a candidate complex for deletion.

*2.2. Kron Reduction of Mathematical Models Corresponding to Reaction Networks*

We describe the model reduction method proposed in [5]. It is based on the Kron reduction [6] and is performed by deleting a set of complexes from the complex graph associated with the CRN. Recall that deleting a complex is equivalent to assuming that it is complex balanced. The complex balancing is carried out by computing the *Schur complement* [7] of the weighted Laplacian matrix associated with the corresponding complex graph.

**Definition 1.** *Let L be an $(m + n) \times (m + n)$ block matrix of the form*

$$L = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

*where A, B, C, and D are $n \times n$, $n \times m$, $m \times n$, and $m \times m$ matrices, respectively. Assume that D is invertible. The Schur complement of D is the $n \times n$ matrix $\mathcal{L}_D$ defined by*

$$\mathcal{L}_D = A - BD^{-1}C.$$

Consider again a CRN of $c$ complexes and $r$ reactions with dynamics modeled by the balance laws (7). Let $V$ be the set of indices corresponding to the complexes of the CRN, i.e., $V = \{1, \cdots, c\}$. Assume that we want to delete the complexes with the set of indices $\widehat{V}$, which is a subset of $V$ of cardinality] $\widehat{c} < c$. Denote by $D(x) \in \mathbb{R}^{\widehat{c} \times \widehat{c}}$ the block matrix of the Laplacian matrix $L(x)$ that corresponds to the set of indices $\widehat{V}$. The complex deletion is carried out by computing the Schur complement $\mathcal{L}_D(x) \in \mathbb{R}^{(c-\widehat{c}) \times (c-\widehat{c})}$ (see, for example, in [7]) of the block matrix $D(x)$. It is proven (see in [5] Proposition 1) that $\mathcal{L}_D(x)$ is again a Laplacian matrix, i.e., it satisfies the properties of Laplacian matrices [19]. On top of that, it is shown that the equation

$$\frac{dx}{dt} = -Z_{\text{red}} \mathcal{L}_D(x) \text{Exp}(Z_{\text{red}} \text{Ln}(x))$$

describes the dynamics of a CRN governed by an enzyme kinetics rate law, i.e., the corresponding reaction rates can be written in the form (6), with smaller number of complexes. Here, $Z_{\text{red}}$ is the complex composition matrix of the reduced CRN defined by removing from $Z$ the columns corresponding to the set of indices $\widehat{V}$. A suitable choice of $\widehat{V}$ ensures that some of the elements in the vector of species concentrations $x$ are conserved in time, i.e., the derivatives of some of the elements of $x$ are zeros. This leads to a reduced number of dependent variables in the corresponding mathematical model. As an example consider the following simple CRN of two unidirectional reactions,

$$X_1 \rightleftharpoons X_2 + X_3 \rightleftharpoons X_4.$$

Let $v_1$ and $v_2$ be the overall rates in the forward direction of the first reaction and the second reaction, respectively. For $i = 1, \cdots, 4$, denote by $x_i$ the concentration of the species $X_i$. In this case, the balance laws (4) can be rewritten as

$$\frac{dx_1}{dt} = -v_1 \qquad\qquad \frac{dx_2}{dt} = \frac{dx_3}{dt} = v_1 - v_2 \qquad\qquad \frac{dx_4}{dt} = v_2.$$

Now suppose that the second complex $X_2 + X_3$ is complex balanced, which is equivalent to assuming that $v_1 = v_2$. From the above-mentioned balance laws we obtain $\dfrac{dx_2}{dt} = \dfrac{dx_3}{dt} = 0$, which means that $x_2$ and $x_3$ are conserved in time.

In [5], the difference between the dynamical behaviors of the original model and the corresponding reduced model is quantified by an error integral $E$. Let $[0, T]$ be the time interval over which we want to observe the difference between the behaviors of the original model and the reduced model. Denote by $J$ the set of indices of the significant species of the network. The error integral $E$ is defined as

$$E(T) = \frac{1}{Tn(J)} \sum_{i \in J} \int_0^T \left| 1 - \frac{y_i(t)}{x_i(t)} \right| dt, \tag{9}$$

where $x_i$ and $y_i$ are the concentrations of the $i$th species in the original model and the reduced model, respectively. Note that in [5] the right endpoint of the time interval has been set manually. However, in our case this right endpoint is chosen to be the settling time of the network, i.e., the time instant at which the considered CRN reaches a sufficiently small prespecified neighborhood of its unique steady state.

The model reduction method proposed in [5] is an iterative procedure that selects a subset of intermediate complexes to be deleted from the corresponding complex graph as follows. First, the candidate complexes for deletion, i.e., the intermediate complexes of the network, are identified.

For each candidate complex the error integral corresponding to the reduced model obtained after its deletion is evaluated. These values of error integrals are used to rank the candidate complexes and then delete the candidate complex with the least value of error integral. In the next step, the above-mentioned procedure is repeated with the reduced model. The iteration is continued until the error integral $E(T)$ exceeds a prespecified threshold value.

One of the advantages of the reduction method [5] is that it does not rely on a priori biological knowledge about the CRN. Moreover, it is an easy-to-implement model reduction method that can be used to simplify huge mathematical models of CRNs by retaining the original kinetics.

## 3. Equivalent Mathematical Models

In this section, we give a detailed explanation of the basic idea of our reduction technique which we briefly introduced in Section 1. We first establish concepts that are relevant to the model reduction procedure.

**Definition 2.** *The index set $I_{\mathcal{C}_j} \subset \{1, \cdots, m\}$, $j \in \{1, \cdots, c\}$, corresponding to the $j^{th}$ complex $\mathcal{C}_j$ of the network, is the set of indices of the species participating in the expression of the complex $\mathcal{C}_j$.*

We thus obtain the following expression for the complex $\mathcal{C}_j$ in terms of its species,

$$\mathcal{C}_j = \sum_{i \in I_{\mathcal{C}_j}} a_{\mathcal{C}_j, i} X_i, \ j = 1, \cdots, c,$$

where $a_{\mathcal{C}_j, i} \in \mathbb{R}_+$ represents the number of moles of the *i*th species in the expression of the complex $\mathcal{C}_j$. It is clear that two complexes $\mathcal{C}_p$ and $\mathcal{C}_q$ share common species if and only if $I_{\mathcal{C}_p} \cap I_{\mathcal{C}_q} \neq \emptyset$. Trivially, $\bigcup_{j=1}^{c} I_{\mathcal{C}_j} = \{1, \cdots, m\}$.

**Definition 3.** *Let $\mathcal{C}$ be a complex of a CRN. For $\varepsilon \in \mathbb{N}$, we define $\varepsilon\mathcal{C}$ as the new complex consisting of the species of $\mathcal{C}$ with their respective number of moles multiplied by $\varepsilon$, i.e., $\varepsilon\mathcal{C} = \sum_{i \in I_{\mathcal{C}}} \varepsilon a_{\mathcal{C}, i} X_i$.*

With the following lemma, we show the equivalence between two mathematical models of a given CRN. Later on, we will see that the application of this lemma allows us to join certain reactions into a single linkage class.

**Lemma 1.** *Let $v(x) = [v_j(x)]_{j=1}^r$ be the vector of overall reaction fluxes in the forward direction of the CRN $\mathcal{R}_j : \mathcal{S}_j \xrightarrow{\longleftrightarrow} \mathcal{P}_j$, $j = 1, \cdots, r$. For every $\varepsilon = [\varepsilon_j]_{j=1}^r \in \mathbb{N}^r$, the reaction network is equivalent, in terms of its mathematical model, to the reaction network $\varepsilon_j \mathcal{S}_j \xrightarrow{\longleftrightarrow} \varepsilon_j \mathcal{P}_j$, $j = 1, \cdots, r$, with the vector of overall reaction fluxes being $\bar{v}(x) = \left[ \frac{1}{\varepsilon_j} v_j(x) \right]_{j=1}^r$.*

**Proof.** Let $X_\alpha$, $\alpha \in \{1, \cdots, m\}$ be a random species of the CRN $\varepsilon_j \mathcal{S}_j \xrightarrow{\longleftrightarrow} \varepsilon_j \mathcal{P}_j$, $j = 1, \cdots, r$. Using (4), the rate of change of the concentration $x_\alpha$ of species $X_\alpha$ is given by

$$\begin{aligned}
\frac{dx_\alpha}{dt} &= \sum_{j=1}^r \left( -\varepsilon_j a_{\mathcal{S}_j, \alpha} \left[ \frac{1}{\varepsilon_j} v_j(x) \right] + \varepsilon_j a_{\mathcal{P}_j, \alpha} \left[ \frac{1}{\varepsilon_j} v_j(x) \right] \right) \\
&= \sum_{j=1}^r \left( a_{\mathcal{P}_j, \alpha} - a_{\mathcal{S}_j, \alpha} \right) v_j(x).
\end{aligned}$$

As the right-hand side of this equation represents the rate of concentration change of species $X_\alpha$ in the original network, the proof of Lemma 1 is complete. $\square$

**Remark 1.** *As the CRNs $\mathcal{R}_j : \mathcal{S}_j \xleftrightarrow{} \mathcal{P}_j$, $j = 1, \cdots, r$, with $v(x) = \left[ v_j(x) \right]_{j=1}^r$, and $\overline{\mathcal{R}}_j : \varepsilon_j \mathcal{S}_j \xleftrightarrow{} \varepsilon_j \mathcal{P}_j$, $j = 1, \cdots, r$, with $\overline{v}(x) = \left[ \frac{1}{\varepsilon_j} v_j(x) \right]_{j=1}^r$, have the same mathematical models, they also possess the same conservation laws.*

In the following proposition, we generalize the idea of the complex graph rewriting procedure, as was roughly illustrated in Section 1. We show that in the presence of a shared species between two complexes, a suitable conservation law can be used to join the corresponding reactions with the shared species as intermediate complexes.

