

RESEARCH

Open Access



A numerical approach for detecting switch-like bistability in mass action chemical reaction networks with conservation laws

Brandon C Reyes¹, Irene Otero-Muras² and Vladislav A Petyuk^{3*}

*Correspondence:

vladislav.petyuk@pnnl.gov
³ Biological Sciences Division,
Pacific Northwest National
Laboratory, Richland, WA
99352, USA

Full list of author information
is available at the end of the
article

Abstract

Background: Theoretical analysis of signaling pathways can provide a substantial amount of insight into their function. One particular area of research considers signaling pathways capable of assuming two or more stable states given the same amount of signaling ligand. This phenomenon of bistability can give rise to switch-like behavior, a mechanism that governs cellular decision making. Investigation of whether or not a signaling pathway can confer bistability and switch-like behavior, without knowledge of specific kinetic rate constant values, is a mathematically challenging problem. Recently a technique based on optimization has been introduced, which is capable of finding example parameter values that confer switch-like behavior for a given pathway. Although this approach has made it possible to analyze moderately sized pathways, it is limited to reaction networks that presume a uniterminal structure. It is this limited structure we address by developing a general technique that applies to any mass action reaction network with conservation laws.

Results: In this paper we developed a generalized method for detecting switch-like bistable behavior in any mass action reaction network with conservation laws. The method involves (1) construction of a constrained optimization problem using the determinant of the Jacobian of the underlying rate equations, (2) minimization of the objective function to search for conditions resulting in a zero eigenvalue, (3) computation of a confidence level that describes if the global minimum has been found and (4) evaluation of optimization values, using either numerical continuation or directly simulating the ODE system, to verify that a bistability region exists. The generalized method has been tested on three motifs known to be capable of bistability.

Conclusions: We have developed a variation of an optimization-based method for the discovery of bistability, which is not limited to uniterminal chemical reaction networks. Successful completion of the method provides an S-shaped bifurcation diagram, which indicates that the network acts as a bistable switch for the given optimization parameters.

Keywords: Systems biology, Signaling pathways, Chemical reaction network theory, Mass action kinetics, Bistability, Switch-like behavior



Background

Cellular decisions via signaling pathways are essential for complex biological systems to function. The key attribute of signaling pathways which are capable of mediating a decision-making process is switch-like behavior. Such behavior assumes that the system has (typically) two stable equilibria and there is a way to switch between them. In bistable switches, two different thresholds for switching back and forth ensure the robustness of the decision. This characteristic dose–response behaviour is called hysteretic. Distortion of these signaling pathways with switch-like behavior manifests in no switching at all, switching at incorrect input signals, irreversible switching, etc. These malfunctions result in incorrect cell decisions and may be one of the underlying causes of developmental disorders, cancer, diabetes, and presumably a number of other pathologies [1–3]. Given that cellular decisions can consist of an immense number of biochemical interactions, smaller network motifs are often considered and can help elucidate the portions of the signaling pathway that are key to the decision-making process [4].

Although discovering essential network motifs can provide a wealth of information, obtaining these configurations is not only difficult, but costly if approached purely from an experimental point of view. For this reason, it is imperative for this process to be coupled with mathematical modeling, which can act as a guide for designing experiments. Analysis of these models is useful as they can determine the existence of bistability in the signaling pathway, an attribute directly tied to the pathway's ability to exhibit switch like behavior. Existence of bistability in chemical reaction networks has been an active area of research since the 1970s. In particular, a wealth of mathematical theory has been oriented towards network motifs that utilize mass action kinetics for the participating reactions. This is due to the fact that mass action law does not employ assumptions on time- and concentration-scale separation as do other kinetics, such as Michaelis-Menten [5–8].

One well-established theoretical framework to preclude multistationarity (and therefore bistability) in a network motif following mass action kinetics was developed by Feinberg, Horn, and Jackson [9, 10]. This theory, aptly named Chemical Reaction Network Theory (CRNT) uses the underlying structure of the reactions in the network to identify key properties. CRNT has produced results such as the Deficiency Zero and One Theorems which preclude bistability for certain network structures, irregardless of the kinetic constant values [9]. Although these theorems are very powerful, it is often the case that more complex networks found in cell signaling do not meet the deficiency requirement of these theorems [11]. To consider networks with higher deficiency, the Deficiency One Algorithm and Advanced Deficiency Algorithm were developed [11, 12]. These algorithms use the structure of the network to construct a system of equalities and inequalities. If these systems are solvable by either linear or nonlinear programming, they can state the existence of multiple positive steady states. However, these methods do not provide any conclusions about switchability between the alternative steady states. For more details on the topic of switchability between steady states we would like to refer the reader to [13]. In addition to the Deficiency One Algorithm and Advanced Deficiency Algorithm, injectivity theory and network concordance tests have also been developed, which attempt to address those networks that are not covered by the aforementioned theory [14, 15].

Besides CRNT and linear/nonlinear programming, there are a number of alternative approaches for gaining insights into an ODE system's behavior. For example, algebraic methods utilizing the Gröbner basis approach [13, 16], cylindrical algebraic decomposition [17] and sign pattern analysis of the Jacobian [18] have been developed in concert with CRNT. The goal of these methods is to obtain analytical solutions to the system of ODEs at the steady state and then further analyze the solutions to detect bistability. A number of methods have also been reviewed and compared in [19]. Additionally, promising approaches have been developed that attempt to determine the parameter regions where multistability may occur using symbolic [20] or numerical [21] methods.

Recently, the optimization-based approach of [22] has gained attention as it provides an efficient procedure to search for bistability with respect to the total amount of a certain chemical moiety (e.g. ligand) that is reflected by a conservation law. For brevity we will refer to this method as the uniterminal approach (where the term uniterminal is defined in the Basic notation section) as the approach is limited to uniterminal networks. This particular approach attempts to find bistability by constructing an efficient optimization problem to find a saddle-node and then uses numerical continuation to confirm if the saddle-node is a saddle-node bifurcation (in general for signaling pathways under study bistability occurs via saddle-node bifurcations). This approach is of considerable interest because it allows (1) evaluation of fairly large pathways and (2) directly imposing bounds on values of the parameters of the network, such as the species' concentrations and kinetic constants. The uniterminal approach is constructed based on the assumption that (1) every reaction is endowed with mass action kinetics, (2) the network admits a strictly positive steady state (where the concentrations of all the species are positive) and (3) the network is uniterminal.

Although this optimization approach in combination with hybrid optimization solvers [23] works quite efficiently, it is limited to networks that are uniterminal. This constraint is a direct result of assumptions made in the formulation of the optimization problem. More specifically, by assuming that the network is uniterminal, the approach is able to form a basis for the deficiency subspace, see section "Deficiency and equilibrium manifold" in [22]. A consequence of this result is the ability to form a square system of equations that define the equilibrium manifold of the ODE system that is compatible with the reaction polyhedron [24]. Furthermore, since this system is square, the approach can construct sufficient conditions for a saddle-node in the presence of mass conservation by minimizing the system's determinant (a scalar value available for only square systems), see section "Sufficient conditions for a saddle-node in presence of mass conservation" in [22]. To address this limiting condition and extend the reach of this optimization approach, we constructed a general technique that can investigate the bistability of reaction networks regardless of their terminality.

The general approach (as depicted in Fig. 1) uses the network structure, represented in SBML format [25], to construct an optimization problem that searches for a saddle-node. This optimization problem differs from the uniterminal approach as it searches for a saddle-node using direct ODE stability analysis, rather than relying on constrictive assumptions. In essence, the optimization problem consists in minimizing an objective function represented by the squared determinant of the Jacobian, so that a minimal value of zero guarantees the presence of a zero eigenvalue. Once a set of parameters conferring a saddle-node is

found by this minimization, then production of a bifurcation diagram is attempted in one of two ways. One way is based on a well-established numerical continuation technique. Alternatively, the ODE system can be directly simulated from different initial conditions to compute the dose–response curve. In this case the dose–response curve serves as a surrogate of a bifurcation diagram except it does not contain an unstable branch. This option, although more computationally intensive, allows one to determine if the system is bistable in cases where numerical continuation is not possible due to an ill-conditioned Jacobian.

Basic notation

Assuming mass action kinetics, a given chemical reaction network can be represented as a system of autonomous ODEs composed of N species and R reactions

$$\dot{\mathbf{c}} = f(\mathbf{c}, \mathbf{k}). \quad (1)$$

Here $\dot{\mathbf{c}}$ denotes the temporal derivative of the species concentrations $\mathbf{c} \in \mathbb{R}^N$, $\mathbf{c} \geq 0$, with each species having concentration c_i for $i = 1, \dots, N$. The vector $\mathbf{k} \in \mathbb{R}^R$, $\mathbf{k} > 0$, represents the kinetic rate constants of the reactions (determined by mass action kinetics), where the individual kinetic rate constants are denoted as k_i for $i = 1, \dots, R$. We also denote the individual reactions as r_i for $i = 1, \dots, R$. In addition to the system of ODEs, we also require that the network have one or more conservation laws, which we denote as follows:

