

RESEARCH ARTICLE

McSNAC: A software to approximate first-order signaling networks from mass cytometry data

Darren Wethington^{1,2,*}, Sayak Mukherjee⁷, Jayajit Das^{1,2,3,4,5,6,*}

¹ Steve and Cindy Rasmussen Institute for Genomic Medicine, Abigail Wexner Research Institute, Nationwide Children's Hospital, Columbus, Ohio 43205, United States

² Biomedical Sciences Graduate Program, College of Medicine, The Ohio State University, Columbus, Ohio 43210, United States

³ Department of Pediatrics, College of Medicine, The Ohio State University, Columbus, Ohio 43210, United States

⁴ Pelotonia Institute for Immuno-Oncology, College of Medicine, The Ohio State University, Columbus, Ohio 43210, United States

⁵ Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, Ohio 43210, United States

⁶ Biophysics Graduate Program, The Ohio State University, Columbus, Ohio 43210, United States

⁷ Battelle Memorial Institute, Columbus, Ohio 43201, United States

* Correspondence: darren.wethington@nationwidechildrens.org; jayajit@gmail.com

Received July 15, 2022; Revised July 27, 2022; Accepted July 27, 2022

Background: Mass cytometry (CyTOF) gives unprecedented opportunity to simultaneously measure up to 40 proteins in single cells, with a theoretical potential to reach 100 proteins. This high-dimensional single-cell information can be very useful in dissecting mechanisms of cellular activity. In particular, measuring abundances of signaling proteins like phospho-proteins can provide detailed information on the dynamics of single-cell signaling processes. However, computational analysis is required to reconstruct such networks with a mechanistic model.

Methods: We propose our Mass cytometry Signaling Network Analysis Code (McSNAC), a new software capable of reconstructing signaling networks and estimating their kinetic parameters from CyTOF data. McSNAC approximates signaling networks as a network of first-order reactions between proteins. This assumption often breaks down as signaling reactions can involve binding and unbinding, enzymatic reactions, and other nonlinear constructions. Furthermore, McSNAC may be limited to approximating indirect interactions between protein species, as cytometry experiments are only able to assay a small fraction of protein species involved in signaling.

Results: We carry out a series of *in silico* experiments here to show (1) McSNAC is capable of accurately estimating the ground-truth model in a scalable manner when given data originating from a first-order system; (2) McSNAC is capable of qualitatively predicting outcomes to perturbations of species abundances in simple second-order reaction models and in a complex *in silico* nonlinear signaling network in which some proteins are unmeasured.

Conclusions: These findings demonstrate that McSNAC can be a valuable screening tool for generating models of signaling networks from time-stamped CyTOF data.

Keywords: single-cell; CyTOF data; signaling network; kinetics; ODE; McSNAC

Author summary: Modeling how cells transfer a signal from an extracellular stimulus to different compartments within the cell is critical to understand how different stimuli result in different cellular responses. We designed a software (McSNAC) that proposes a minimal mathematical model to describe biochemical signaling events learning from time-stamped cytometry data. This software makes linear approximations which hold true in many scenarios, but can also break down under several conditions. We explore this approximation in depth to provide guidelines regarding its applicability. In addition, we provide a user interface for non-technical users to analyze their time-stamped cytometry data with this approach.