**Proposition 1.** *Let $v(x) = \left[ v_j(x) \right]_{j=1}^r$ be the vector of overall reaction rates in the forward direction of the CRN $\mathcal{R}_j : \mathcal{S}_j \xleftrightarrow{} \mathcal{P}_j$, $j = 1, \cdots, r$. Suppose that, for certain $j_1, j_2 \in \{1, \cdots, r\}$, two complexes, $\mathcal{P}_{j_1}$ and $\mathcal{S}_{j_2}$, of different reactions share a common species $X_\alpha$, $\alpha \in \{1, \cdots, m\}$. Define $I = (I_{\mathcal{P}_{j_1}} \cup I_{\mathcal{S}_{j_2}}) \setminus \{\alpha\}$, where $I_{\mathcal{P}_{j_1}}$ and $I_{\mathcal{S}_{j_2}}$ are the sets of indices corresponding to the complexes $\mathcal{P}_{j_1}$ and $\mathcal{S}_{j_2}$, respectively. If for every $\beta \in I$, there exists a conservation law in which the species $X_\beta$ is participating, but not the remaining species corresponding to I, then we can join the two reactions, $\mathcal{R}_{j_1}$ and $\mathcal{R}_{j_2}$, into a single linkage class $\mathcal{R}'$,*

$$\mathcal{R}' : a_{\mathcal{S}_{j_2}, \alpha} \mathcal{S}_{j_1} \xleftrightarrow{\overline{v}_{j_1}} \mathcal{C}_\alpha \xleftrightarrow{\overline{v}_{j_2}} a_{\mathcal{P}_{j_1}, \alpha} \mathcal{P}_{j_2}, \tag{10}$$

*with the corresponding reaction rates being*

$$\overline{v}_{j_1}(x) = \frac{v_{j_1}(x)}{a_{\mathcal{S}_{j_2}, \alpha}} \quad \text{and} \quad \overline{v}_{j_2}(x) = \frac{v_{j_2}(x)}{a_{\mathcal{P}_{j_1}, \alpha}}.$$

*The intermediate complex $\mathcal{C}_\alpha$ is the shared species $X_\alpha$ with number of moles $a_{\mathcal{C}_\alpha, \alpha} = a_{\mathcal{P}_{j_1}, \alpha} a_{\mathcal{S}_{j_2}, \alpha}$, i.e., $\mathcal{C}_\alpha = a_{\mathcal{P}_{j_1}, \alpha} a_{\mathcal{S}_{j_2}, \alpha} X_\alpha$.*

**Proof.** For every $\beta \in I$, choose a vector $\xi \in \ker(S^\top)$ such that the corresponding conservation law (5) satisfies the assumptions of the proposition. Note that this vector strictly depends on $\beta$. More precisely, its $\beta$th component $\xi_\beta$ is not equal to zero. Moreover, note that all the other components of $\xi$ corresponding to $I$ are zeros. Denote by $\widetilde{\xi}$ the vector $\xi$ with its $\beta$th element replaced by zero. Using (5) we represent $x_\beta$ in terms of the concentrations of the other species:

$$x_\beta = \frac{1}{\xi_\beta} \left( -\widetilde{\xi}^\top x + \xi^\top x_0 \right),$$

where $x_0$ is the vector of species' initial concentrations. By substituting $x_\beta$ as in the above equation in the expression of the vector $v(x)$ of the reaction rates, we can rewrite the reactions $\mathcal{S}_{j_k} \xleftrightarrow{} \mathcal{P}_{j_k}$, $k = 1, 2$, in the following forms

$$\mathcal{S}_{j_1} \xleftrightarrow{} a_{\mathcal{P}_{j_1}, \alpha} X_\alpha, \qquad a_{\mathcal{S}_{j_2}, \alpha} X_\alpha \xleftrightarrow{} \mathcal{P}_{j_2}.$$

Application of Lemma 1 with

$$\varepsilon_j = \begin{cases} 1, & j \neq j_1, j_2, \\ a_{\mathcal{S}_{j_2}, \alpha} & j = j_1, \\ a_{\mathcal{P}_{j_1}, \alpha} & j = j_2 \end{cases} \tag{11}$$

completes the proof. □

As (10) represents a linkage class with more than one reaction, the method proposed in [5] can be meaningfully applied to delete the intermediate complex $\mathcal{C}_\alpha$ from the corresponding complex

graph. In the reduced network, the linkage class (10) can be replaced by a single reaction $\overline{\mathcal{R}}$ :
$a_{\mathcal{S}_{j_2},\alpha}\mathcal{S}_{j_1} \xleftarrow{\quad} a_{\mathcal{P}_{j_1},\alpha}\mathcal{P}_{j_2}$ using the principle of complex balancing.

**Remark 2.** *We would like to point out the three major steps of the complex graph rewriting procedure described in the proof of Proposition 1. Later on, these steps will be used to automate our model reduction procedure.*

M1. *Elimination of certain species from the network (rewriting the corresponding concentrations in terms of the concentrations of the other species), after which certain complexes are composed of the same species.*
M2. *Application of Lemma 1 in order to make such complexes identical to each other.*
M3. *Joining the reactions with identical complexes into a single linkage class.*

## 4. Automatic Reduction Procedure

In this section, we describe the automatic model reduction procedure step by step. We show how to apply Proposition 1 for any kind of CRN independently of its governing laws to join certain reactions in an automatic way. We then apply the strategy in [5] to reduce the number of complexes in the complex graph corresponding to the equivalent network.

Inputs: The list of inputs required for the model reduction procedure is as follows.

- The complex composition matrix $Z \in \mathbb{R}^{m \times c}$, where $m$ and $c$ are the number of the species and complexes of the network, respectively.
- The incidence matrix $B \in \mathbb{R}^{c \times r}$ of the network, where $r$ is the number of reactions of the network.
- The vector $k \in \mathbb{R}^r_+$ of rate constants of the reactions.
- The vector $d(x) \in \mathbb{R}^r_+$ of rational terms in the expressions of reaction rates.
- The vector $x_0 \in \mathbb{R}^m_+$ of initial concentrations.
- The threshold value of the error integral (9), i.e., the maximum admissible value of $E$.

Outputs: The automatic reduction procedure provides the following outputs.

- The mathematical model and the complex graph corresponding to the original network.
- The mathematical model and the complex graph corresponding to the reduced network.
- The final value of the error integral.

We divide our model reduction procedure into seven steps. In order to clearly explain the reduction procedure throughout this section, we will demonstrate it step by step for the following example of six unidirectional reactions:
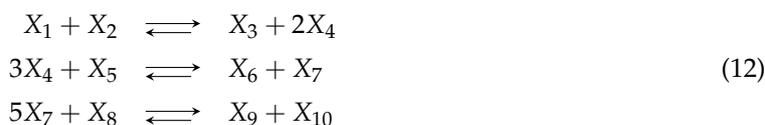
$$\begin{aligned}
X_1 + X_2 &\rightleftharpoons X_3 + 2X_4 \\
3X_4 + X_5 &\rightleftharpoons X_6 + X_7 \\
5X_7 + X_8 &\rightleftharpoons X_9 + X_{10}
\end{aligned} \tag{12}$$

with the corresponding reaction rates being

$$v_1(x) = k_1 x_1 x_2,$$
$$v_3(x) = \frac{k_2 x_4^3 x_5}{1 + x_4 x_5 + x_6 x_7},$$
$$v_5(x) = \frac{k_3 x_7^5 x_8}{1 + x_7 + x_8 + x_9 + x_{10}},$$

$$v_2(x) = k_{-1} x_3 x_4^2,$$
$$v_4(x) = \frac{k_{-2} x_6 x_7}{1 + x_4 x_5 + x_6 x_7},$$
$$v_6(x) = \frac{k_{-3} x_9 x_{10}}{1 + x_7 + x_8 + x_9 + x_{10}}.$$

Step 1: Mathematical model of the network.

In the first step, we determine the system of differential Equation (7) that underlies the dynamics of the species' concentration vector $x \in \mathbb{R}^m_+$.

Step 2: Settling time of the network.

Each CRN is uniquely described by the system (7) obtained from its structure. We can then commence considering its model reduction. An important tool, which is crucial in our reduction procedure, is the settling time of the considered CRN defined in Section 2.2. We show how to automatically compute it for a given CRN.