$$\mathbf{C} = g(\mathbf{c}),$$

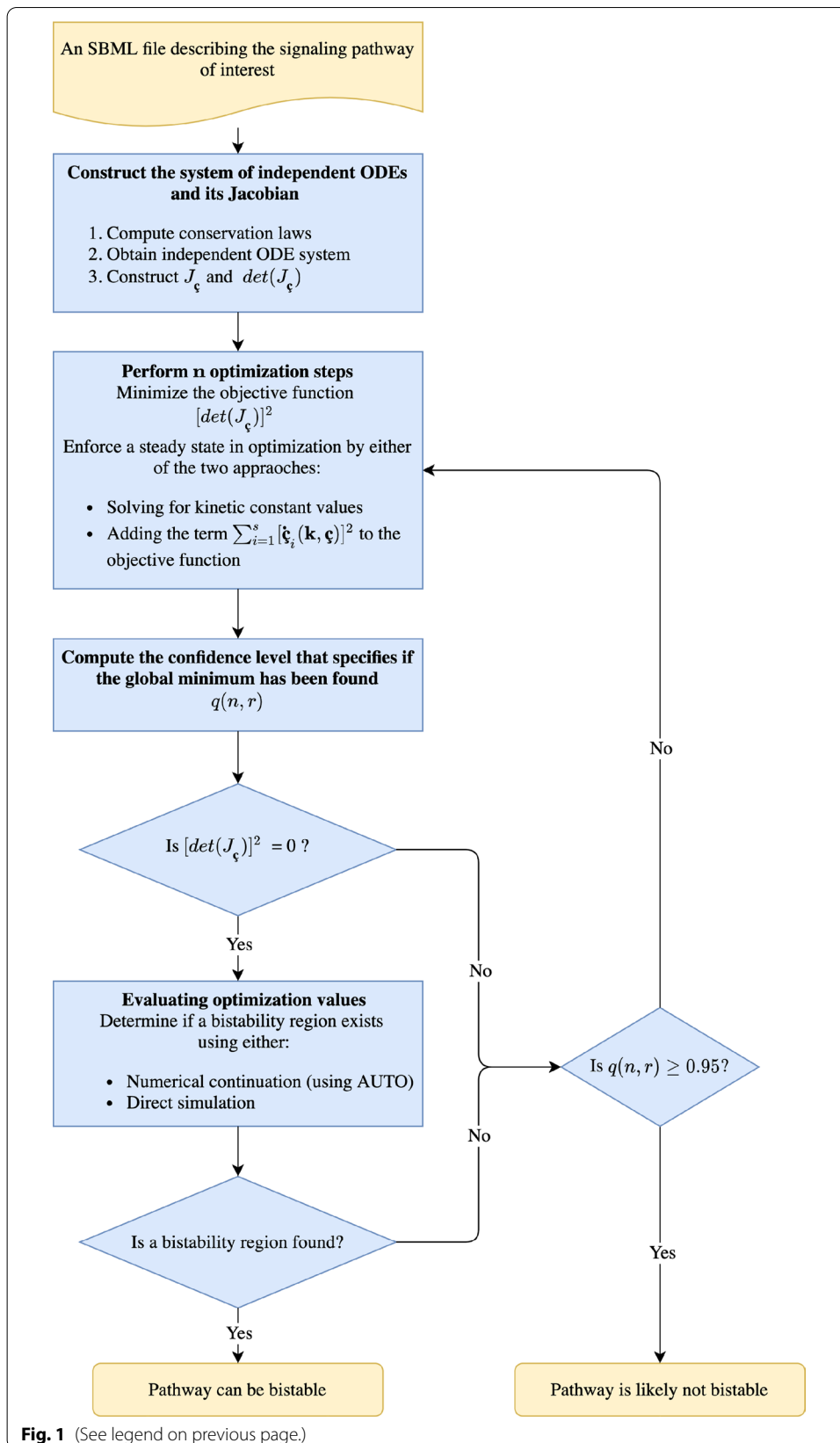
where $\mathbf{C} \in \mathbb{R}^\lambda$, λ denoting the number of conservation laws, and each conservation law is given as C_i for $i = 1, \dots, \lambda$.

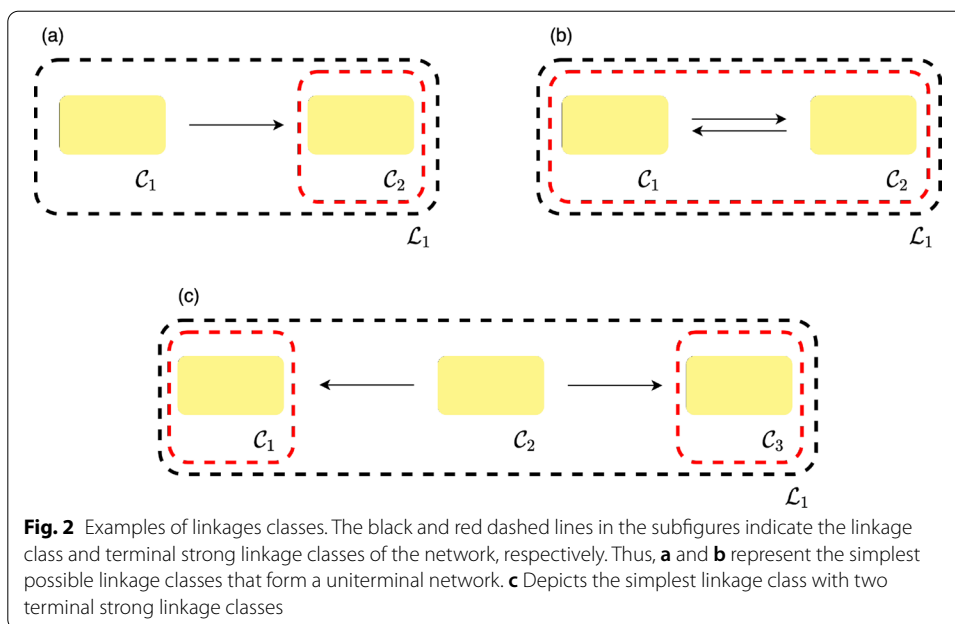
The following section is heavily based on the CRNT terminology, such as complexes and linkage classes. Briefly, a complex (in CRNT terminology) is a list of reactants or products for a given reaction. They are denoted as \mathcal{C}_i for $i = 1, \dots, M$. Complexes and corresponding reactions constitute the vertices and edges of the chemical reaction network or graph (also known as a C-graph). Disconnected subgraphs of this network are called linkage classes. For more formal definitions and descriptions we would strongly encourage the reader to review the original lectures on CRNT [26] and the prior publication [22].

For a given network we let ℓ be the number of linkage classes and denote them as \mathcal{L}_i for $i = 1, \dots, \ell$. By inspecting a linkage class further, one can then determine if a given network is uniterminal using Definitions 1, 2, and 3. The simplest depictions for a network being uniterminal and biterminal (Definition 4) are depicted in Fig. 2. Note that in the uniterminal case a linkage class containing a single complex is itself uniterminal as a single complex is strongly linked to itself [9]. Although this is assumed by CRNT, this is of little or no interest in application (as it essentially means no reaction occurs) and is simply stated for theoretical completeness.

(See figure on next page.)

Fig. 1 Workflow of the general approach for bistability detection in mass action chemical reaction networks. The approach is based on finding conditions that produce a Jacobian evaluated at a steady state, which has one zero eigenvalue. The symbol $\boldsymbol{\zeta}$ denotes the independent species' concentrations. A confidence level defining if a global minimum of the optimization problem has been found is denoted as $q(n, r)$





Definition 1 (*Strongly linked nodes*) Two nodes C_i, C_j are said to be strongly linked if there is a directed path from C_i to C_j and also a directed path from C_j to C_i .

Definition 2 (*Terminal strong linkage class*) A terminal strong linkage class is a maximal set of nodes within a linkage class such that there is no edge pointing to any other set of nodes that are strongly linked.

Definition 3 (*Uniterminal network*) A network graph is uniterminal if every linkage class in the graph contains only one terminal strong linkage class.

Definition 4 (*Biterminal network*) A network graph is biterminal if it contains at least one linkage class that has two terminal strong linkage classes.

Bistability in reaction networks

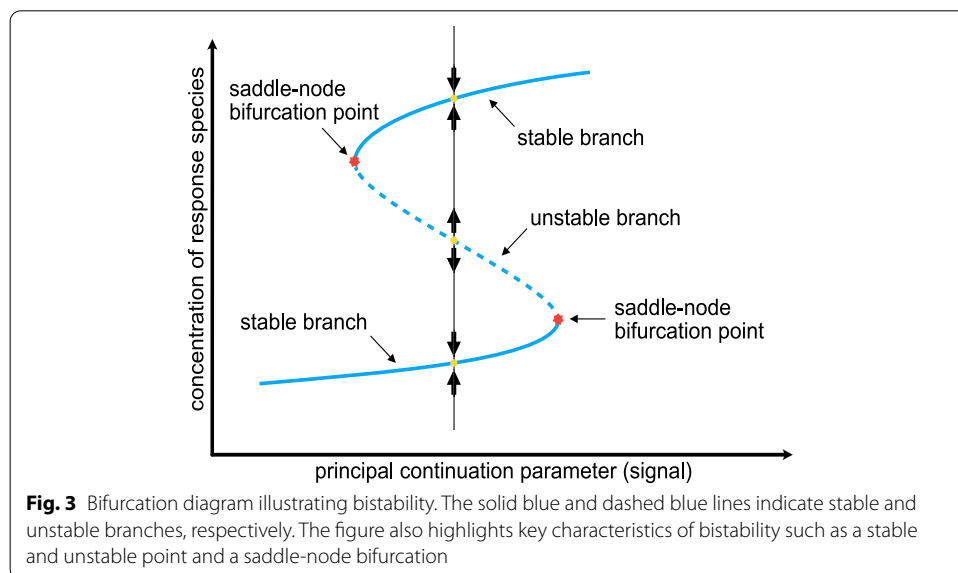
Provided system (1), it's possible that the chemical reactions will resolve to different equilibrium states depending on the starting conditions. Intuitively, such non-linear phenomenon can occur when the equilibrium concentration of one species is a polynomial function with degree greater than one with respect to an individual species. In biological systems, this behavior may present itself in a situation where a particular value of a signal species produces more than one solution with respect to a response species. In particular, bistability resulting from two stable branches connected by an unstable branch mimics switch-like action. Identification of such behavior is the focus of this manuscript.

Determining the stability of an ODE system is achieved by considering the eigenvalues of the Jacobian at a steady state point. The Jacobian of the ODE system is a matrix

representation of the first-order partial derivatives of the ODE system with respect to state variables. By constructing the Jacobian one can approximate any ODE system with a system of linear ODEs at a given point. The stability at the steady state is defined by the sign of the real part of the Jacobian’s eigenvalues. If any of the eigenvalues have a positive real component, the deviation from the steady state will increase in time, which implies an unstable steady state. In the case where all of the eigenvalues have a negative real component, the deviation decreases with time, indicating a stable steady state. In some cases, such as a bistable system, a single zero eigenvalue indicates that a steady state is momentarily transitioning from stable to unstable or vice versa. This transition happens as a consequence of varying a parameter of the ODE system.

In a mathematical sense, this transition from a stable state to an unstable state can be exhibited by a saddle-node bifurcation. Sufficient conditions for a saddle-node bifurcation are given by Theorem 1, where the point must also be a saddle-node according to Definition 5. For our purposes, F in Definition 5 corresponds to the linearly independent system of ODEs with a variable signal, u is the independent species’ concentrations, and ϵ_0 is the signal of interest. Throughout the text we refer to linearly independent ODEs as independent ODEs.

It should be noted that over the years the naming of a saddle-node bifurcation point has become somewhat convoluted, in other literature it is also referred to as a turning point, fold bifurcation point, or limit point bifurcation [27]. In a bistable system, two saddle-node bifurcation points delimit the range of the signal (or bifurcation parameter) for which two stable branches of steady states exist. It is this bistability phenomenon that we are particularly interested in identifying. An example scenario for bistability is depicted in Fig. 3, where the parameter being varied for the bifurcation analysis is the signal of the reaction network, the solid blue line denotes a stable branch, and the dashed blue line represents an unstable branch.



Definition 5 ([28] *Saddle-node*) When considering an n -dimensional system of ODEs, F , we say that $u_0 \in \mathbb{R}^n$ is a saddle-node for $F : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ at ϵ_0 if $F(u_0, \epsilon_0) = 0$, the linear transformation $DF(u_0, \epsilon_0) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ has zero as an eigenvalue with algebraic multiplicity of one, and all other eigenvalues have nonzero real parts, where $DF(u_0, \epsilon_0)$ denotes the Jacobian of F with respect to u evaluated at $u = u_0$ and $\epsilon = \epsilon_0$.

