SUPPLEMENTARY MATERIAL 1:

Dynamic modeling and analysis of cancer cellular network motifs

Mathieu Cloutier[1] and Edwin Wang[1,2*]

1. Computational Chemistry and Bioinformatics Group, Biotechnology Research Institute, National Research Council Canada, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada
2. Center for Bioinformatics, McGill University, Montreal, Canada

This SM presents the Ordinary Differential Equations (ODE) models used to produce the simulations in the associated manuscript. The generic model for a 3 nodes network motif will be presented and the kinetic parameters for the 6 different cases presented in Figure 2 will be given in detail. Some additional details on parameter estimation are also provided as a complement to Figure 3 of the manuscript.

Bracketed references numbers in this document correspond to the references list of the article.

ODE model for a 3 nodes network motif

The generic 3 nodes network motif we consider is presented in Figure S1.

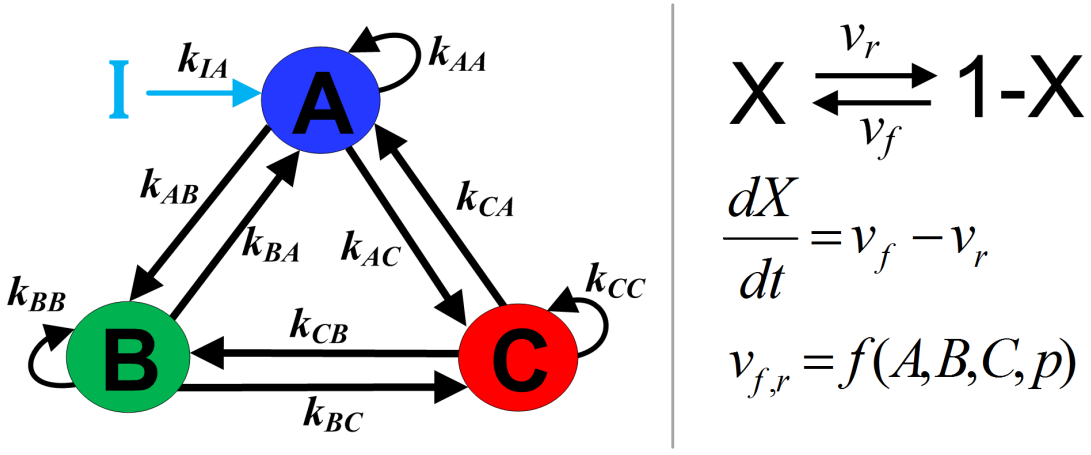

Figure S1 Generic 3 nodes network motif. Topology of possible interactions (left) and differential equation for one node (labelled X, right). The change between active (X) and inactive (1-X) form is dynamically regulated by the state of the system.

In this system, each of the nodes A, B and C will dynamically change (with values bounded between 0 and 1) depending on its input from the other nodes. The numerical value of each node thus represents its activity (i.e. as in a signalling cascade). In this framework, we neglect the degradation of the biomolecules and the total amount of active and inactive node is 1 at all times. However, by considering a reverse reaction (i.e. v_r in Figure S1) dependent on the node's activity (or concentration), we get an effect similar to degradation, whereas the activity of a node will tend to 0 in the absence of stimulation by its neighbours.

It is also important to mention that many biological networks do not have such limitations on the amount of molecule. In metabolic networks, metabolites can accumulate to very high levels. However, highly regulated metabolic systems, such as glycolysis, tend to operate in a relatively narrow range of concentrations.

Thus, our representation of the 3 nodes motifs should not be interpreted as a completely generic case of all possible network motifs in biological systems and detailed modelling of a specific case should obviously include all relevant mechanisms. However, the conclusions we can draw from the modelling of this 3 nodes system apply to many biological systems, bearing in mind the aforementioned limitations.

The differential equation for each node (X = A, B or C) will have the following form:

$$\frac{dX}{dt} = \left[k_{IX} \cdot I + k_{f,AX} \cdot A + k_{f,BX} \cdot B + k_{f,CX} \cdot C + k_{fX}\right] \cdot \frac{(1 - X)}{(1 - X) + K_{a,X}}$$

$$-\left[k_{r,AX} \cdot A + k_{r,BX} \cdot B + k_{r,CX} \cdot C + k_{rX}\right] \cdot \frac{X}{X + K_{d,X}}$$

... where k_f represents the rate constant for activation (i.e. $k_{f,AX}$ is for the activation of X by A) and k_r is the rate constant for inhibition. The parameter k_{rX} is the deactivation rate constant for X, which will push the system towards an inactive state in the absence of external stimuli (except, of course if a bistability is present), inversely we can also consider a basal activation rate (k_{fX}). The parameters $K_{a,X}$ and $K_{d,X}$ are the Michaelis-Menten constants for the activation and deactivation of X, respectively. Finally, note that, k_{IX}, the activation by input signal I, is valid only for node A.