It has been proved in [20] that for a class of networks called zero deficiency networks, there exists a unique point satisfying the following system,

$$
\begin{cases}
\dfrac{dx}{dt} = 0, \\
\xi_i^\top x = \xi_i^\top x_0, \ i = 1, \cdots, l,
\end{cases}
\tag{13}
$$

where $\xi_i$, $i = 1 \cdots, l$, form a maximal set of linearly independent conservation laws. In [11,21,22], the same result has been proved for a related class of networks called complex balanced networks. It states that such a network is asymptotically stable around its unique steady state corresponding to a particular initial concentration $x_0$. We assume that the same property is true for the CRNs considered for our reduction procedure as mentioned in the introduction. Solving the system (13), we find the unique steady state $x_* \in \mathbb{R}_+^m$. The settling time of the CRN can then be automatically computed by the so called *bound function* defined as follows. For a fixed $\delta \ll 1$, the $\delta$-bound function is given as

$$
b_\delta(x) = \begin{cases}
1, & \text{if } (1-\delta)x_* \leq x \leq (1+\delta)x_*, \\
0, & \text{otherwise.}
\end{cases}
$$

Note that the bound function is computed component-wise. For an appropriate choice of $\delta$, it is an easy numerical task to find the vector $\tau \in \mathbb{R}_+^m$ of settling time, such that $\tau_i$, $i = 1, \cdots, m$, is the time when the bound function corresponding to the $i$th species becomes 1 and remains so.

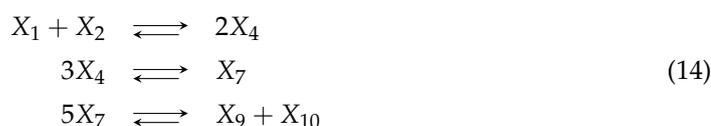Step 3: Selecting species to be eliminated from the network.

Next, we identify the set $I$ of indices corresponding to the species whose elimination from the network results in joining certain reactions. This can be done by using the index sets $I_{\mathcal{C}_j}$, $j = 1, \cdots, c$, corresponding to the complexes $\mathcal{C}_j$ of the network, which can be easily extracted from the complex composition matrix $Z$. For a given pair of complexes $\mathcal{C}_p$ and $\mathcal{C}_q$ that share at least one species $X_\alpha$, define $I_{p,q} = \left( I_{\mathcal{C}_p} \cup I_{\mathcal{C}_q} \right) \setminus \{\alpha\}$. If $\mathcal{C}_p$ and $\mathcal{C}_q$ share no species, set $I_{p,q} = \varnothing$. It is reasonable to define the set $I$ of indices to be eliminated from the network as $I = \bigcup_{p,q} I_{p,q}$. We note that if there exists a complex $\mathcal{C}$, which is made up of a single species $X_\beta$ with $\beta \in I$, i.e., $\mathcal{C} = X_\beta$, the elimination of $X_\beta$ from the network will end up eliminating the entire complex $\mathcal{C}$, meaning that $\mathcal{C}$ becomes the zero complex.

The above-mentioned procedure allows us to determine the set of indices $I$ in a fully automated manner. For example, in the case of the network (12), we obtain $I = \{3, 5, 6, 8\}$.
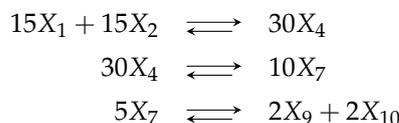
Step 4: Mathematical model of the equivalent network.

In this step, we show how to determine the mathematical model of the equivalent network obtained after eliminating the species corresponding to $I$ from the original network. For every $\beta \in I$, we find the corresponding conservation law that can be used with Proposition 1 to eliminate the concentration $x_\beta$ from the mathematical model of the original network. It can be done by computing $\ker(S_\beta^\top)$, where $S_\beta$ is the matrix obtained by replacing all the rows of the stoichiometric matrix $S$ that correspond to the elements of $I \setminus \{\beta\}$, with zero rows. If the assumptions of Proposition 1 are satisfied, then it can be used to join certain reactions. The mathematical model of the equivalent network is then determined according to the steps described in Remark 2. We give a detailed explanation of how $Z$, $B$, $k$, $d$, and $x_0$ for the corresponding model change after each step.

M1. As the eliminated species are no longer participating in the equivalent network, we delete the corresponding rows from $Z$. Subsequently, the vector $\widetilde{x} \in \mathbb{R}^{\widetilde{m}}$ of species' concentrations is defined by deleting the $\beta$th, $\beta \in I$, element of $x$. Similarly, we obtain the vector $\widetilde{x}_0 \in \mathbb{R}^{\widetilde{m}}$ of initial concentrations of the equivalent network. The vector $\widetilde{d}(\widetilde{x})$ of the rational functions can be derived from $d(x)$ in the following way. For every $\beta \in I$, if the species $X_\beta$ is participating in the substrate complex of $j$th reaction, then $\widetilde{d}_j(\widetilde{x}) = x_\beta^{-S_{\beta j}} d_j(x)$, which ensures that the reaction fluxes of the network obtained at this step still obey the Equation (6). For the network (12), it is clear that the species $X_3$ is participating in the substrate complex of the second reaction. After rewriting the concentration function $x_3$ in terms of concentrations of the other species, we multiply $d_2$ by $x_3$. Thus, $\widetilde{d}_2(\widetilde{x}) = x_3 d_2(x)$. Similarly, we have $\widetilde{d}_3(\widetilde{x}) = x_5 d_3(x)$, $\widetilde{d}_4(\widetilde{x}) = x_6 d_4(x)$, and $\widetilde{d}_5(\widetilde{x}) = x_8 d_5(x)$. We therefore write the reactions of the network (12) in the following form.

$$
\begin{aligned}
X_1 + X_2 &\rightleftharpoons 2X_4 \\
3X_4 &\rightleftharpoons X_7 \\
5X_7 &\rightleftharpoons X_9 + X_{10}
\end{aligned}
\tag{14}
$$

M2. After the elimination of species, the columns of $Z$ corresponding to the complexes that share species become multiples of each other. Consequently, in order to make these columns identical to each other, we multiply each of them by the corresponding constant from (11). Likewise, we divide each rate constant $k_j$ by the corresponding constant given in (11). In the case of the network (12), the vector of rate constants of its corresponding equivalent network is $\widetilde{k} = \left[ \frac{k_1}{15}, \frac{k_{-1}}{30}, \frac{k_2}{30}, \frac{k_{-2}}{10}, \frac{k_3}{10}, \frac{k_{-3}}{2} \right]$. We therefore rewrite the reactions (14) in the following form.

$$
\begin{aligned}
15X_1 + 15X_2 &\rightleftharpoons 30X_4 \\
30X_4 &\rightleftharpoons 10X_7 \\
5X_7 &\rightleftharpoons 2X_9 + 2X_{10}
\end{aligned}
$$

M3. We now delete the duplicate columns of $Z$ and keep only one of them in order to find the complex composition matrix $\widetilde{Z}$ of the equivalent network. Suppose that $\widetilde{Z} \in \mathbb{R}^{\widetilde{m} \times \widetilde{c}}$, where $\widetilde{m}$ and $\widetilde{c}$ are the number of species and complexes in the equivalent model, respectively. It is clear that, if $n(I) = n_1$ and the number of duplicate columns is $n_2$, then $\widetilde{m} = m - n_1$ and $\widetilde{c} = c - n_2 + 1$.

Let the $p$th and $q$th complexes be a pair of identical complexes. We first replace the $p$throw of $B$ with

$$
\widetilde{B}_{pj} =
\begin{cases}
B_{qj}, & \text{if } B_{pj} = 0 \text{ and } B_{qj} \neq 0, \\
B_{pj}, & \text{otherwise,}
\end{cases}
$$

and delete its $q$th row. We then repeat the same technique for all the pairs of identical complexes to obtain the incidence matrix $\widetilde{B}$ of the equivalent network. It is clear that $\widetilde{B} \in \mathbb{R}^{\widetilde{c} \times r}$.

The complex composition matrix $\widetilde{Z} \in \mathbb{R}^{6 \times 4}$ and the incidence matrix $\widetilde{B} \in \mathbb{R}^{4 \times 6}$ of the equivalent network corresponding to the network (12) are

$$\widetilde{Z} = \begin{bmatrix} 15 & 0 & 0 & 0 \\ 15 & 0 & 0 & 0 \\ 0 & 30 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 2 \end{bmatrix}, \qquad \widetilde{B} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

Thus, the equivalent network corresponding to the network (12) becomes

$$15X_1 + 15X_2 \rightleftharpoons 30X_4 \rightleftharpoons 10X_7 \rightleftharpoons 2X_9 + 2X_{10}. \tag{15}$$
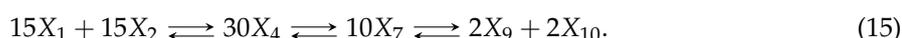
Define $\widetilde{K} := \text{diag}(\widetilde{k})$, where $\widetilde{k}$ is the vector of rate constants of the equivalent network. Then, the Laplacian matrix $\widetilde{L}(\widetilde{x}) \in \mathbb{R}^{\widetilde{c} \times \widetilde{c}}$ of the equivalent network is $\widetilde{L}(\widetilde{x}) = \widetilde{B}\widetilde{K}\widetilde{D}(\widetilde{x})\widetilde{\Delta}^{\top}$, where $\widetilde{D}(\widetilde{x}) = \text{diag}(\widetilde{d})$ and $\widetilde{\Delta}$ is the outgoing matrix of the equivalent network. The mathematical model of the equivalent network is

$$\begin{cases} \dfrac{d\widetilde{x}}{dt} = -\widetilde{Z}\widetilde{L}(\widetilde{x})\widetilde{C}(\widetilde{x}), \\ \widetilde{x}(0) = \widetilde{x}_0, \end{cases} \tag{16}$$

where $\widetilde{C}(\widetilde{x}) = \text{Exp}(\widetilde{Z}^{\top}\text{Ln}(\widetilde{x}))$.