Theorem 1 contains some technical terminology that requires a brief introduction. The smoothness of a function depends on the number of continuous derivatives that exist over the function’s domain. The range of a matrix is the span of its column vectors. The kernel (or nullspace) of a matrix is the linear subspace of the domain of the map which is mapped to the zero vector. Lastly, the Fréchet derivative is the generalized form of the derivative of a real-valued function. Further details on these mathematical concepts can be found in the original text [28] and elsewhere.

Theorem 1 ([28] *Saddle-node Bifurcation Theorem*) *Suppose that $F : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ is a smooth function, $u = u_0$ is a saddle-node for F at $\epsilon = \epsilon_0$, and the kernel of the linear transformation $D_u F(u_0, \epsilon_0) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is spanned by the nonzero vector $k \in \mathbb{R}^n$. If $D_\epsilon F(u_0, \epsilon_0) \in \mathbb{R}^n$ and $D_{uu} F(u_0, \epsilon_0)(k, k) \in \mathbb{R}^n$ are both nonzero and both not in the range of $D_u F(u_0, \epsilon_0)$, then there is a saddle-node bifurcation at $u = u_0$. Here we define $D_{uu} F(u_0, \epsilon_0)(k, k)$ as the second Fréchet derivative of F evaluated at (u_0, ϵ_0) in the directions given by k and k . Additionally, $D_\epsilon F(u_0, \epsilon_0)$ denotes the directional derivative of F with respect to the last vector of the canonical basis.*

Results

Bistability in a biterminal futile signaling cycle

Using the general technique established in the Methods section, we will continue by considering a non-uniterminal example. For this example we will be considering a key reaction network found predominantly in eukaryotic signaling systems, namely a futile signaling cycle, that exhibits bistability when featuring a two-state kinase, as presented in [29]. This reaction network is represented in Fig. 4.

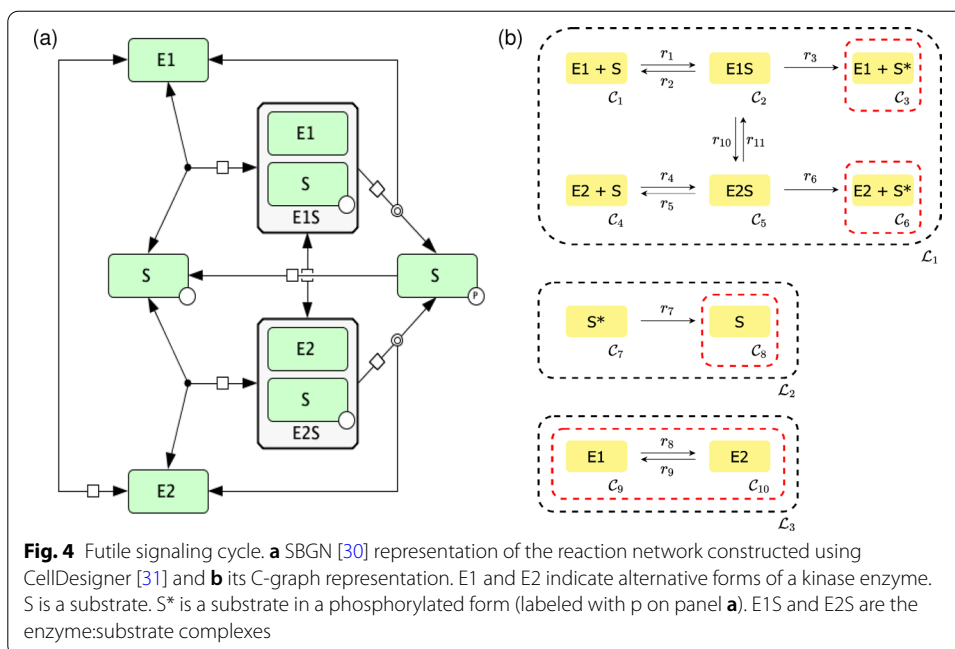
From Fig. 4b, one can see that the linkage class \mathcal{L}_1 has two terminal strong linkage classes (due to reactions r_3 and r_6). Thus, we have a biterminal linkage class and we cannot apply the theory presented in [22]. Letting

$$c_1 = [S], c_2 = [E1], c_3 = [E1S], c_4 = [S^*], c_5 = [E2], c_6 = [E2S],$$

and performing the steps outlined in the Methods section, we obtain the following independent ODEs

$$\begin{aligned} \dot{c}_3 &= k_1(E_{tot} - c_5 - c_3 - c_6)(S_{tot} - c_4 - c_3 - c_6) - k_2c_3 - k_3c_3 - k_{10}c_3 + k_{11}c_6 \\ \dot{c}_4 &= -k_7c_4 + k_3c_3 + k_6c_6 \\ \dot{c}_5 &= -k_4c_5(S_{tot} - c_4 - c_3 - c_6) + k_5c_6 + k_6c_6 + k_8(E_{tot} - c_5 - c_3 - c_6) - k_9c_5 \\ \dot{c}_6 &= k_4c_5(S_{tot} - c_4 - c_3 - c_6) - k_5c_6 - k_6c_6 + k_{10}c_3 - k_{11}c_6, \end{aligned}$$

with conservation laws



$$E_{tot} = c_2 + c_5 + c_3 + c_6$$

$$S_{tot} = c_1 + c_3 + c_4 + c_6,$$

and fix particular kinetic rate constants to expressions (as described in the Methods section) in order to ensure a steady state in the ODE system

$$k_1 = \frac{k_7c_4 + k_2c_3 - k_6c_6 + k_{10}c_3 - k_{11}c_6}{(E_{tot} - c_5 - c_3 - c_6)(S_{tot} - c_4 - c_3 - c_6)}$$

$$k_3 = \frac{k_7c_4 - k_6c_6}{c_3}$$

$$k_4 = \frac{k_5c_6 + k_6c_6 - k_{10}c_3 + k_{11}c_6}{c_5(S_{tot} - c_4 - c_3 - c_6)}$$

$$k_8 = \frac{k_9c_5 - k_{10}c_3 + k_{11}c_6}{E_{tot} - c_5 - c_3 - c_6}.$$

After performing the optimization routine, we obtain the following values from our decision vector

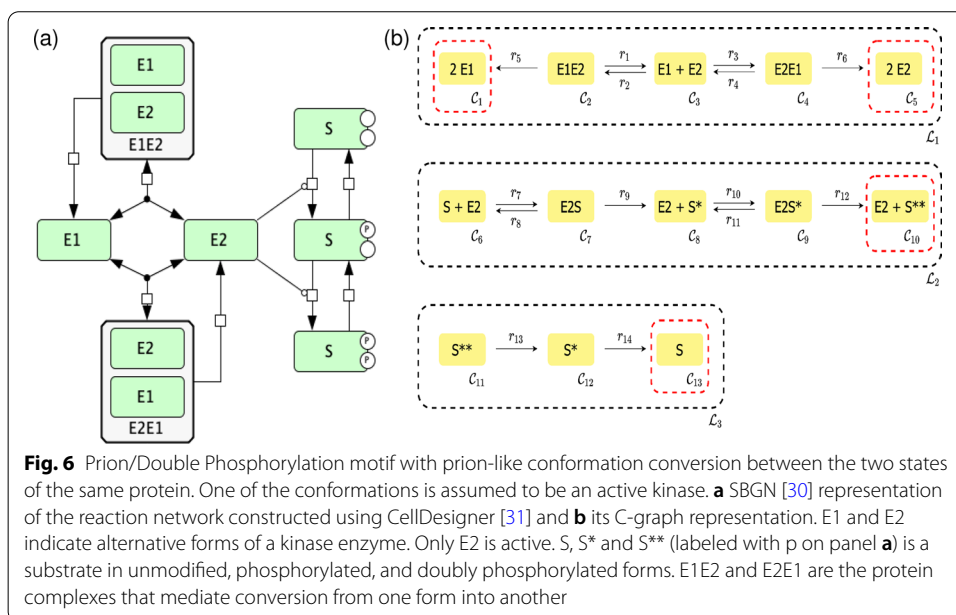
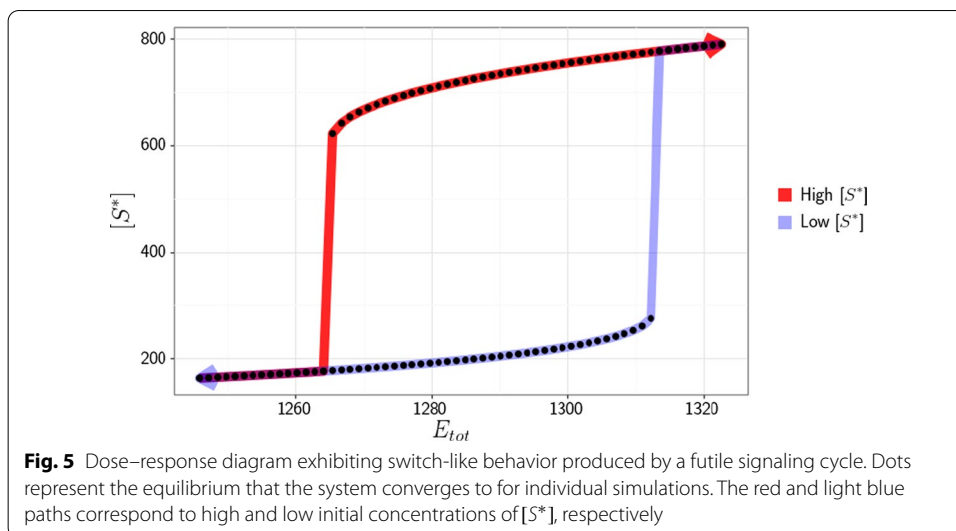
$$k_1 = 0.738028, k_2 = 3.316128, k_3 = 0.001800, k_4 = 0.050541, k_5 = 0.029997,$$

$$k_6 = 0.53685, k_8 = 2.310671, k_9 = 0.001081, k_{10} = 0.008850, k_{11} = 3.437878,$$

$$k_7 = 0.065141, c_3 = 957.287983, c_4 = 605.457182, c_5 = 136.645046,$$

$$c_6 = 70.255623, E_{tot} = 1265.114144, \text{ and } S_{tot} = 1672.513735.$$

Using the values for the species' concentrations and kinetic rate constants found, we can now begin to construct the dose–response curve. This is done by using several different concentrations of species S* for the initial value and simulating the ODE system until it has reached a steady state. For this particular example, we obtain the dose–response diagram presented in Fig. 5.



Bistability in a biterminal Prion/Double Phosphorylation motif

Next we consider the hypothetical mechanism for prion-like conformation conversion between two states of a protein described in Fig. 6. Kinase can be in two conformations E1 and E2. Conversion between E1/E2 proceeds through a prion-like mechanism, that is catalyzed by the enzyme E in the corresponding conformation. Only one conformation of kinase, E2, is active and phosphorylates substrate S in two-steps.