Using this equation for nodes A, B and C, it is possible to represent all the possible topologies for the 3 nodes network motif. More specifically, each link in Figure S1 is present only if an activation and/or inhibition occur, which is in turn determined by the values of the k_f /k_r parameters. The resulting model with 3 differential equations will thus have up to 31 parameters, but in most practical cases, many interactions are absent (i.e. the k_f or k_r is 0) and the number of parameters to identify is much lower. References [14, 22, 23, 33, 34, 36, 37] of the manuscript present interesting thoughts and examples on network determination and

parameter estimation for dynamic systems. References [36, 37] present very interesting examples of parameter determination for relatively large systems using time series of experimental data. References [19, 83] show the other end of the spectrum, with the exploratory analysis of the parameter space for small 3 nodes systems.

Table S1 of this document presents the parameters we used to produce simulations in Figure 2. These networks were set-up by manually tuning the parameters in order to produce illustrative examples. However, it must be kept in mind that the parameters space can be explored extensively while retaining features such as bistability [52] or adaptation [19]. Thus, the exact parameters values for Figure 2 are not critical for our argument.

Parameters determination for a given topology and experimental data

Figure 3 of the manuscript presents a synthetic example of parameter determination from time profile data points. In that specific example, we are not addressing the issue of determining the network topology. For metabolic or signaling pathways, these topologies are relatively well known or can be determined from data [84]. For gene regulation networks, the problem of network determination is more complex, but methods are available to determine topology from data [32]. The example in Figure 3 is rather a case of finding kinetic parameters from time profile data given that we know the topology.

The data in Figure 3 was generated to represent the widely observed adaptive response in signaling [34, 81, 82]. For the normal case (red dots in Figure 3A), a curve was generated with a peak at around 10 units of time post-stimulation, followed by first-order decay towards a low value. The curve for the second case (with inhibitor, purple dots in Figure 3A) was generated in similar fashion, but with a much slower decay to reproduce the loss of adaptation. We then sample these curves and add 10% random noise.

As we are trying to model an adaptive response, only two broad classes of models will be relevant: the negative feedback and negative feedforward loops [19]. We also assume that node 'A' simply relays the input signal rapidly and so we fix k_{IA} = 1, $K_{a,A}$ = 1, $K_{d,A}$ = 1 and $k_{r,A}$ = 0.1. We thus have two sets of parameters (p) to find, depending of the motif (H1 or H2 in Figure 3). These parameters sets are defined as follows:

$$H1\text{: } p = [k_{f,AC};\ k_{f,CB};\ k_{r,BC};\ k_{r,A};\ k_{r,B};\ k_{r,C};\ K_{a,A};\ K_{d,A};K_{a,B};\ K_{d,B};K_{a,C};\ K_{d,C};]$$

$$H2\text{: } p = [k_{f,AC};\ k_{f,AB};\ k_{r,BC};\ k_{r,A};\ k_{r,B};\ k_{r,C};\ K_{a,A};\ K_{d,A};K_{a,B};\ K_{d,B};K_{a,C};\ K_{d,C};]$$

...where H1 is the set of parameters for the negative feedback motif and H2 is for the feedforward loop.

The problem of parameters determination is then formulated as an optimization problem. More specifically, the objective is to find the set of parameters (p) that minimizes the sum of squared errors between simulations and data points, defined as follows:

$$\sum_{\mathrm{i}=1}^{n}\left[\mathrm{y}_{\mathrm{i}}^{\mathrm{SIM}}-\mathrm{y}_{\mathrm{i}}^{\mathrm{MES}}\right]^{2}$$

...where y^{SIM} and y^{MES} are the simulated and measured values for the i=1..n data points in Figure 3A. Thus, the set of parameters that will minimize this error will provide the best 'fit' between model and data.