Step 5: Independent subnetworks.

In some cases, the equivalent network may consist of independent subnetworks, which contain sets of reactions that do not share any common species. In this situation, we determine all the independent subnetworks of the equivalent network. First, we determine the linkage classes of the equivalent network using its incidence matrix $\widetilde{B}$. The index sets corresponding to the complexes of the equivalent network are then used to determine the indices of the species participating in each linkage class. More precisely, if $\mathcal{C}_{j_1}, \mathcal{C}_{j_2}, \cdots, \mathcal{C}_{j_p}$ are the complexes of the same linkage class $\mathcal{L}$, then the set of indices of $\mathcal{L}$ is $I_{\mathcal{L}} = \bigcup_{i=1}^{p} I_{\mathcal{C}_{j_i}}$. It is clear that two linkage classes $\mathcal{L}_1$ and $\mathcal{L}_2$ are dependent, if $I_{\mathcal{L}_1} \cap I_{\mathcal{L}_2} \neq \emptyset$. Mutually dependent linkage classes form an independent subnetwork of the equivalent network.

Let $\mathcal{N}_j, j = 1, \cdots, l$, be the independent subnetworks of a network having more than one reaction. As they are independent, we can consider each of them as a separate network. Thus, we consider the reduction of each of them separately. We note that in the case of the network (12), there is only one subnetwork in the corresponding equivalent network.

Step 6: Selecting complexes for deletion.

In this step, we apply the reduction procedure of [5] to meaningfully delete certain complexes from the equivalent network. The first step of this procedure is to determine the candidate (intermediate) complexes for deletion. We note that the intermediate complexes are the ones that participate in more than one reaction, with each reversible reaction considered as a single reaction. It is reasonable to consider all the other complexes as important complexes. We consider the constituent species of the important complexes as the important species. In [5], the candidate complexes for deletion have been chosen manually. Here, we show how to determine such complexes in an automatic way. Let $\widetilde{B}_{\text{out}}$ be the incidence matrix of the equivalent network, with each reversible reaction considered as a single reaction. $\widetilde{B}_{\text{out}}$ can be found by deleting the columns of $\widetilde{B}$ corresponding to the reverse reactions of reversible reactions. The intermediate complexes can then be determined from $\widetilde{B}_{\text{out}}$ by detecting the rows that contain more than one nonzero element.

In the case of the network (15), we have

$$
\widetilde{B}_{\text{out}} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.
$$

The candidate complexes for deletion from the linkage class (15) are $30X_4$ and $10X_7$. The important species are $X_1$, $X_2$, $X_9$, and $X_{10}$, which are the constituent species of the important complexes $15X_1 + 15X_2$ and $2X_9 + 2X_{10}$.

Let $J_i$ be the set of indices of the important species of $\mathcal{N}_i$. As explained in [5], we rank the candidate (intermediate) complexes according to the error integral $E_i(T_i)$ defined in (9). Here, $T_i = \max\limits_{j \in J_i} \tau_j$ is the settling time of $\mathcal{N}_i$. We delete from $\mathcal{N}_i$ the optimal combination of complexes, i.e., the combination with the maximum value of $E_i(T_i)$ that does not exceed the predefined threshold.

Step 7: Mathematical model of the reduced network.

Let $M_{\text{del}}^i$ be the set of indices of complexes deleted from $\mathcal{N}_i$, $i = 1, \cdots, l$. Define $M_{\text{del}} = \bigcup_{i=1}^{l} M_{\text{del}}^i$. In the final step, we show how to determine the mathematical model of the reduced network after deleting the complexes with indices $M_{\text{del}}$ from the equivalent network. In order to obtain the complex composition matrix $\widehat{Z}$ of the reduced network, we delete the columns of $\widetilde{Z}$ corresponding to $M_{\text{del}}$. Define $M_{\text{red}} = \{1, \cdots, c\} \setminus M_{\text{del}}$, which is the set of indices corresponding to the complexes remaining in the reduced network. According to Proposition 1 in [5], the Laplacian matrix $\widehat{L}(\widetilde{x})$ of the reduced network is defined as the Schur complement of $\widetilde{L}(\widetilde{x})$ corresponding to $M_{\text{red}}$. The mathematical model of the reduced network is then given by

$$
\begin{cases} \dfrac{d\widetilde{x}}{dt} = -\widehat{Z}\widehat{L}(\widetilde{x})\widehat{C}(\widetilde{x}), \\ \widetilde{x}(0) = \widetilde{x}_0, \end{cases} \tag{17}
$$

where $\widehat{C}(\widetilde{x}) = \text{Exp}\left(\widehat{Z}^\top \text{Ln}(\widetilde{x})\right)$.

## 5. Application to Real-Life Reaction Networks

In this section, we apply our automatic model reduction method to reduce two computational models of biological processes that consist of several linkage classes, with common species shared between at least two of them. These models have been retrieved from the BioModels database [10], which is a repository of computational models of biological processes available online at https://www.ebi.ac.uk/biomodels/. For both models considered throughout this section, the threshold value of the error integral for stopping the iterative reduction procedure has been set to 0.15. In general, this threshold can be chosen depending on the desired closeness of the reduced model to the original model. Detailed explanations of the reduction procedure of these models are provided in Appendix A.

### 5.1. Neural Stem Cell Regulation

We consider a mathematical model of neural stem cell regulation (NSCR). The complex graph corresponding to the model is given in the left-hand panel of Figure 1. For the corresponding detailed mathematical model we refer to [23]. The concentrations of the important species participating in the original model are represented in Figure 2.

We note that there is one linkage class with more than one reversible reaction. The remaining linkage classes of the network contain only one reversible reaction. In this case, the method proposed in [5] can be meaningfully applied to delete the intermediate complexes $X_7$, $X_8$, and $X_{13}$. Further reduction of the network using the same method would eliminate reactions from the network.

To overcome this shortcoming and to obtain a meaningful model reduction of the network, we apply our new reduction technique.

Step 3 of our reduction procedure selects the species $X_5$, $X_6$, and $X_{20}$ to be eliminated from the mathematical model, which allows us to join three linkage classes into a single one. Detailed explanation of the elimination procedure can be found in Appendix A.1. In particular, we use three conservation laws to complete the aforementioned elimination of species. However, as a downside of our elimination procedure, the obtained reduced model is valid only for trajectories for which the three conserved quantities have fixed values as in (A2). For example, the first of these three conserved quantities is the total concentration of the pool of Notch [24] and Notch transmembrane [25], and it is reasonable to assume that in an experimental set-up this pool has a fixed concentration. The complex graph corresponding to the equivalent network of NSCR is shown in the middle panel of Figure 1.
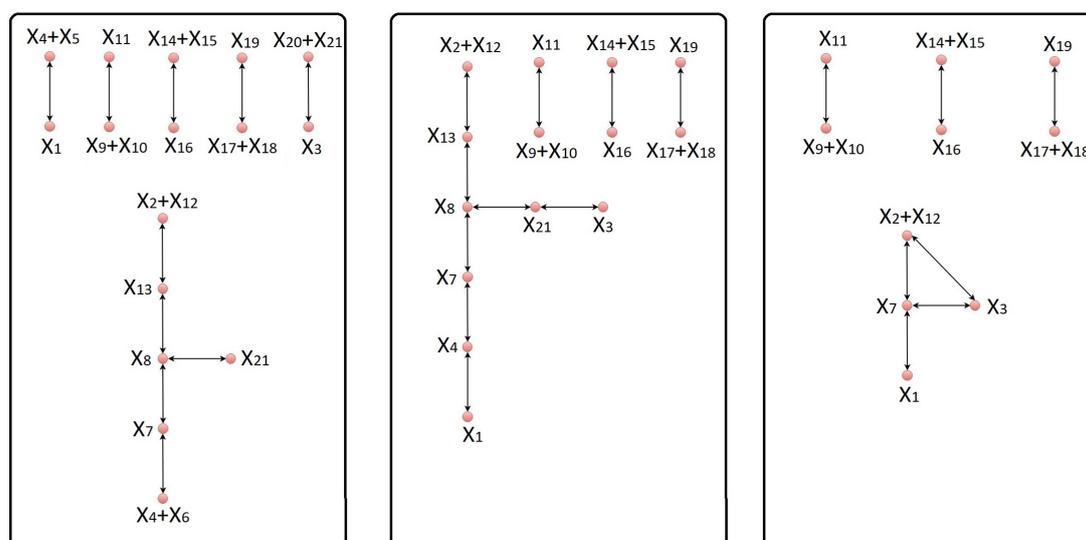


**Figure 1.** The left-hand panel is the complex graph corresponding to the original model of neural stem cell regulation. The complex graph corresponding to the equivalent model obtained after eliminating the species $X_5$, $X_6$, and $X_{20}$ from the original model is given in the middle panel. Deletion of the intermediate complexes $X_4$, $X_8$, $X_{13}$, and $X_{21}$ leads to a reduced network with the corresponding complex graph represented in the right-hand panel. The difference between the original model and the reduced model, as measured by the error integral, is 4.85%.