From inspection, one can see that \mathcal{L}_1 is a linkage class with two terminal strong linkage classes (due to reactions r_5 and r_6). Thus, we have a biterminal linkage class and we cannot apply the theory presented in [22]. Letting

$$c_1 = [E1], c_2 = [E2], c_3 = [E1E2], c_4 = [E2E1], c_5 = [S],$$

$$c_6 = [S^*], c_7 = [SE2], c_8 = [S^{**}], c_9 = [S^*E2],$$

and performing the steps outlined in the Methods section, we obtain the following independent ODEs

$$\begin{aligned} \dot{c}_2 &= k_1c_3 - c_2(k_2 + k_3)(E_{tot} - c_2 - c_9 - c_7 - 2c_3 - 2c_4) + (k_4 + 2k_6)c_4 \\ &\quad + (k_9 + k_8)c_7 - k_7c_2(S_{tot} - c_6 - c_8 - c_9 - c_7) + (k_{12} + k_{11})c_9 - k_{10}c_6c_2 \\ \dot{c}_3 &= k_2c_2(E_{tot} - c_2 - c_9 - c_7 - 2c_3 - 2c_4) - k_1c_3 - k_5c_3 \\ \dot{c}_4 &= k_3c_2(E_{tot} - c_2 - c_9 - c_7 - 2c_3 - 2c_4) - k_4c_4 - k_6c_4 \\ \dot{c}_6 &= k_9c_7 - k_{14}c_6 + k_{11}c_9 - k_{10}c_6c_2 + k_{13}c_8 \\ \dot{c}_7 &= -k_9c_7 - k_8c_7 + k_7c_2(S_{tot} - c_6 - c_8 - c_9 - c_7) \\ \dot{c}_8 &= k_{12}c_9 - k_{13}c_8 \\ \dot{c}_9 &= -k_{12}c_9 - k_{11}c_9 + k_{10}c_6c_2, \end{aligned}$$

with conservation laws

$$S_{tot} = c_5 + c_6 + c_7 + c_8 + c_9$$

$$E_{tot} = c_1 + c_2 + 2c_3 + 2c_4 + c_7 + c_9.$$

An alternative way to ensure the steady state in the optimization problem is to directly enforce the time derivatives of the concentrations to be zero. That is, instead of constraining some of the kinetic constants with symbolic expressions, we enforce $\dot{c}_i = 0$ for $i = 1, \dots, s$ explicitly, where c_i denotes a concentration of an independent species in the objective function. Specifically, the objective function will have the additional term $\sum_{i=1}^s [\dot{c}_i(\mathbf{k}, \mathbf{c})]^2$, which ensures a steady state. For details please refer to “Perform n optimizations” box in Fig. 1 and equation (4) in Section 4.6. This approach can be helpful in cases where the Jacobian is ill-conditioned. Here, for demonstration purposes we choose not to fix the kinetic rate constants and instead use the aforementioned robust, but more computationally intensive approach. Performing the optimization routine, we obtain the following values from our decision vector

$$\begin{aligned} k_1 &= 27.963833, k_2 = 2.417993, k_3 = 2.121228, k_4 = 48.342142, k_5 = 0.910340, \\ k_6 &= 1.802118, k_7 = 17.019827, k_8 = 92.473965, k_9 = 0.021611, k_{10} = 0.782488, \\ k_{11} &= 3.692336, k_{12} = 0.205743, k_{13} = 0.063297, k_{14} = 0.235401, c_1 = 14.749224, \\ c_2 &= 18.117522, c_3 = 22.377604, c_4 = 11.304051, c_5 = 27.001718, \\ c_6 &= 8.264281, c_7 = 90.016969, c_8 = 97.695329, c_9 = 30.056006, \\ S_{tot} &= 253.034303, \text{ and } E_{tot} = 220.303031. \end{aligned}$$

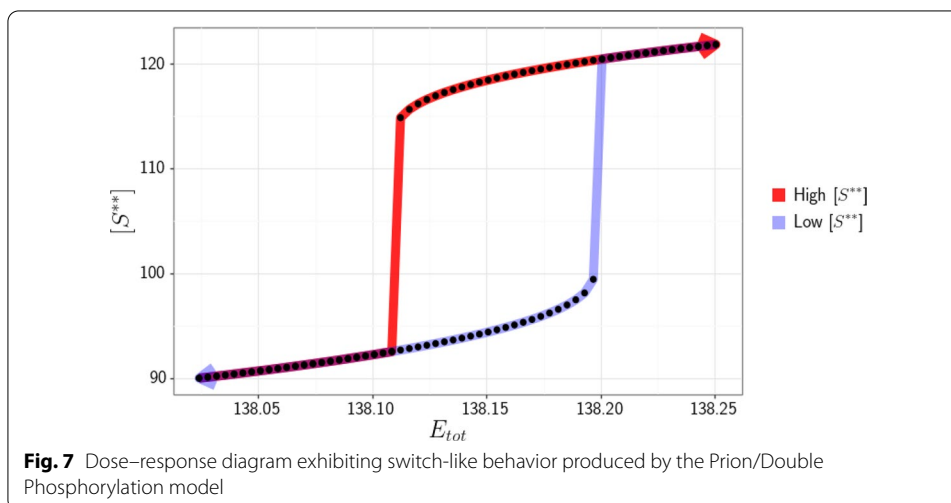
Using the values for the species’ concentrations and kinetic rate constants found, we can now directly compute the bifurcation diagram with continuation methods or alternatively obtain the dose–response curve by direct simulation. This is done by using several different concentrations of species S^{**} for the initial value and simulating the ODE system until it has reached a steady state. For this particular example, we obtain the dose–response diagram presented in Fig. 7.

Conclusions

We have developed a new general technique to detect bistability in any mass conserving reaction network. The general technique established builds on an existing approach that only allows for bistability detection in reaction networks that are uniterminal. This is achieved by first constructing an objective function using CRNT that is not dependent on the number of terminal strong linkage classes in each linkage class. Then, we formulate an optimization approach that searches for a saddle-node. Once a saddle-node is detected, the method performs either numerical continuation or direct simulation to identify if the particular set of parameters that produced the saddle-node also produce a saddle-node bifurcation. Lastly, if a saddle-node bifurcation is found, numerical continuation or direct simulation will elucidate whether or not there is another saddle-node bifurcation. The need to find two saddle-node bifurcation points is necessary as they confer back and forth switch-like behavior. Various examples provided verify the general technique and its ability to identify bistability.

This technique in its entirety is available in an updated version of CRNT4SBML (<https://crnt4sbml.readthedocs.io/>), a Python package that analyzes SBML files (the systems biology community standard for representing reaction networks) and then utilizes mathematical theories to help detect the existence of bistability in cell signaling pathways [32]. The models in SBML format and the corresponding Python code for the analysis presented in the Results and Methods sections are available at https://github.com/PNNL-Comp-Mass-Spec/CRNT4SBML/tree/master/2021_BMC_Bioinformatics_paper_code.

Although the general technique is a useful method for the detection of bistability, there are certain difficulties it may encounter. The technique can become computationally intensive for large reaction networks with very high dimensional search spaces to be explored. To overcome this, we included an option that allows the user to perform the optimization routine using parallel computing techniques (enabled by the use of MPI). This is possible since the optimization routine starts from different independent initial starting points. Additionally, even if a large portion of the objective function’s domain is explored, we cannot preclude bistability if a zero is not



found. We address this uncertainty by computing the probability that the minimum objective function value achieved is equal to the true global minimum. In the future, we would like to couple the general technique with other methods that utilize the Gröbner basis for the ODE system. This coupling could produce further insight into the conditions of bistability for particular parameters produced by the optimization.

Methods

The introduced approach is more general than those methods developed in [22] as it does not require the network to be uniterminal. Without loss of generality and for the sake of simplicity, we will be using the well-known Edelstein network, which is uniterminal. Note, the steps of the approach would be exactly the same regardless of terminality (uniterminal, biterminal, etc.) of the network. Originally, the Edelstein network was introduced in [33] and we choose to use the reduced form presented in Lecture 3 of [26]. The Edelstein network is shown in Fig. 8.

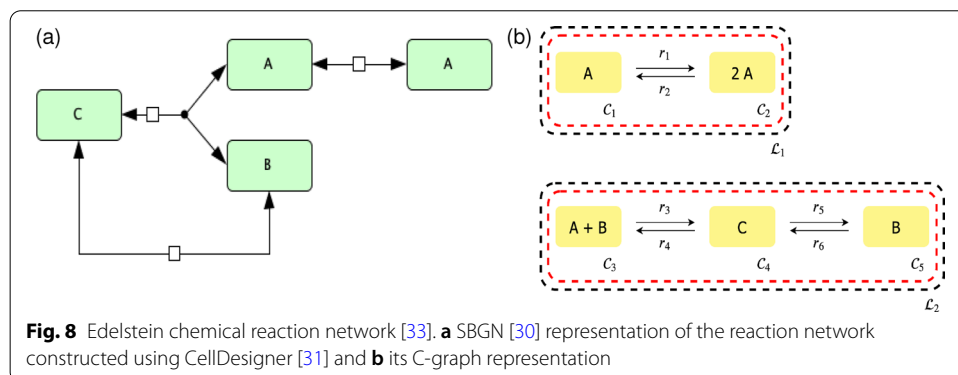
In the following subsections we develop our framework using aspects of CRNT. Here we have $\mathbf{k} = (k_1, k_2, k_3, k_4, k_5, k_6)^T$ and $\mathbf{c} = (c_1, c_2, c_3)^T$ with $c_1 = [A], c_2 = [B], c_3 = [C]$.

Step 1: Constructing the full ODE system

The first step of the approach is to construct the ODEs describing species’ concentration dynamics for the given reaction network. Assuming mass action kinetics, we obtain the following representation for our autonomous ODEs:

$$\dot{\mathbf{c}} = Y A \Psi(\mathbf{c}) \iff \begin{cases} \dot{c}_1 = k_1 c_1 - k_2 c_1^2 - k_3 c_1 c_2 + k_4 c_3 \\ \dot{c}_2 = (k_4 + k_5) c_3 - k_3 c_1 c_2 + k_6 c_2 \\ \dot{c}_3 = k_3 c_1 c_2 - (k_4 + k_5) c_3 + k_6 c_2 \end{cases} \quad (2)$$

The molecularity matrix Y is a $N \times M$ matrix where Y_{ij} corresponds to the molecularity of species i in complex j . For our example the complexes are $C_1 = A, C_2 = 2A, C_3 = A + B, C_4 = C$, and $C_5 = B$, which provide the molecularity matrix:



$$Y = \begin{matrix} & A & 2A & A+B & C & B \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

A is the $M \times M$ kinetic constant matrix, where the diagonal elements of A contain the negative of the sum of the kinetic rate constants corresponding to the reactions going out from C_i . While the off-diagonal elements, say A_{ij} , contain the kinetic rate constants of the reactions going from C_j to C_i . Here i and j correspond to the i th row and j th column of A, respectively.

$$A = \begin{matrix} & A & 2A & A+B & C & B \\ \begin{matrix} A \\ 2A \\ A+B \\ C \\ B \end{matrix} & \begin{pmatrix} -k_1 & k_2 & 0 & 0 & 0 \\ k_1 & -k_2 & 0 & 0 & 0 \\ 0 & 0 & -k_3 & k_4 & 0 \\ 0 & 0 & k_3 & -(k_4 + k_5) & k_6 \\ 0 & 0 & 0 & k_5 & -k_6 \end{pmatrix} \end{matrix}$$

The vector $\psi(\mathbf{c}) = (\psi_1, \dots, \psi_M)^T$ defines the mass action monomials (a product of species' concentrations) associated with each complex

$$\psi_j(\mathbf{c}) = \prod_{i=1}^N c_i^{Y_{ij}}, \quad j = 1, \dots, M.$$

For our example we obtain:

$$\psi(\mathbf{c}) = (c_1, c_1^2, c_1c_2, c_3, c_2)^T.$$

Step 2: Computing the mass conservation laws

The next step is to compute the mass conservation laws that govern the network. To do this, we first construct the species stoichiometric matrix S with dimension $N \times R$. Its entries can be constructed from the Y matrix by noticing that every reaction from C_i to C_j has an associated vector $Y_j - Y_i$ where Y_j and Y_i are the j th and i th columns of Y , respectively. In our example, this produces the stoichiometric matrix S :

$$S = \begin{matrix} & r_1 & r_2 & r_3 & r_4 & r_5 & r_6 \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{pmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 & -1 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{pmatrix} \end{matrix}$$

We also denote the rank, that is the maximal number of linearly independent columns, of S as s , for notational convenience. For our example, $s = 2$.