Many computational tools are available to solve this kind optimization problem and here we used common algorithms and techniques, as described below. For relatively simple systems, such as the one presented here, and given that enough data points are available, it is often possible to find a unique set of parameters that will minimize the sum of squared residuals and this improves our confidence in the model. However, the issues of parameters estimation, model validation and verification of predictive capacity constitute a very broad topic which is outside the scope of this work. The interested reader is referred to references [22, 29, 30, 36, 37, 58, 59] of the main text for further details and examples on this topic.

Finally, note that the models (H1 and H2 in Figure 3) were calibrated for the normal case (red dots in Figure 3) and the presence of an inhibitor (purple dots in Figure 3) was simulated by reducing the activation of node B through an increase in parameter $K_{a,B}$ (Michaelis-Menten constant for activation of node B).

Numerical methods

The ODEs models of the network motifs were implemented in the Matlab computing environment (The MathWorks Inc., Natick, MA, USA) using the Systems Biology toolbox [60].

The following algorithms, included in the Systems Biology toolbox [60] were used for parameters optimization:

- Simulated annealing algorithm (global optimization)
- Simplex algorithm (local optimization)

Multiple tests were performed with different initial sets of parameters and convergence to similar optima (see parameters sets in Table S1) was generally observed for H1 (negative feedback motif) and H2 (feedforward motif).

Table S1 Parameters of the 6 network motifs presented in Figures 2 and 3

	Network motifs							
Parameter	Fig. 2A (1)	Fig. 2A (2)	Fig. 2B (1)	Fig. 2B (2)*	Fig. 2C (1)	Fig. 2C (2)	Fig. 3 (H1)	Fig. 3 (H2)
k_{IA}	1	1	1	1	2	1	1	1
$k_{f,AA}$	0	0	0	0	0	0	0	0
$k_{f,BA}$	0	0	0	0	0	0	0	0
$k_{f,CA}$	0	0	0	0	0	0	0	0
$k_{r,AA}$	0	0	0	0	0	0	0	0
$k_{r,BA}$	0	0	0	0	0	10	0	0
$k_{r,CA}$	0	0	0	0	0	0	0	0
$k_{f,A}$	0	0	0	0	0	0	0	0
$k_{r,A}$	0.1	0.5	0.5	5	0.2	0.1	0.1	0.1
$K_{a,A}$	1	1	1	5	1	0.1	1	1
$K_{d,A}$	1	1	1	0.05	1	0.1	1	1
$k_{f,AB}$	0	4	0	0	1**	0	0	0.05
$k_{f,BB}$	0	0	0	0	0	0	0	0
$k_{f,CB}$	0.2	0	1	0	0	0.1	0.052	0
$k_{r,AB}$	0	0	0	0	0	0	0	0
$k_{r,BB}$	0	0	0	0	0	0	0	0
$k_{r,CB}$	0	0	0	5	0	0	0	0
$k_{f,B}$	0	0	0	0.21	0	0	0	0
$k_{r,B}$	0.01	2	0.1	0	0.1	0.05	0.02	0.16
$K_{a,B}$	2.5	1	1	1	1	0.1	4.8	4.8
$K_{d,B}$	1	1	1	0.25	5	0.1	2.1***	2.1***
$k_{f,AC}$	1	0.75	1	1	0	0.25	0.74	0.74
$k_{f,BC}$	0	0	1	0	5**	0	0	0
$k_{f,CC}$	0	0	0	0	0	0	0	0
$k_{r,AC}$	0	0	0	0	0	0	0	0
$k_{r,BC}$	1	2	0	5	0	0	0.91	0.93
$k_{r,CC}$	0	0	0	0	0	0	0	0
$k_{f,C}$	0	0	0	0	0	0	0	0
$k_{r,C}$	0.05	0.05	0.1	0.25	1	0.1	0.05	0.05
$K_{a,C}$	0.1	1	1	0.5	0.1	0.1	0.02	0.02
$K_{d,C}$	1	1	1	0.25	0.1	0.1	1.07***	1.07***

*In all simulations, the initial values of A, B and C is 0, except for Fig. 2B(2) where the initial values are A = 0.5, B = 0.6 and C = 0.07.

**For this simulation, the sharp transition is achieved using the Hill kinetic equation with exponent 10. We thus replace '$k_{f,AB}\cdot A$' by '$k_{f,AB}\cdot(A^{10}/(0.1^{10}+A^{10}))$' and '$k_{f,BC}\cdot B$' by '$k_{f,BC}\cdot(B^{10}/(0.8^{10}+B^{10}))$'

***In the case an inhibitor of B/C is applied, the Michaelis-Menten constant is increased by a factor of ≈10 to reduce the activation of the corresponding node.