Finally, applying the procedure described in Step 6 to this equivalent network, we delete the optimal combination of the intermediate complexes, which consists of $X_4$, $X_8$, $X_{13}$, and $X_{21}$. The complex graph corresponding to the reduced network is given in the right-hand panel of Figure 1. Even though 33.33% of the species have been deleted from the original model, the error integral is only 4.85%. The concentrations of the important species participating in the reduced model are represented in Figure 2.

### 5.2. Hedgehog Signaling Pathway

Subsequently, we have used our new technique of model reduction to a very different type of biochemical reaction network, namely, a model of hedgehog signaling pathway (HSP). The corresponding detailed mathematical model can be found in [23].
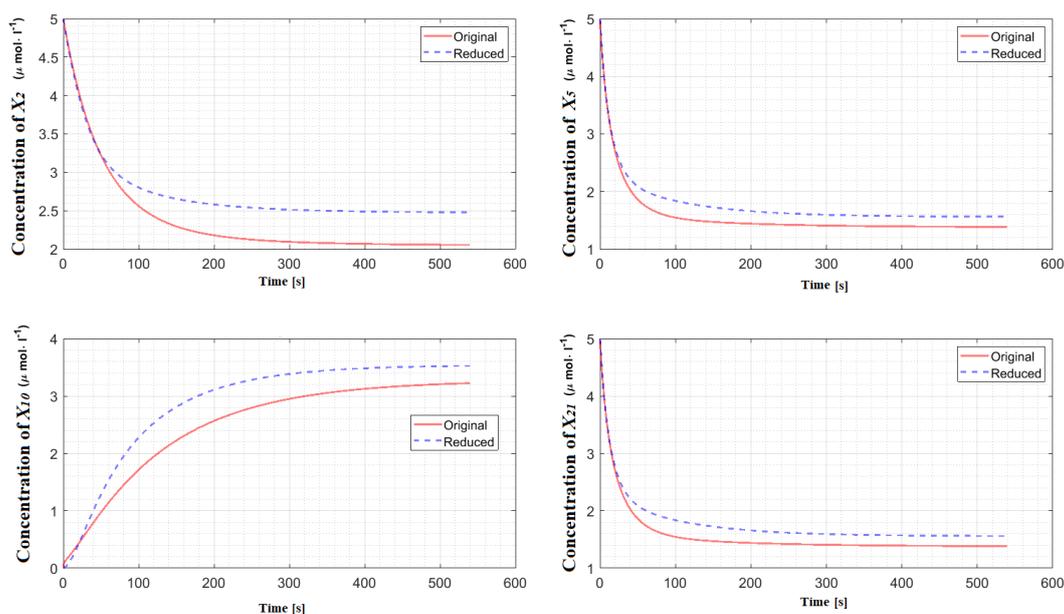
**Figure 2.** Concentrations of the important species of neural stem cell regulation in the original model and in the reduced model. The difference between these models, as measured by the error integral, is 4.85%.

There are nine reversible and two irreversible reactions in the original network. The complex graph corresponding to the original network is given in the left-hand panel of Figure 3. The concentrations of the important species participating in the original model are represented in Figure 4. We observe that it contains two linkage classes consisting of more than one reaction. The remaining six linkage classes consist of only one reversible reaction. In this case, the reduction method proposed in [5] can be meaningfully applied to delete four intermediate complexes, namely, $X_4$, $X_{15}$, $X_{18}$, and $X_{19}$. Further reduction using the same procedure would lead to the deletion of reactions, which is not desirable in the sense of preserving the original behavior. However, the use of our new approach results in a meaningful reduction without causing a significant change in the original behavior. Detailed explanation of the species elimination procedure is included in Appendix A.2.

Step 3 of our reduction procedure chooses the species $X_1$, $X_{11}$, and $X_{12}$ to be eliminated from the network, i.e., $I = \{1, 11, 12\}$. Similar to the model of NSCR, we use three conservation laws to eliminate these species. Again, as a downside of our reduction procedure, our final reduced model of HSP is valid only for trajectories for which the three conserved quantities have fixed values given by (A6). For example, the first of these conserved quantities is the total concentration of the pool of ADP (adenosine diphosphate) and ATP (adenosine triphosphate), and it is reasonable to assume that in an experimental set-up this pool has a fixed concentration. The complex graph corresponding to the equivalent network is shown in the middle panel of Figure 3.

Finally, the procedure described in Step 6 results in $X_4$, $X_8$, $X_{13}$, and $X_{18}$ to be the optimal combination of the intermediate complexes for deletion, i.e., the maximal combination of complexes with an error integral of at most 15%. Deletion of these complexes by the procedure of [5] leads to the reduced network with the corresponding complex graph given in right-hand panel of Figure 3. We remark that, even though 33.33% of the species have been eliminated from the model, the resulting error integral is only 6.59%. The concentrations of the important species participating in the reduced model are represented in Figure 4.
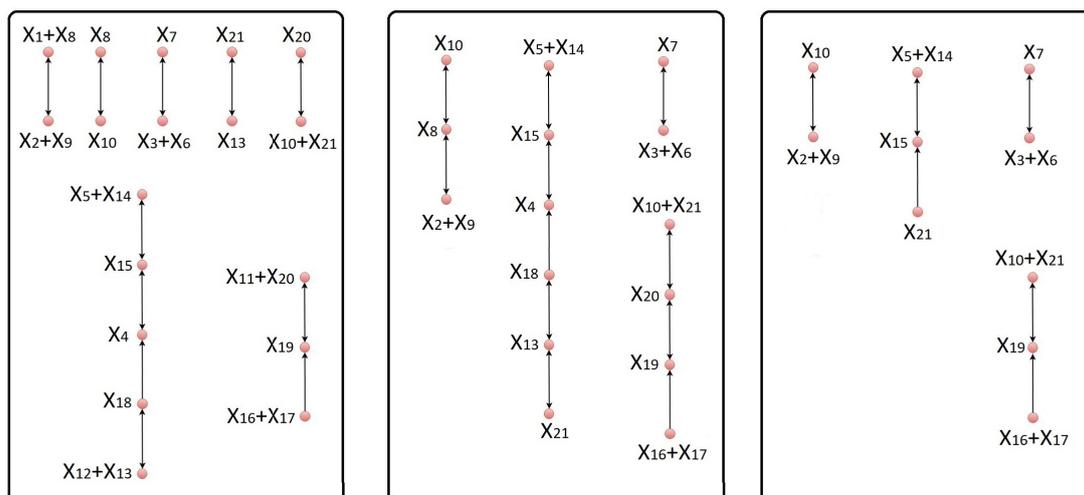
**Figure 3.** The left-hand panel is the complex graph of the original model of hedgehog signaling pathway. The complex graph corresponding to the equivalent model obtained after eliminating the species $X_1$, $X_{11}$, and $X_{12}$ from the original model is given in the middle panel. Deletion of the intermediate complexes $X_4$, $X_8$, $X_{13}$, and $X_{18}$ leads to a reduced model with the corresponding complex graph shown in the right-hand panel. The difference between the original model and the reduced model, as measured by the error integral, is 6.59%.
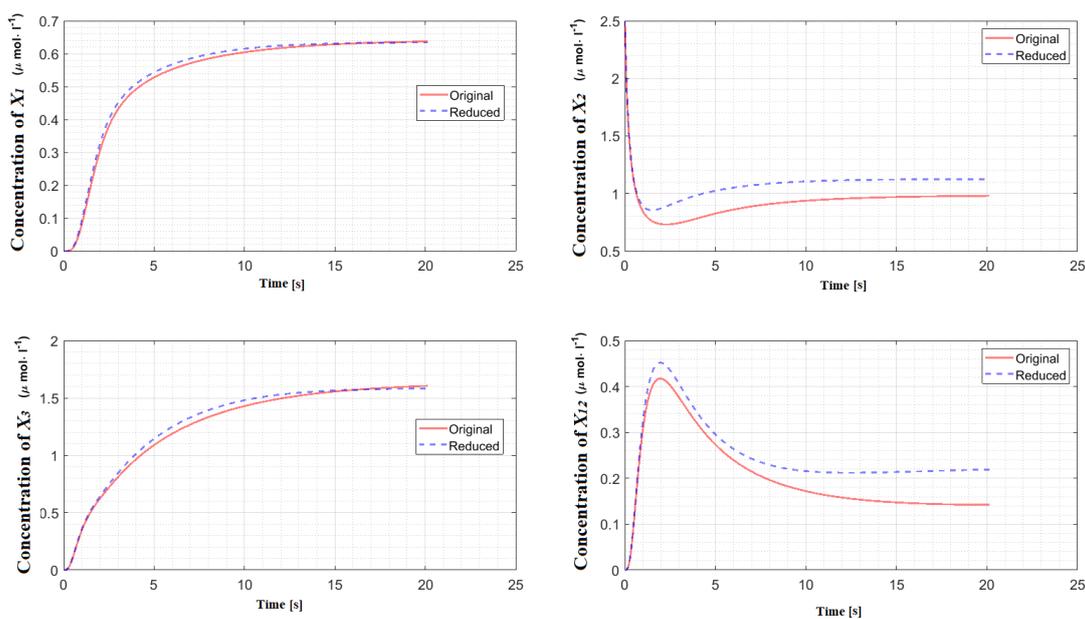


**Figure 4.** Concentrations of the important species of hedgehog signaling pathway in the original model and in the reduced model. The difference between these models, as measured by the error integral, is 6.59%.