Note, that the number of mass conservation relations is $\lambda = N - s$. Matrix B (of dimension $N \times \lambda$) that defines such relations is defined as the nullspace of S^T , such that $S^T B = 0$. It should be also noted that the matrix spanning the nullspace of S^T is not uniquely determined, and we choose B such that all of its entries are nonnegative. This choice of B is always possible provided that each conservation law represents the

conservation of a chemical or moiety. This is because an individual species' concentration cannot be negative. Thus, there must be a way for a basis vector, representing a conservation law, to be expressed only with nonnegative values.

For most reaction networks, constructing a B matrix with nonnegative entries can be done by using linear programming. The construction of the linear programming problem requires a firm understanding of polyhedra. Thus, we strongly recommend that the reader consider Sections 2.2 (in particular Subsection 2.2.3) and 2.4 of [34], which describe polyhedral sets and cones, and how the nullspace of the stoichiometric matrix can be described as a polyhedral cone, respectively. Additionally, it should be noted that the mathematical definitions and theorems in the suggested sections of [34] were obtained from [35], see for example Sections 1, 2, and 19 of [35].

We now briefly describe the linear programming problem construction. The optimization problem is formed by first considering the intersection of $Null(S^T)$ and the nonnegative vectors of \mathbb{R}^N , which is an infinite cone. In order to obtain a finite set of this intersection, one can consider an intersection of the infinite cone with the set of vectors that have the sum of their coordinates equal to one. The solution to this intersection is a finite convex polytope. This conclusion can be seen by considering Definition 2.29 and 2.33 of [34]. It should be noted that [34] presents Enumeration Problems in general (Section 2.3) as a way to search for the vertices of this convex polytope. We instead search for these vertices beginning from random directions via the optimization problem (3), which is realized by the Simplex optimization method [36].

$$\begin{aligned} & \text{Minimize } \mathbf{w}^T \mathbf{x} \\ & \text{subject to:} \\ & \quad -\tilde{B}\mathbf{x} \leq \mathbf{0}, \\ & \quad \left(\sum_{i=1}^N \tilde{B}_{i,1}, \dots, \sum_{i=1}^N \tilde{B}_{i,\lambda} \right)^T \mathbf{x} = 1, \\ & \quad -\infty \leq \mathbf{x} \leq \infty \quad \mathbf{x} \in \mathbb{R}^\lambda, \\ & \quad \mathbf{w}^T \in \mathbb{R}^\lambda. \end{aligned} \tag{3}$$

In optimization problem (3), \mathbf{w}^T is the vector of random search directions and \tilde{B} is the initial B matrix with at least one negative entry. For our purposes, \tilde{B} is initially set to the basis vectors that compose the span of $Null(S^T)$. The endpoints (i.e. the minimized \mathbf{x} from (3)) of these vertices are then used to form a basis vector (consisting of only nonnegative elements) of the nullspace. Thus, by taking the dot product, $\tilde{B} \cdot \mathbf{x}$, we can obtain a vector with nonnegative entries, with values between zero and one. To obtain integer entries, one can then divide the vector by the smallest nonzero entry of the vector. By repeating this process multiple times and obtaining λ unique basis elements, one can then form a B matrix with nonnegative entries, as outlined in Algorithm 1. It should be noted that λ is representative of the number of conserved chemical moieties. Thus, by definition we are certain that $\lambda > 0$ if the system contains conservation laws. A noteworthy remark of Algorithm 1 is that it does not guarantee that one will find all λ basis vectors, given a fixed number of iterations. Thus, if the number of the found basis vectors is less than the pre-computed $\lambda = N - s$, the number of iterations should be increased. For further details of Algorithm 1, we refer the reader to the Supplementary, Additional File 1. In addition to this approach, there exists alternative ways to obtain a B matrix with nonnegative entries, see [37–39].

Set $\tilde{B} = \text{Null}(S^T)$.

Set the number of random search directions to $niter = (N + 1)(\lambda + 1)10$.

if \tilde{B} has at least one negative entry **then**

for $i = 1$ to $niter$ **do**

 Generate the random vector $\mathbf{w} \sim \mathcal{N}(0, 1)$.

 We solve optimization problem (3) using the Simplex algorithm for a given \mathbf{w} and obtain \mathbf{x} .

 Set $\mathbf{z} = \tilde{B} \cdot \mathbf{x}$.

 Divide \mathbf{z} by $\min\{\mathbf{z}\}$, where $\min\{\mathbf{z}\} \neq 0$, to the set of possible basis vectors.

end

 Create B^T from the set of possible basis vectors, where the rows correspond to λ unique basis vectors.

else

 Divide the rows of \tilde{B}^T by the minimum nonzero value in each row. In practice, this is sufficient enough to transform the matrix into the form with all integer values because the material conservation laws are written with integer coefficients.

end

Test the linear independence of B^T by confirming that there are λ pivots in the row-reduced echelon form of B^T .

Algorithm 1: Pseudocode for obtaining a B matrix with nonnegative entries.

Using S^T and the procedure outlined in Algorithm 1, we obtain the B matrix below:

$$B^T = c_1 \begin{pmatrix} A & B & C \\ 0 & 1 & 1 \end{pmatrix},$$

where C_1 again stands for the first conservation law. To obtain an explicit statement for the conservation laws, we simply take $B^T \mathbf{c}$, giving the following conservation law:

$$C_1 = c_2 + c_3.$$

Step 3: Determining the independent ODE system

For any given reaction network, the number of independent ODEs describing the system’s dynamics is equal to the $\text{rank}(S) = s \leq R$ [9]. Since we will ultimately consider the Jacobian of the ODE system with respect to the concentrations \mathbf{c} , it is necessary to remove those ODEs that can be represented as linear combinations of other ODEs in the system. We will let the independent system of ODEs be represented as follows:

$$\dot{\boldsymbol{\zeta}} = \hat{f}(\mathbf{c}, \mathbf{k}),$$

where $\boldsymbol{\zeta} \in \mathbb{R}^s$ represents those species’ concentrations that form an independent ODE system. To obtain the independent system, we will first see if $s = N$, if this is true, this implies that there are no conservation laws. This scenario is out of the scope of this

manuscript. If $s < N$ then we have conservation laws and we must continue by finding the independent system. It is at this point where knowledge of the response in the bifurcation analysis is used to determine which ODEs should remain in the independent ODE system.

For our particular example, we take C_1 (the sum of c_2 and c_3) as the input signal, and c_1 as the output or readout of the system's response. Since c_1 (the concentration of species A) is taken as our system's response, it is necessary for \dot{c}_1 to be included in $\dot{\mathbf{c}}$. Using this fact and conservation law C_1 , we can see that c_2 has a dependence on c_3 of the form $c_2 = C_1 - c_3$. For this simple example, this information is sufficient enough to eliminate \dot{c}_2 from $\dot{\mathbf{c}}$. Indeed if we consider the full ODE system (2), we see that $\dot{c}_2 = -\dot{c}_3$. Thus, for our particular example we have $\dot{\mathbf{c}}$ given by:

$$\begin{aligned} \dot{c}_1 &= k_1c_1 - k_2c_1^2 - k_3c_1c_2 + k_4c_3 \\ \dot{c}_3 &= k_3c_1c_2 - (k_4 + k_5)c_3 + k_6c_2. \end{aligned}$$

For larger systems the above methodology may not be straightforward. Thus, we suggest a procedure that exhaustively looks through combinations of species (excluding the response species), one from each conservation law, removes corresponding ODEs and checks if the rank of the remaining ODE system remains the same. The pseudocode for this procedure is provided in Algorithm 2.

Let y be the desired response species.

Let PDS be a list of length λ initialized with empty lists as its elements.

for $i = 1$ to λ **do**

 Set the i th element of PDS to the list of all species (not including y) in C_i .

end

Create DS , an exhaustive list containing unique unordered combinations, by choosing one of the species from each of the PDS lists. Remove combinations that have repeating species.

for each element in DS **do**

 Let P be the matrix formed by choosing the rows of YA corresponding to those species not in the element of DS .

if $rank(P) = s$ **then**

 Set \mathfrak{c} to those species not in the element of DS .

 Construct $\dot{\mathfrak{c}}$ by using those species in \mathfrak{c} .

break

end

end

Algorithm 2: Pseudocode for finding a linearly independent ODE system using exhaustive search. Where PDS denotes the lists of possible dependent species for each conservation law and DS contains all lists that represent our dependent species.

Referring to Definition 5, we see that u and ϵ_0 correspond to \mathfrak{c} and the input signal in our example, respectively, with $u = (c_1, c_3)^T$ and $\epsilon_0 = C_1$. From this definition we see that some modifications to $\dot{\mathfrak{c}}$ must be made, in particular, we need to include the conservation law C_1 . To eliminate all the dependent species (DS) we express them in terms of

ζ and the corresponding conservation law. In our example, this is done by replacing c_2 with $C_1 - c_3$. For our example, we then have $\dot{\zeta}$ given as follows:

$$\begin{aligned} \dot{c}_1 &= k_1c_1 - k_2c_1^2 - k_3c_1(C_1 - c_3) + k_4c_3 \\ \dot{c}_3 &= k_3c_1(C_1 - c_3) - (k_4 + k_5)c_3 + k_6(C_1 - c_3). \end{aligned}$$

The system above is then equivalent to $F((c_1, c_3)^T, C_1)$.

Step 4 : Preserving the steady state

Now that we have our ODE system in terms of our input signal (or bifurcation parameter), we can now continue by constructing the expressions that are necessary to check the requirements of Definition 5. To do this, the first item we consider is the statement that $F(u_0, \epsilon_0) = 0$, this denotes that the ODE system is at a steady state. Thus, we must ensure that $\dot{\zeta} = \mathbf{0}$. One straightforward way of doing this is to add the term $\sum_{i=1}^s [\dot{\zeta}_i(\mathbf{k}, \zeta)]^2$ to the objective function (equation (4) in Section 4.6), which represents the sum of the squared derivatives. Although this formulation is robust, the trade-off includes increased computational time.

To avoid this issue, we suggest an alternative, based on symbolic constraints, formulation of the optimization problem. In this section, we propose splitting the kinetic constants into two sets, one set that is free ($\mathbf{k} \in \mathbb{R}^{R-s}$) and another set that will be fixed using symbolic expressions ($\tilde{\mathbf{k}} \in \mathbb{R}^s$). We analytically solve for specific kinetic constants, $\tilde{\mathbf{k}}$, to enforce a steady state. By solving for these expressions analytically, the domain of feasible points is reduced, providing a simpler optimization problem. This can be performed systematically because we are using mass action kinetics to form our ODEs. Furthermore, we are always guaranteed a unique analytical expression for a steady state (in terms of a particular choice of kinetic constants). To see this, consider the stoichiometric matrix S , which has columns corresponding to the reactions. Based on the law of mass action, the rate of the reaction is expressed as the product of the concentrations of the reactants, with each concentration raised to a power equal to the stoichiometric coefficient, multiplied by a corresponding kinetic constant. Thus, there is a one-to-one correspondence between the notation of reactions and kinetic constants. As stated in step 3 (Subsection 4.3), $\text{rank}(S) = s \leq R$, for any given reaction network. By definition, s corresponds to the number of linearly independent columns of S and since the columns of S correspond to the kinetic constants, we are provided with a set of kinetic constants that form a linearly independent system of equations.

In practice, it is quite simple to determine the expressions for $\tilde{\mathbf{k}}$, that should be chosen to create a linear system. One particular way to choose these kinetic rate constants is to put S into row reduced echelon form (RREF) and choose those columns that contain pivots. In our example:

$$S = \begin{matrix} & r_1 & r_2 & r_3 & r_4 & r_5 & r_6 \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{pmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 & -1 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{pmatrix} & \xrightarrow{\text{RREF}} & \begin{matrix} A \\ B \\ C \end{matrix} & \begin{pmatrix} 1 & -1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

the pivots are in the columns 1 and 3, resulting in $\tilde{\mathbf{k}} = (k_1, k_3)^T$.

Given we are looking for a steady state, we must consider the following linear system:

$$U\tilde{\mathbf{k}} = \mathbf{b},$$

Here $U \in \mathbb{R}^{s \times s}$ corresponds to the coefficient matrix produced from $\dot{\zeta}$ for a particular choice of $\tilde{\mathbf{k}}$. Vector $\mathbf{b} \in \mathbb{R}^s$ are those terms of $\dot{\zeta}$ that do not contain kinetic constants from $\tilde{\mathbf{k}}$. For our example we have the following system for our steady state:

$$\begin{pmatrix} c_1 & -c_1(C_1 - c_3) \\ 0 & c_1(C_1 - c_3) \end{pmatrix} \begin{pmatrix} k_1 \\ k_3 \end{pmatrix} = \begin{pmatrix} k_2c_1^2 - k_4c_3 \\ (k_4 + k_5)c_3 - k_6(C_1 - c_3) \end{pmatrix}.$$

Solving for k_1 and k_3 provides necessary conditions for a steady state

$$k_1 = \frac{k_2c_1^2 + k_5c_3 - k_6(C_1 - c_3)}{c_1}$$

$$k_3 = \frac{(k_4 + k_5)c_3 - k_6(C_1 - c_3)}{c_1(C_1 - c_3)}.$$

Note that for some systems, this may force a kinetic rate constant equal to zero. However, other choices of $\tilde{\mathbf{k}}$ can yield nonzero values. To account for these cases, if zero values are found, we exhaustively search through all distinct combinations of independent columns to choose $\tilde{\mathbf{k}}$ combinations and search for ones that yield no zero entries. If no combination provides nonzero values this means that no steady state exists when $\mathbf{k} > 0$ and $\mathbf{c} > 0$. In such a scenario we suggest checking the Deficiency Zero and One theorems [26]. To obtain a network that can assume a steady state with $\mathbf{k} > 0$ and $\mathbf{c} > 0$, one can consider removing the reactions that have kinetic constants equal to zero in the aforementioned computation.

Step 5: Ensuring a zero eigenvalue

From step 4, we have necessary conditions for a steady state. Thus, the next item we must satisfy is that $DF(u_0, \epsilon_0)$ must have a zero eigenvalue with a multiplicity of one and all other eigenvalues have nonzero real parts, as stated in Definition 5. Effectively, $DF(u_0, \epsilon_0)$ is the Jacobian of the right hand side of our independent ODE system with respect to ζ , we denote this as J_ζ . Given that we would like to formulate this as an optimization problem, it is easy to see that satisfying the criteria that an eigenvalue of zero must have a multiplicity of one and all other eigenvalues be nonzero, will create an expensive optimization problem because we would have to calculate the eigenvalues each time the objective function is evaluated. For this reason, we will simply search for a zero eigenvalue and then check afterwards if the eigenvalue criteria are satisfied.

To find a zero eigenvalue, consider the eigenvalue problem with eigenvalue $z \in \mathbb{C}$ and corresponding eigenvector $\mathbf{v} \in \mathbb{C}^s \setminus \{0\}$:

$$J_\zeta \mathbf{v} = z\mathbf{v}.$$

The characteristic polynomial can then be formed by taking the determinant and setting it equal to zero:

$$\det(J_\zeta - zI) = 0.$$

However, since $z = 0$, we have the following problem that reassures us that at least one eigenvalue is equal to zero:

$$\det(J_{\mathcal{C}}) = 0.$$

It is this criteria that we will use to formulate our optimization problem.

Step 6: Searching for a Saddle-node using optimization

Using the definitions from the previous steps we can create the optimization problem in two alternative ways. In both formulations, the optimization problems search for species' concentrations \mathbf{c} and kinetic rate constants \mathbf{k} that provide a system in a steady state ($\dot{\mathbf{c}} = 0$) and at least one zero eigenvalue of the associated Jacobian $J_{\mathcal{C}}$. One formulation enforces a steady state by explicitly requiring that the derivatives of the concentrations be zero through an additional term in the objective function:

$$\begin{aligned} &\text{Minimize} \quad [\det(J_{\mathcal{C}})(\mathbf{k}, \mathbf{c})]^2 + \sum_{i=1}^s [\dot{c}_i(\mathbf{k}, \mathbf{c})]^2 \\ &\text{subject to:} \quad \mathbf{c}_L \leq \mathbf{c} \leq \mathbf{c}_U \quad \mathbf{c} \in \mathbb{R}^N, \\ &\quad \quad \quad \mathbf{k}_L \leq \mathbf{k} \leq \mathbf{k}_U \quad \mathbf{k} \in \mathbb{R}^R. \end{aligned} \tag{4}$$

Here, \mathbf{c}_L and \mathbf{c}_U are the lower and upper bounds for the species' concentrations, respectively and similarly, \mathbf{k}_L and \mathbf{k}_U are the lower and upper bounds for the kinetic rate constants, respectively. This results in a more complex objective function, which is due to the addition of more nonlinear terms in the objective function and a higher dimensional decision vector. This approach is more robust as it samples from a broader space of solutions, but computationally intensive. Another approach implicitly enforces a steady state by solving for kinetic constants (as in Step 4):

$$\begin{aligned} &\text{Minimize} \quad [\det(J_{\mathcal{C}})(\mathbf{x})]^2 \\ &\text{subject to:} \quad \mathbf{c}_L \leq \mathbf{c} \leq \mathbf{c}_U \quad \mathbf{c} \in \mathbb{R}^N, \\ &\quad \quad \quad \mathbf{k}_L \leq \mathbf{k} \leq \mathbf{k}_U \quad \mathbf{k} \in \mathbb{R}^R. \end{aligned} \tag{5}$$

In this version of the optimization problem, the decision vector is $\mathbf{x} = (\{\mathbf{k}\} - \{\tilde{\mathbf{k}}\}, \{\mathbf{c}\})^T$, where $\{\mathbf{k}\} - \{\tilde{\mathbf{k}}\}$ corresponds to the set of kinetic rate constants in \mathbf{k} that are not in $\tilde{\mathbf{k}}$ and $\{\mathbf{c}\}$ is the set of species' concentrations. The kinetic rate constants $\tilde{\mathbf{k}}$ are then found by using the solutions found from solving the linear system in step 4, which reassure a steady state occurs, and the conservation constants C_1, \dots, C_λ are found using the conservation laws in step 2.

Since the approach is aimed towards analysis of biological pathways, we can apply bounds to the optimization problem based on the BioNumbers database [40]. To bound the species' concentrations, one can choose the typical range of protein concentrations in a cell, 5×10^{-13} to 5×10^{-7} M. The complex formation kinetic constants can be bounded between 10^4 and $10^8 \text{ M}^{-1}\text{s}^{-1}$. A common range for complex dissociation constants is from 10^{-5} to 10^{-3} s^{-1} . Finally, the enzyme catalysis kinetic constants range from 10^{-3} to 1 s^{-1} . These ranges are the defaults in the CRNT4SBML Python package. However, they can be overwritten by the user to more narrow ranges if some more accurate

prior knowledge exists. For further details we refer the reader to the corresponding CRNT4SBML documentation page.

Due to their nature, like the optimization problem in [22], (5) and (4) are non-convex and multi-modal. For this reason, a global optimization procedure should be used to search for the optimal solution of (5) and (4). For the results obtained throughout the manuscript, we utilize the global optimization algorithm Dual Annealing, starting from different random initial starting points that are bounded by the species' concentrations and kinetic constant values. Dual Annealing is an optimization algorithm that combines Simulated Annealing [41] and additionally applies a local search on accepted locations [42]. In particular, we utilize the Nelder-Mead simplex algorithm for all local searches. Given that we are looking specifically for a zero eigenvalue, the objective function of the optimization should obtain a minimum of zero. All points that have a zero value (within machine precision) are considered for further evaluation. The next evaluation step is to check for i) a saddle-node using Definition 5 and discard points that produce more than one eigenvalue that is zero (which might indicate a codim 2 bifurcation such as the Bogdanov-Takens bifurcation [43]), and ii) a saddle-node bifurcation point using Theorem 1. This allows one to remove unnecessary runs of numerical continuation. Although explicitly checking the criteria for a saddle-node bifurcation can reduce the runtime of the approach, it should be noted that it is sufficient to conduct numerical continuation.