## 6. Conclusions and Discussion

In this paper, we have illustrated a new approach to the model reduction of CRNs, which extends the method proposed in [5]. Our method is based on eliminating concentrations of certain species from the mathematical model corresponding to the network using the conservation laws obtained from the model. This elimination allows us to rewrite the complex graph of the network in a preferred form, for which the model reduction method proposed in [5] becomes meaningfully applicable. Even though

the procedure results in reducing the order of the corresponding mathematical model, new parameters depending on the vector of initial concentrations are added to the model.

We have implemented an automated process for our model reduction method, which is divided into seven steps. We created a MATLAB file corresponding to each step and all files are included in a single MATLAB library. The information of a given CRN is provided by the user and subsequently used to reduce the corresponding mathematical model in a fully automated manner. The error integral defined in [5] is used to measure the difference between the behaviors of the original model and the reduced model. An important tool in the definition of the error integral is the settling time of the network, which was computed manually in [5]. We have shown how to automatically determine the settling time of a given CRN using its steady state. We have also been able to determine the intermediate complexes in a fully automated manner using the incidence matrix of the network. The MATLAB library consisting of all the files corresponding to our step-wise reduction procedure is provided as Supplementary Material.

We have successfully applied our model reduction method on two real-life examples: neural stem cell regulation and the hedgehog signaling pathway. For each case, we have used our techniques to enable the reduction method proposed in [5] to be applied to a significant extent. We have been able to automatically join several linkage classes of the network by eliminating concentrations of certain species from the corresponding mathematical model. Table 1 provides the percentage of deleted species and the value of the corresponding error integral for each reaction network. The detailed explanation of the reduction procedure corresponding to these models can be found in the Supplementary Material.

**Table 1.** The amount of deleted species and the value of the error integral (in%) corresponding to each example after its model reduction using our method.

| Model | Deleted Species (%) | Error Integral (%) |
|-------|---------------------|--------------------|
| NSCR  | 33.33               | 4.85               |
| HSP   | 33.33               | 6.59               |

The major tools in our automated reduction method are the linearly independent conservation laws, which are used to eliminate the concentrations of species with index set $I$ from the network. The identification of the set $I$ is straightforward. However, in some cases, not all the selected species can be eliminated from the network because of the lack of suitable conservation laws, i.e, in the case when the number of the linearly independent conservation laws is less than the number of elements in $I$. In such cases, we select $I$ in a way that the number of its elements is equal to the number of linearly independent conservation laws of the network. This adjustment ensures that the species with index set $I$ can be eliminated from the network. Note that, even though this elimination rewrites the complex graph of the network in a preferred form (i.e., increases the number of intermediate complexes in it), new parameters (conserved quantities) are added in the corresponding mathematical model.

Although the procedure of elimination of species using conservation laws described in this paper improves the extent of applicability of the model reduction method in [5], it is no longer valid for all trajectories of the original model. As mentioned earlier, the reduced model is valid only for trajectories with fixed values of these conserved quantities. This aspect is similar to the case of the Michaelis–Menten approximation of enzyme kinetic networks which depends on a conserved quantity, namely, the total enzyme concentration.

One of the main advantages of the model reduction procedure proposed in [5] is that it preserves the governing laws of the CRN. For instance, if a CRN is governed by Michaelis–Menten kinetics, then the reduced CRN by the approach in [5] will again be governed by Michaelis–Menten kinetics. However, this is not the case for our automatic reduction method due to the rewriting procedure, as we have explained in [8].

In [5], the behaviors of the species' concentrations participating in the reduced model are compared with the ones participating in the original model using the error integral (9) defined over a manually set timescale. In this paper, the same procedure has been adopted with the significant modification that the time length of the trajectory is automatically determined for models with unique steady state. We only consider CRNs that are asymptotically stable around a unique steady state. The importance of asymptotically stability arises in determining the settling time automatically. Note that for CRNs that are not asymptotically stable, the settling time cannot be automatically determined. For instance, in the case of oscillatory networks, it is not clear what time frame should be considered for comparing the dynamics therefore its automatic determination is not straightforward. However, this does not mean that the application of our reduction procedure for such CRNs is not possible. In the case of CRNs that are not asymptotically stable, the determination of the time-scale for comparison of the dynamics of the reduced and the original CRNs can still be done manually based on the user preference.

The error integral, as defined in [5], strictly depends on a particular trajectory. Considering an ensemble of trajectories would seem a better choice for more robust reduction. However, in this case it is not clear which subset of trajectories to choose from an infinite number of trajectories for comparing the original model and the reduced models. Moreover, such a reduction requires a higher computational effort, which is not desirable as we aim at using a model reduction method that is computationally less intensive.

The main principle of our model reduction method is similar to the one of QSSA, which is currently the state-of-the-art reduction technique for networks consisting of enzyme-catalyzed reactions. Our method, as well as QSSA applied for deriving the Michaelis–Menten kinetics, make use of combinations of conservation laws and the principle of complex balancing to reduce the number of complexes from the corresponding complex graph. We use the conservation laws to join certain linkage classes of a given network, which allows us to apply the procedure of [5] for complementary model reduction. On the other hand, when applied for deriving the Michaelis–Menten kinetics, QSSA uses the total enzyme (which is distributed between the enzyme and the intermediate enzyme complexes) in terms of a conservation law to eliminate the concentrations of enzyme species from the corresponding mathematical model.

Our approach to the model reduction is essentially QSSA applied to complexes rather than species. It would seem as though the settling time of complexes would be another choice (rather than the error integrals) for selecting the complexes to be deleted from the network. This could provide a more robust ranking of candidate complexes to be deleted as such a ranking might be independent of the chosen trajectory. However, even though the settling time can be used for ranking the candidate complexes according to their speed of reaching steady state, it is not clear which subset of candidate complexes should be considered ultimately for deletion because we still need a measure for quantifying the difference between the reduced model and the original model.

**Abbreviations**

The following abbreviations are used in this manuscript.

CRN  biochemical reaction networks
HSP  hedgehog signaling pathway
NSCR  neural stem cell regulation
ODE  ordinary differential equation
QSSA  quasi-steady-state assumption
ADP  adenosine diphosphate
ATP  adenosine triphosphate

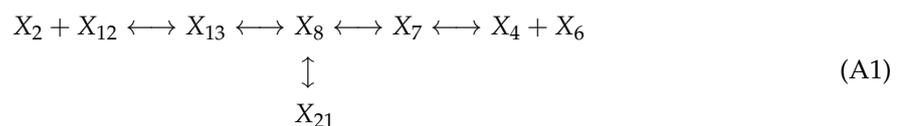**Appendix A. Detailed Explanation of the Reduction Procedure**

We provide a detailed explanation of our automatic reduction procedure, which has been applied to reduce the mathematical models of NSCR and HSP retrieved from the BioModels database [10]. The complete mathematical models can be found in [23].