Bayesian stopping rule

If the optimization satisfies the condition for a saddle-node, one can state that a saddle-node exists. However, the inverse is not true: if we can't find conditions that result in a zero eigenvalue, this doesn't guarantee that a saddle-node does not exist. This is a consequence of using stochastic optimization instead of deterministic methods (based for example on interval analysis), which will provide this guarantee [44]. The issue with methods such as these is that they become computationally intractable for mass action reaction systems with more than two or three parameters, whereas stochastic optimization provides in general a good overall efficiency [45]. One of the key uncertainties in stochastic optimization is how exhaustively the procedure searched the space of all the possible solutions. To address this uncertainty we compute the probability that the achieved minimum value is the true global minimum. This way we have a quantitative estimate of the thoroughness of the optimization procedure. The probability calculation is based on the unified Bayesian stopping rule in [46] and Theorem 4.1 of [47], where the rule was first established.

For n starting optimization points, let f_k be the achieved local minimum for the k -th decision vector, f^* the true global minimum and $\tilde{f} = \min\{f_1, \dots, f_n\}$. Our objective is to estimate the probability that \tilde{f} is f^* . Let α_k and α^* denote the probabilities that a single run of the optimization routine has converged to f_k and f^* , respectively. Assuming that $\alpha^* \geq \alpha_k$ for all local minimum values f_k we may then estimate the lower bound of the probability that $\tilde{f} = f^*$ is as follows:

$$Pr[\tilde{f} = f^*] \geq q(n, r) = 1 - \frac{(n + a + b - 1)!(2n + b - r - 1)!}{(2n + a + b - 1)!(n + b - r - 1)!}$$

Here a and b are the parameters of the Beta distribution $\beta(a, b)$, where we use $a = 1$ and $b = 5$ as suggested in [46]. The term $q(n, r)$ is the confidence level, where r is the number of f_k for $k = 1, \dots, n$ that are in the neighborhood of \tilde{f} .

We say that f_k is in the neighborhood of \tilde{f} if the relative difference of f_k and \tilde{f} is less than a predefined tolerance ϵ , which is by default set to 10% in the CRNT4SBML tool:

$$\frac{|\tilde{f} - f_k|}{\tilde{f}} \leq \epsilon.$$

Given the formulation of the optimization, the lowest possible minimal value is zero. Thus, if \tilde{f} reaches a numerical zero value then we set $q(n, r) = 1.0$, skipping the computation of $q(n, r)$. Conventionally $q(n, r) \geq 0.95$ is considered an acceptable confidence level to make the conclusion that \tilde{f} is the global minimum of the objective function. If no zero value is found after achieving a high confidence level ($q(n, r) \geq 0.95$), then we suggest stopping the optimization routine as it is unlikely that there is a solution. Note that the accuracy of this Bayesian stopping rule has not been thoroughly evaluated for our optimization problem and method.

Step 7: Numerical continuation and direct simulation

Once the optimization problem is solved, we have an independent ODE system with a steady state and at least one zero eigenvalue of the Jacobian. To ensure that this is a saddle-node bifurcation we need to check if this is the only zero eigenvalue and that Theorem 1 is satisfied. In practice, we check directly for the presence of the bifurcation diagram with switch-like behavior. The existence of such a bifurcation diagram rules out undesired findings such as a Jacobian with additional zero eigenvalues or eigenvalues with zero real parts. To generate the bifurcation diagram, we utilize either the technique of numerical continuation or directly simulate the ODEs. When performing numerical continuation, we use the values provided by the optimization

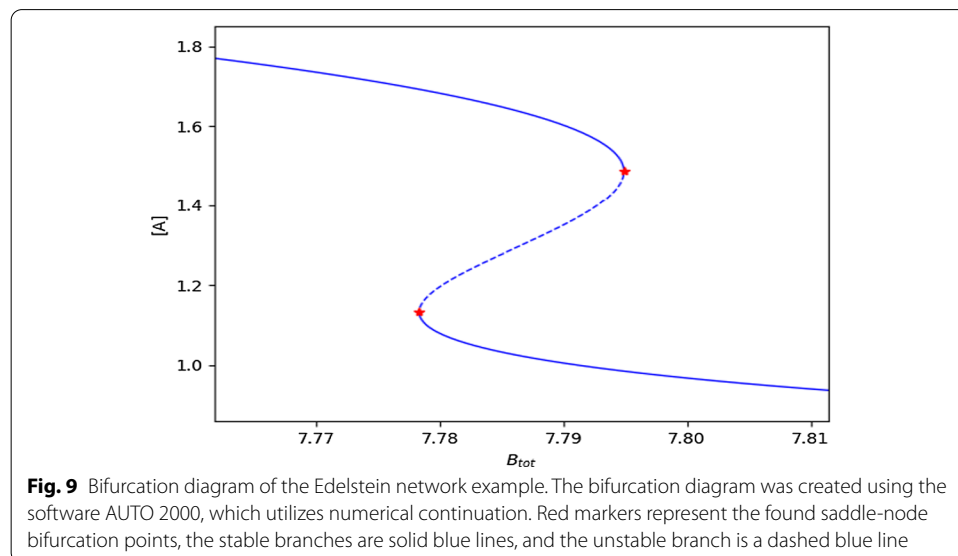
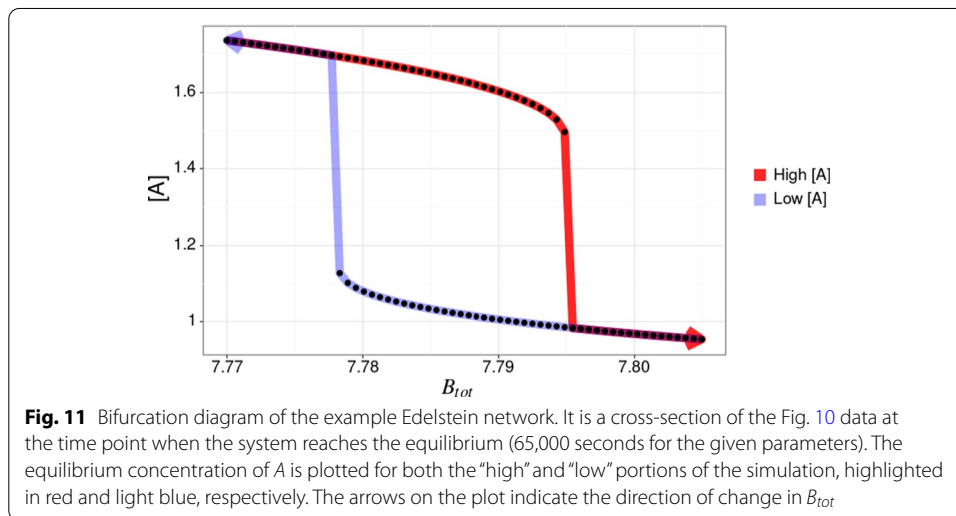
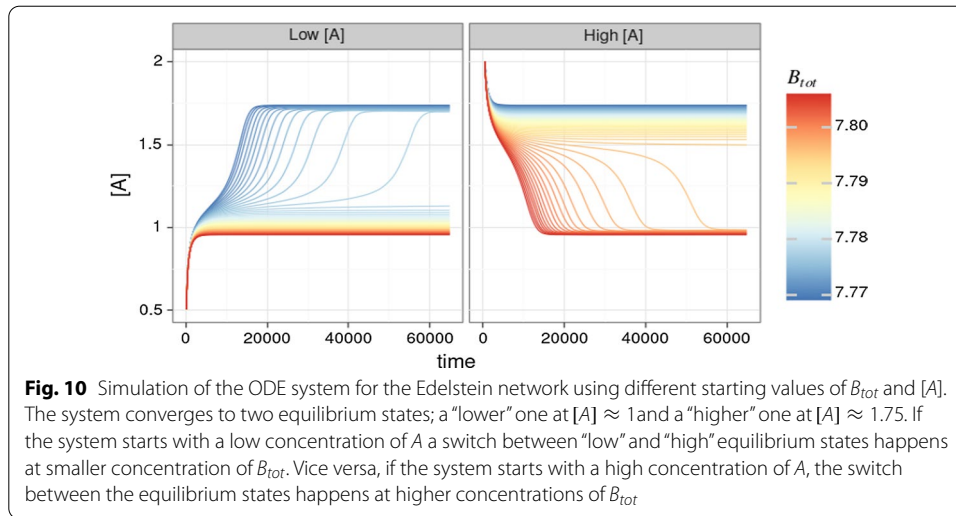


Fig. 9 Bifurcation diagram of the Edelstein network example. The bifurcation diagram was created using the software AUTO 2000, which utilizes numerical continuation. Red markers represent the found saddle-node bifurcation points, the stable branches are solid blue lines, and the unstable branch is a dashed blue line



routine and then leverage the established tool AUTO 2000 [48]. It is made accessible through the libroadrunner python library and its extension rrplugins [49]. To demonstrate numerical continuation, we use the following values provided by the optimization routine, which provide Fig. 9:

$$\begin{aligned}
 k_1 &= 0.05024595, k_2 = 0.01029913, k_3 = 0.03592955, k_4 = 0.01027423, \\
 k_5 &= 0.01272131, k_6 = 0.01006076, c_1 = 1.48685156, c_2 = 2.07275022, \\
 c_3 &= 5.72214036, \text{ and } C_1 = 7.79489058.
 \end{aligned}$$

It should be noted that it is possible that the optimization routine finds a certain combination of kinetic rate constant values that force the Jacobian of the system to be ill-conditioned or even singular, even if species’ concentrations are varied. The situation when the Jacobian is always singular precludes the numerical continuation approach. To overcome this type of situation we offer the option of direct simulation, which varies the user defined signal and initial conditions values for the ODE system and then integrates

the ODEs until a steady state occurs. The steady state is obtained when the species' concentrations do not change between the ODE integration steps, within a predefined tolerance (here 1e-06). Given the direct simulation method is numerically integrating the system of ODEs, this method will often take longer than the numerical continuation routine. Although this is the case, direct simulation is more robust and may be able to provide a bifurcation diagram when numerical continuation cannot.

Considering high and low concentrations of species *A* (response) for the initial value, we obtain the ODE simulations in Fig. 10 by direct simulation, where $B_{tot} = C_1$.

If we then consider the left and right plot at time 65,000 as the equilibrium, we can plot the values of B_{tot} vs the concentration of species *A*. The dose–response diagram obtained by direct simulation depicted in Fig. 11 mirrors the bifurcation diagram obtained by numerical continuation in Fig. 9, thus cross-validating both approaches.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-021-04477-x>.

Additional file 1. Deriving a Conservation Law Matrix (B) with Non-Negative Integer Values.

Acknowledgements

We would like to thank Matthew E. Monroe for supporting us with computational resources and infrastructure and Alexandre M. Tartakovsky for helpful discussions.

Authors' contributions

BR and VP developed the methodology. BR developed the code. IOM consulted on the development of the method. All authors read and approved the final manuscript.

Funding

This work was supported by the "Capturing Pathway Activity and Linking Outcomes through Multi-omic Data Integration" project (PI. Jason McDermott) under the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory.

Availability of data and materials

https://github.com/PNNL-Comp-Mass-Spec/CRNT4SBML/tree/master/2021_BMC_Bioinformatics_paper_code

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Advanced Computing, Math, and Data Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA.

²BioProcess Engineering Group, IIM-CSIC (Spanish National Research Council), 36208 Vigo, Spain. ³Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA.

Received: 5 November 2020 Accepted: 11 November 2021

Published online: 04 January 2022

References

1. Massoner P, Ladurner-Rennau M, Eder IE, Klocker H. Insulin-like growth factors and insulin control a multifunctional signalling network of significant importance in cancer. *Br J Cancer*. 2010;103:1479.
2. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol*. 2001;2(2):127–37. <https://doi.org/10.1038/35052073>.
3. Fischer OM, Hart S, Gschwind A, Ullrich A. EGFR signal transactivation in cancer cells. *Biochem Soc Trans*. 2003;31(6):1203–8. <https://doi.org/10.1042/bst0311203>.

4. Atay O, Doncic A, Skotheim JM. Switch-like transitions insulate network motifs to modularize biological networks. *Cell Syst*. 2016;3(2):121–32. <https://doi.org/10.1016/j.cels.2016.06.010>.
5. Schoeberl B, Eichler-Jonsson C, Gilles ED, Müller G. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nat Biotechnol*. 2002;20(4):370–5.
6. Klipp E, Liebermeister W. Mathematical modeling of intracellular signaling pathways. *BMC Neurosci*. 2006;7(1):10.
7. Gilbert D, Heiner M, Breitling R, Orton R. In: Seger R, editor. *Computational modelling of kinase signalling cascades*. Totowa: Humana Press; 2010. p. 369–384.
8. Gunawardena J. Time-scale separation—Michaelis and Menten's old idea, still bearing fruit. *FEBS J*. 2014;281(2):473–88. <https://doi.org/10.1111/febs.12532>.
9. 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(10):2229–68. [https://doi.org/10.1016/0009-2509\(87\)80099-4](https://doi.org/10.1016/0009-2509(87)80099-4).
10. Horn F, Jackson R. General mass action kinetics. *Arch Ration Mech Anal*. 1972;47(2):81–116. <https://doi.org/10.1007/BF00251225>.
11. Ellison PR. The advanced deficiency algorithm and its applications to mechanism discrimination. PhD thesis, The University of Rochester; 1998.
12. Feinberg M. Chemical reaction network structure and the stability of complex isothermal reactors ii. Multiple steady states for networks of deficiency one. *Chem Eng Sci*. 1988;43(1):1–25.
13. Arkun Y. Detection of biological switches using the method of Gröbner bases. *BMC Bioinform*. 2019;20(1):615.
14. Craciun G, Feinberg M. Multiple equilibria in complex chemical reaction networks: I. The injectivity property. *SIAM J Appl Math*. 2005;65(5):1526–46.
15. Shinar G, Feinberg M. Concordant chemical reaction networks and the species-reaction graph. *Mathe Biosci*. 2012;241. <https://doi.org/10.1016/j.mbs.2012.08.002>.
16. Martín J, González M. Revealing regions of multiple steady states in heterogeneous catalytic chemical reaction networks using Gröbner basis. *J Symb Comput*. 2017;80:521–37. <https://doi.org/10.1016/j.jsc.2016.07.024>.
17. Bradford R, Davenport JH, England M, Errami H, Gerdt V, Grigoriev D, Hoyt C, Košta M, Radulescu O, Sturm T, Weber A. A case study on the parametric occurrence of multiple steady states. In: *Proceedings of the 2017 ACM on international symposium on symbolic and algebraic computation. ISSAC'17*, p. 45–52. Association for Computing Machinery, New York, NY, USA; 2017. <https://doi.org/10.1145/3087604.3087622>.
18. Conradi C, Flockerzi D. Switching in mass action networks based on linear inequalities. *SIAM J Appl Dyn Syst*. 2012;11:110–34.
19. Feliu E, Wiuf C. A computational method to preclude multistationarity in networks of interacting species. *Bioinformatics*. 2013;29(18):2327–34. <https://doi.org/10.1093/bioinformatics/btt400>.
20. Conradi C, Feliu E, Mincheva M, Wiuf C. Identifying parameter regions for multistationarity. *PLoS Comput Biol*. 2017;13(10):1–25. <https://doi.org/10.1371/journal.pcbi.1005751>.
21. Feliu E, Sadeghimanesh A. Kac–Rice formulas and the number of solutions of parametrized systems of polynomial equations. 2020. [arXiv:2010.00804](https://arxiv.org/abs/2010.00804).
22. Otero-Muras I, Yordanov P, Stelling J. Chemical reaction network theory elucidates sources of multistability in interferon signaling. *PLoS Comput Biol*. 2017;13(4).
23. Egea JA, Henriques D, Cokelaer T, Villaverde AF, MacNamara A, Danciu D-P, Banga JR, Saez Rodriguez J. Benchmarking optimization methods for parameter estimation in large kinetic models. *Bioinformatics*. 2019;35(5):830–8.
24. Otero-Muras I, Banga JR, Alonso AA. Characterizing multistationarity regimes in biochemical reaction networks. *PLoS ONE*. 2012;7(7):39194. <https://doi.org/10.1371/journal.pone.0039194>.
25. Keating SM, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, Bergmann FT, Finney A, Gillespie CS, Helikar T, Hoops S, Malik-Sheriff RS, Moodie SL, Moraru II, Myers CJ, Naldi A, Olivier BG, Sahle S, Schaff JC, Smith LP, Swat MJ, Thieffry D, Watanabe L, Wilkinson DJ, Blinov ML, Begley K, Faeder JR, Gómez HF, Hamm TM, Inagaki Y, Liebermeister W, Lister AL, Lucio D, Mjolsness E, Proctor CJ, Raman K, Rodriguez N, Shaffer CA, Shapiro BE, Stelling J, Swainston N, Tanimura N, Wagner J, Meier-Schellersheim M, Sauro HM, Palsson B, Bolouri H, Kitano H, Funahashi A, Hermjakob H, Doyle JC, Hucka M, members SLC. Sbml level 3: an extensible format for the exchange and reuse of biological models. *Mol Syst Biol*. 2020;16(8):9110. <https://doi.org/10.15252/msb.20199110>.
26. Feinberg M. Lectures on chemical reaction networks. Notes of lectures given at the Mathematics Research Center, University of Wisconsin; 1979. <https://crnt.osu.edu/LecturesOnReactionNetworks>.
27. Seydel R. *Practical bifurcation and stability analysis. Interdisciplinary applied mathematics*. New York: Springer; 2009.
28. Chicone C. *Ordinary differential equations with applications. Texts in applied mathematics*. New York: Springer; 2006.
29. Feng S, Sáez M, Wiuf C, Feliu E, Soyer O. Core signalling motif displaying multistability through multi-state enzymes. *J. R. Soc. Interface*. 2016;13. <https://doi.org/10.1098/rsif.2016.0524>.
30. ...Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villéger A, Boyd SE, Calzone L, Courtot M, Dogrusoz U, Freeman TC, Funahashi A, Ghosh S, Jouraku A, Kim S, Kolpakov F, Luna A, Sahle S, Schmidt E, Watterson S, Wu G, Goryanin I, Kell DB, Sander C, Sauro H, Snoep JL, Kohn K, Kitano H. The systems biology graphical notation. *Nat Biotechnol*. 2009;27(8):735–41. <https://doi.org/10.1038/nbt.1558>.
31. Funahashi A, Matsuoka Y, Jouraku A, Morohashi M, Kikuchi N, Kitano H. CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Proc Inst Radio Eng*. 2008;96(8):1254–65. <https://doi.org/10.1109/JPROC.2008.925458>.
32. Reyes BC, Otero-Muras I, Shuen MT, Tartakovsky AM, Petyuk VA. CRNT4SBML: a Python package for the detection of bistability in biochemical reaction networks. *Bioinformatics*. 2020;36:3922–4. <https://doi.org/10.1093/bioinformatics/btaa241>.
33. Edelstein BB. Biochemical model with multiple steady states and hysteresis. *J Theor Biol*. 1970;29(1):57–62. [https://doi.org/10.1016/0022-5193\(70\)90118-9](https://doi.org/10.1016/0022-5193(70)90118-9).
34. Terzer M. *Large scale methods to enumerate extreme rays and elementary modes*. Zurich: ETH; 2009.

35. Rockafellar RT. Convex analysis. Princeton landmarks in mathematics and physics. Princeton: Princeton University Press; 1970.
36. Dantzig GB. Linear programming and extensions. Princeton: Princeton University Press; 1991.
37. Schilling CH, Letscher D, Palsson BØ. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol.* 2000;203(3):229–48. <https://doi.org/10.1006/jtbi.2000.1073>.
38. Murabito E, Simeonidis E, Smallbone K, Swinton J. Capturing the essence of a metabolic network: a flux balance analysis approach. *J Theor Biol.* 2009;260(3):445–52. <https://doi.org/10.1016/j.jtbi.2009.06.013>.
39. Fleming RMT, Thiele I, Provan G, Nasheuer H. Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism. *J Theor Biol.* 2010;264(3):683–92. <https://doi.org/10.1016/j.jtbi.2010.02.044>.
40. Milo R, Jorgensen P, Moran U, Weber G, Springer M. Bionumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* 2009;38:750–3. <https://doi.org/10.1093/nar/gkp889>.
41. Xiang Y, Sun DY, Fan W, Gong XG. Generalized simulated annealing algorithm and its application to the Thomson model. *Phys Lett A.* 1997;233(3):216–20. [https://doi.org/10.1016/S0375-9601\(97\)00474-X](https://doi.org/10.1016/S0375-9601(97)00474-X).
42. Xiang Y, Gong X. Efficiency of generalized simulated annealing. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Top.* 2000;62:4473–6. <https://doi.org/10.1103/PhysRevE.62.4473>.
43. Kuznetsov YA. Elements of applied bifurcation theory. 2nd ed. New York: Springer; 1998.
44. Moore RE. Reliability in computing: the role of interval methods in scientific computing. San Diego: Academic Press Professional Inc; 1988.
45. Villaverde A.F, Frölich F, Weindl D, Hasenauer J.R. B.J. Benchmarking optimization methods for parameter estimation in large kinetic models. *Bioinformatics.* 2019;35(5):830–8.
46. Bolton HPJ, Groenwold AA, Snyman JA. The application of a unified Bayesian stopping criterion in competing parallel algorithms for global optimization. *Comput Math Appl.* 2004;48(3):549–60. <https://doi.org/10.1016/j.camwa.2003.09.030>.
47. Snyman JA, Fatti LP. A multi-start global minimization algorithm with dynamic search trajectories. *J Optim Theory Appl.* 1987;54(1):121–41. <https://doi.org/10.1007/BF00940408>.
48. Doedel EJ, Paenroth RC, Champneys AR, Fairgrieve TF. Auto 2000: continuation and bifurcation software for ordinary differential equations (with homcont). Technical report, California Institute of Technology; 1997.
49. Somogyi ET, Bouteiller J-M, Glazier JA, König M, Medley JK, Swat MH, Sauro HM. libRoadRunner: a high performance SBML simulation and analysis library. *Bioinformatics.* 2015;31(20):3315–21. <https://doi.org/10.1093/bioinformatics/btv363>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.