*Appendix A.1. Reduction of Neural Stem Cell Regulation*

We consider a mathematical model of NSCR [23]. There are 21 species participating in the ten reversible reactions of the network. The list of these reactions is given below.

$$\mathcal{R}_1 : X_1 \longleftrightarrow X_4 + X_5 \qquad\qquad \mathcal{R}_6 : X_{13} \longleftrightarrow X_2 + X_{12}$$
$$\mathcal{R}_2 : X_4 + X_6 \longleftrightarrow X_7 \qquad\qquad \mathcal{R}_7 : X_3 \longleftrightarrow X_{20} + X_{21}$$
$$\mathcal{R}_3 : X_7 \longleftrightarrow X_8 \qquad\qquad\qquad \mathcal{R}_8 : X_{17} + X_{18} \longleftrightarrow X_{19}$$
$$\mathcal{R}_4 : X_8 \longleftrightarrow X_{13} \qquad\qquad\qquad \mathcal{R}_9 : X_{14} + X_{15} \longleftrightarrow X_{16}$$
$$\mathcal{R}_5 : X_8 \longleftrightarrow X_{21} \qquad\qquad\qquad \mathcal{R}_{10} : X_9 + X_{10} \longleftrightarrow X_{11}$$

The reaction rates with corresponding rate constants can be found in [23]. We note that there is one linkage class with more than one reaction. This linkage class consists of the five reversible reactions $\mathcal{R}_j, j = 2, \cdots, 6$, and contains three intermediate single-species complexes, $X_7$, $X_8$, and $X_{13}$.

$$X_2 + X_{12} \longleftrightarrow X_{13} \longleftrightarrow X_8 \longleftrightarrow X_7 \longleftrightarrow X_4 + X_6$$
$$\updownarrow \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A1)}$$
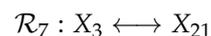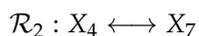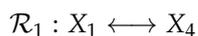$$X_{21}$$

As there are three intermediate complexes, the model reduction method proposed in [5] can be meaningfully applied to delete these species from the linkage class (A1). However, further reduction of the model using the same method is not meaningful as explained in the main paper. We demonstrate how to use our new approach of model reduction, which enables a complementary reduction in a meaningful way.
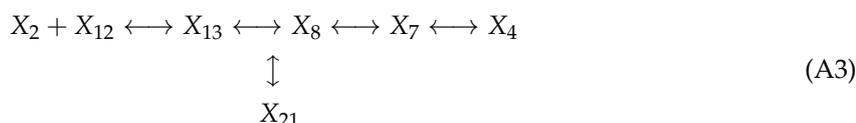
Step 3 of our reduction procedure chooses the species $X_5$, $X_6$, and $X_{20}$ to be eliminated from the mathematical model using the following conservation laws.

$$
\begin{aligned}
x_1 + x_5 &= 5, \\
x_1 - x_4 - x_6 &= 0, \\
x_3 + x_{20} &= 5.
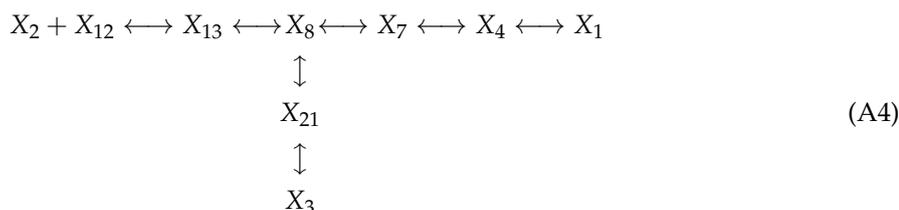\end{aligned}
\qquad\qquad \text{(A2)}
$$

Note that the conserved quantities, i.e., the constants given in the right hand side of the Equation (A2), strictly depend on the vector of species' initial concentrations, meaning that for different values of initial concentrations' we will have different values for these constants. After the aforementioned elimination we can rewrite the reactions $\mathcal{R}_1$, $\mathcal{R}_2$, and $\mathcal{R}_7$ in the following form.

$$\mathcal{R}_1 : X_1 \longleftrightarrow X_4 \qquad \mathcal{R}_2 : X_4 \longleftrightarrow X_7 \qquad \mathcal{R}_7 : X_3 \longleftrightarrow X_{21}$$

Similarly, we can rewrite the linkage class (A1) as

$$
\begin{array}{c}
X_2 + X_{12} \longleftrightarrow X_{13} \longleftrightarrow X_8 \longleftrightarrow X_7 \longleftrightarrow X_4 \\
\updownarrow \\
X_{21}
\end{array}
\tag{A3}
$$

We therefore join $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_7$, and the linkage class (A3) into a single linkage class:

$$
\begin{array}{c}
X_2 + X_{12} \longleftrightarrow X_{13} \longleftrightarrow X_8 \longleftrightarrow X_7 \longleftrightarrow X_4 \longleftrightarrow X_1 \\
\updownarrow \\
X_{21} \\
\updownarrow \\
X_3
\end{array}
\tag{A4}
$$

The equivalent network consists of four independent subnetworks, namely, $\mathcal{R}_8$, $\mathcal{R}_9$, $\mathcal{R}_{10}$, and the linkage class (A4). We apply the reduction procedure of [5] to the linkage class (A4) to eliminate the optimal combination of its intermediate complexes. We remark that the first three independent subnetworks remain unchanged after reducing the linkage class (A4). There are five intermediate complexes, namely, $X_4$, $X_7$, $X_8$, $X_{13}$, and $X_{21}$, which are candidate complexes for deletion using the method proposed in [5]. The iterative test described in Step 6 in the main paper selects $\{X_4, X_8, X_{13}, X_{21}\}$, to be the optimal combination of intermediate complexes for deletion. Therefore, the principle of complex balancing is used to delete these complexes from the linkage class (A4). Even though seven species (33.33%) have been deleted from the original model, the error integral is only 4.85%. From the structure of the Laplacian matrix $\hat{L}(\widetilde{x})$ (see Equation (15) of the main paper) we derive the complex graph of the reduced linkage class:
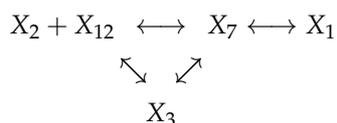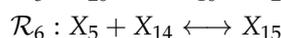
$$
\begin{array}{c}
X_2 + X_{12} \longleftrightarrow X_7 \longleftrightarrow X_1 \\
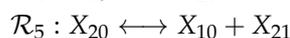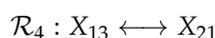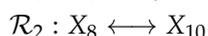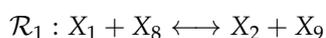\searrow \quad \nearrow \\
X_3
\end{array}
$$

Table A1 provides a quantitative comparison of the original model of NSCR and the corresponding reduced model. Comparison of concentrations of important species participating in the original model and the reduced model is illustrated in Figure 4 of the main manuscript.
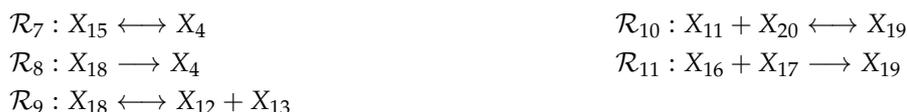
**Table A1.** Quantitative comparison of the original model and the reduced model of neural stem cell regulation.

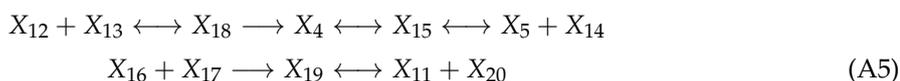|                | Species | Reactions |
| -------------- | ------- | --------- |
| Original Model | 21      | 10        |
| Reduced Model  | 14      | 7         |

*Appendix A.2. Reduction of Hedgehog Signaling Pathway*

The second example considered for our model reduction procedure is a model of HSP. The model contains nine reversible and two irreversible reactions. The list of these reactions is given below.

$$\mathcal{R}_1 : X_1 + X_8 \longleftrightarrow X_2 + X_9 \qquad\qquad \mathcal{R}_4 : X_{13} \longleftrightarrow X_{21}$$
$$\mathcal{R}_2 : X_8 \longleftrightarrow X_{10} \qquad\qquad\qquad\quad\; \mathcal{R}_5 : X_{20} \longleftrightarrow X_{10} + X_{21}$$
$$\mathcal{R}_3 : X_7 \longleftrightarrow X_3 + X_6 \qquad\qquad\quad \mathcal{R}_6 : X_5 + X_{14} \longleftrightarrow X_{15}$$

$$\mathcal{R}_7 : X_{15} \longleftrightarrow X_4$$
$$\mathcal{R}_8 : X_{18} \longrightarrow X_4$$
$$\mathcal{R}_9 : X_{18} \longleftrightarrow X_{12} + X_{13}$$

$$\mathcal{R}_{10} : X_{11} + X_{20} \longleftrightarrow X_{19}$$
$$\mathcal{R}_{11} : X_{16} + X_{17} \longrightarrow X_{19}$$

The reaction rates with the corresponding rate constants can be found in [23]. We observe that it contains two linkage classes having more than one reaction. These linkage classes consist of reactions $\mathcal{R}_i, j = 6, \cdots, 9$, and $\mathcal{R}_j, j = 10, 11$, respectively, as shown below.

$$X_{12} + X_{13} \longleftrightarrow X_{18} \longrightarrow X_4 \longleftrightarrow X_{15} \longleftrightarrow X_5 + X_{14}$$
$$X_{16} + X_{17} \longrightarrow X_{19} \longleftrightarrow X_{11} + X_{20} \tag{A5}$$
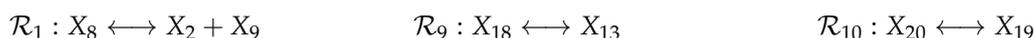
The remaining five linkage classes consist of only one reversible or irreversible reaction. Clearly, in (A5) there are four intermediate complexes, namely, $X_4, X_{15}, X_{18}$, and $X_{19}$. In this case, the reduction method proposed in [5] can be meaningfully applied to delete the optimal combination of these complexes. Further reduction using the same procedure would lead to the deletion of reactions, which is not desirable in the sense of preserving the original behavior. However, we show how to use our new reduction technique for a meaningful complementary reduction without causing a significant change in the original behavior.
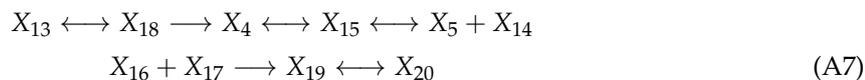
Step 3 of our reduction procedure selects $X_1, X_{11}$, and $X_{12}$ to be eliminated from the network using the following conservation laws.

$$x_1 + x_2 = 2,$$
$$x_{11} + x_{16} + x_{19} = 3, \tag{A6}$$
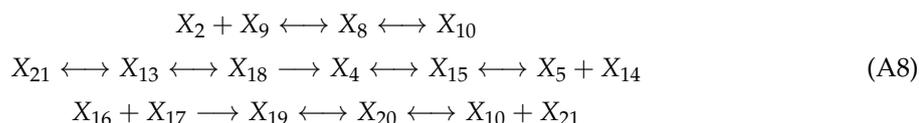$$x_{12} + x_4 + x_5 + x_{15} + x_{18} = 4.5.$$

Note that, similar to the case of the NSCR model, the constants given in the right hand side of the Equation (A6) strictly depend on the vector of species' initial concentrations, meaning that for different values of initial concentrations, we will have different values for these constants. After the aforementioned elimination, we can rewrite the reactions $\mathcal{R}_1, \mathcal{R}_9$, and $\mathcal{R}_{10}$ in the following form.

$$\mathcal{R}_1 : X_8 \longleftrightarrow X_2 + X_9 \qquad \mathcal{R}_9 : X_{18} \longleftrightarrow X_{13} \qquad \mathcal{R}_{10} : X_{20} \longleftrightarrow X_{19}$$

Similarly, we rewrite the linkage classes (A5) in the following form.

$$X_{13} \longleftrightarrow X_{18} \longrightarrow X_4 \longleftrightarrow X_{15} \longleftrightarrow X_5 + X_{14}$$
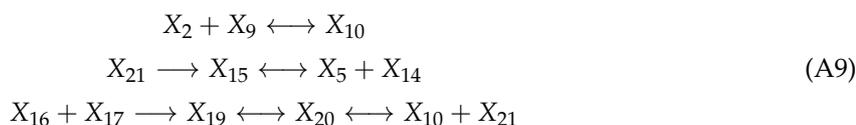$$X_{16} + X_{17} \longrightarrow X_{19} \longleftrightarrow X_{20} \tag{A7}$$

We therefore join the reactions $\mathcal{R}_1$ and $\mathcal{R}_2$; reaction $\mathcal{R}_4$ and the first linkage class of (A7); and reaction $\mathcal{R}_5$ and the second linkage class of (A7). As a result, we obtain three linkage classes with more than one reaction:

$$X_2 + X_9 \longleftrightarrow X_8 \longleftrightarrow X_{10}$$
$$X_{21} \longleftrightarrow X_{13} \longleftrightarrow X_{18} \longrightarrow X_4 \longleftrightarrow X_{15} \longleftrightarrow X_5 + X_{14} \tag{A8}$$
$$X_{16} + X_{17} \longrightarrow X_{19} \longleftrightarrow X_{20} \longleftrightarrow X_{10} + X_{21}$$

The equivalent network contains six intermediate complexes $X_4, X_8, X_{15}, X_{18}, X_{19}$, and $X_{20}$, which form the set of candidate complexes for deletion from (A8) by the approach in [5]. Finally, the procedure described in Step 6 in the main paper selects $\{X_4, X_{18}, X_8, X_{13}\}$ to be the optimal

combination of the intermediate complexes for deletion, i.e., the maximal combination of complexes with an error integral of at most 15%. The reduced linkage classes are

$$X_2 + X_9 \longleftrightarrow X_{10}$$
$$X_{21} \longrightarrow X_{15} \longleftrightarrow X_5 + X_{14} \qquad \text{(A9)}$$
$$X_{16} + X_{17} \longrightarrow X_{19} \longleftrightarrow X_{20} \longleftrightarrow X_{10} + X_{21}$$

Even though 33.33% of the species have been eliminated from the network, the resulting error integral is only 6.59%. Table A2 provides a quantitative comparison of the original model of HSP and the corresponding reduced model. A comparison of concentrations of important species participating in the original model and the reduced model is illustrated in Figure 4 of the main manuscript.

**Table A2.** Quantitative comparison of the original model and the reduced model of hedgehog signaling pathway.

|                | Species | Reactions |
| -------------- | ------- | --------- |
| Original Model | 21      | 11        |
| Reduced Model  | 14      | 7         |

## References

1. Radulescu, O.; Gorban, A.N.; Zinovyev, A.; Noel, V. Reduction of dynamical biochemical reactions networks in computational biology. *Front. Genet.* **2012**, *3*, 131. [CrossRef] [PubMed]
2. Snowden, T.J.; van der Graaf, P.H.; Tindall, M.J. Methods of model reduction for large-scale biological systems: a survey of current methods and trends. *Bull. Math. Biol.* **2017**, *79*, 1449–1486. [CrossRef]
3. Cornish-Bowden, A. *Fundamentals of Enzyme Kinetics*; Wiley-Blackwell: Weinheim, Germany, 2012.
4. Segel, I. *Biochemical Calculations*; Wiley: Hoboken, NJ, USA, 1975.
5. Rao, S.; Van der Schaft, A.; Van Eunen, K.; Bakker, B.M.; Jayawardhana, B. A model reduction method for biochemical reaction networks. *BMC Syst. Biol.* **2014**, *8*, 52. [CrossRef] [PubMed]
6. Kron, G. *Tensor Analysis of Networks*; Macdonald: New York, NY, USA, 1939.
7. Zhang, F. *The Schur Complement and Its Applications*; Springer Science & Business Media: Berlin, Germany, 2006; Volume 4.
8. Gasparyan, M.; Van Messem, A.; Rao, S. A Novel Technique for Model Reduction of Biochemical Reaction Networks. *IFAC-PapersOnLine* **2018**, *51*, 28–31. [CrossRef]
9. Noether, E. Invariant variation problems. *Transp. Theory Stat. Phys.* **1971**, *1*, 186–207. [CrossRef]
10. Le Novere, N.; Bornstein, B.; Broicher, A.; Courtot, M.; Donizelli, M.; Dharuri, H.; Li, L.; Sauro, H.; Schilstra, M.; Shapiro, B.; et al. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* **2006**, *34*, D689–D691. [CrossRef] [PubMed]
11. Van der Schaft, A.; Rao, S.; Jayawardhana, B. A network dynamics approach to chemical reaction networks. *Int. J. Control* **2016**, *89*, 731–745. [CrossRef]
12. Feinberg, M. Lectures on chemical reaction networks. In *Notes of Lectures Given at the Mathematics Research Center*; University of Wisconsin-Madison: Madison, WI, USA, 1979; p. 49.
13. Michaelis, L.; Menten, M.L. *Die kinetik der Invertinwirkung*; Universitätsbibliothek Johann Christian Senckenberg: Frankfurt, Germany, 2007.
14. Monod, J.; Wyman, J.; Changeux, J.P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **1965**, *12*, 88–118. [CrossRef]
15. Hill, A.V. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol.* **1910**, *40*, 4–7.
16. Hofmeyr, J.H.; Cornish-Bowden, H. The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models. *Bioinformatics* **1997**, *13*, 377–385. [CrossRef] [PubMed]
17. Cleland, W.W. Derivation of rate equations for multisite ping-pong mechanisms with ping-pong reactions at one or more sites. *J. Biol. Chem.* **1973**, *248*, 8353–8355. [PubMed]

18. Koshland, D.; Némethy, G.; Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **1966**, *5*, 365–385. [CrossRef] [PubMed]

19. Bapat, R.B. *Graphs and Matrices*; Springer: New York, NY, USA, 2010; Volume 27.

20. Feinberg, M. Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency zero and deficiency one theorems. *Chem. Eng. Sci.* **1987**, *42*, 2229–2268. [CrossRef]

21. Horn, F.; Jackson, R. General mass action kinetics. *Arch. Ration. Mech. Anal.* **1972**, *47*, 81–116. [CrossRef]

22. Rao, S.; van der Schaft, A.; Jayawardhana, B. A graph-theoretical approach for the analysis and model reduction of complex-balanced chemical reaction networks. *J. Math. Chem.* **2013**, *51*, 2401–2422. [CrossRef]

23. Sivakumar, K.C.; Dhanesh, S.B.; Shobana, S.; James, J.; Mundayoor, S. A systems biology approach to model neural stem cell regulation by notch, shh, wnt, and EGF signaling pathways. *Omics J. Integr. Biol.* **2011**, *15*, 729–737. [CrossRef] [PubMed]

24. Cau, E.; Blader, P. Notch activity in the nervous system: to switch or not switch? *BMC Neural Dev.* **2009**, *4*, 36. [CrossRef] [PubMed]

25. Catherine, D.; Zhenwei, L.; Ji-Hun, K.; Sanders, C. Notch Transmembrane Domain: Secondary Structure and Topology. *Biochemistry* **2015**, *54*, 3565–3568.