INTRODUCTION

Single cells respond to stimulus *via* receptors that are usually bound to the plasma membrane. These receptors bind to cognate ligands generated by a stimulation, such as viral infection of cells in the local environment, and initiate a series of biochemical signaling reactions leading to activation of many genes that can generate a variety of cell responses such as secretion of specific proteins (*e.g.*, cytokines), cell proliferation, or cell death [1–3]. Signaling reactions are composed of biochemical reactions involving a large number of proteins (~ thousands), and it is often challenging to determine key signaling regulators of specific cell responses among a large set of interacting proteins [4,5]. This problem is further confounded as single cells can contain large cell-cell variations of protein abundances where protein abundances are usually distributed broadly (*e.g.*, lognormally) in a cell population, even in a clonal population [6]. Since biochemical reaction propensities depend on concentrations of participating molecules, and biochemical reactions are intrinsically stochastic in nature due to thermal fluctuations [7], single cells in clonal cell populations stimulated by the same ligand can show large variations in responses making it difficult to determine key regulators in the face of large variations. Recent advances in single cell cytometry, such as Mass Cytometry, or Cytometry by Time-of-Flight (CyTOF), are currently capable of measuring up to 40 proteins simultaneously in thousands to millions of single cells, with a theoretical limit as high as 100 proteins [8,9]. Thus, measuring signaling protein abundances at different points post stimulation using CyTOF provides time-stamped snapshot data regarding signaling kinetics and cell-cell variation of the kinetics. However, it is difficult to intuitively determine key regulators that give rise to a specific response in a single cell or a group of single cells with this data alone for the following reasons. First, though causal relationships between measured proteins, such as phosphorylation of protein A inducing phosphorylation of protein B, might be known from previous experiments, activation of key regulator proteins can often be produced by multiple proteins interacting synergistically. Second, protein abundances vary widely across cells, and proteins in the same individual cell are not measured across time, thus if specific proteins are activated during signaling kinetics in individual cells remains unknown.

Some of the above challenges are also present in deciphering mechanisms underlying signaling kinetics using population data or traditional flow cytometry data measuring few proteins. Mechanistic mathematical models composed of ordinary differential equations (ODEs) or stochastic models describing biochemical

signaling reactions have been used to address these difficulties [10,11]. These models work well when the number of protein species is small and biochemical reaction propensities for reacting proteins are well characterized. However, it is often not the case for newly discovered cell types (*e.g.*, myeloid derived suppressor cells) [12] or situations where multiple types of receptors (*e.g.*, cytokine receptors and primary cell receptor) synergize [13]. It is challenging to build such mechanistic models in such situations where protein species interacting physically are usually not measured simultaneously and it is unclear what form (*e.g.*, first, second or higher order reaction) of reaction propensities should be used for describing the interaction between two protein species separated by many intermediate reactions.

Machine learning has also been employed to characterize ODEs to predict the effects of cellular perturbations [14]. This model is quite useful for mechanistically predicting the effects of perturbations, but does not have a mechanism to incorporate the single-cell resolution offered by CyTOF. It is important to note that signaling processes happen at the single-cell resolution, and thus single-cell experiments contain crucial information on signaling dynamics. This information can be used to identify how cells vary in their protein expressions, and which proteins co-vary with each other in single cells. For these reasons, it is crucial to have a method to form accurate mechanistic models for single-cell signaling networks.

To this end, we developed a first-order mass action reaction kinetics model given by linear ODEs to describe signaling kinetics over a time interval — *e.g.*, time between two successive mass cytometry measurements [13,15]. The solution of the model describes the time evolution in a closed form and accounts for cell-cell variations of protein abundances. In this paper we report development of a software package Mass cytometry Signaling Network Analysis Code (McSNAC) that takes flow cytometry data at two successive time points as input, fits the model involving first-order kinetics, and generates parameter estimates for the kinetic rates. The model can then be used for generating predictions regarding changes in protein abundances. We use McSNAC to model synthetic cytometry data to demonstrate: (1) Parameters in the model can be well estimated. (2) Models that include the ground-truth model describing first-order reaction kinetics provide better fits than models without the ground-truth. (3) The first-order models in McSNAC are able to make accurate qualitative predictions to perturbations in protein abundances. (4) The models in McSNAC generate better predictions to perturbations when directionality of the underlying reaction is captured in the model.

Model

McSNAC describes kinetics of single cell protein abundances using first-order chemical reaction kinetics. Consider p number of protein species $\{X_1, \dots, X_p\}$ measured in cytometry experiments that interact during a signaling process. In McSNAC the abundance $x_i^{(\alpha)}$ of protein X_i in each cell indexed by α changes with time given by the linear ODEs shown below:

$$\frac{dx_i^{(\alpha)}}{dt} = k_{i \rightarrow i} x_i^{(\alpha)} - \sum_{j=1}^p k_{i \rightarrow j} x_i^{(\alpha)} + \sum_{j=1}^p k_{j \rightarrow i} x_j^{(\alpha)} \quad (1)$$

where the first-order rate that produces species X_j from X_i is given by $k_{i \rightarrow j}$ ($k_{i \rightarrow j} \geq 0$ for $i \neq j$). $k_{i \rightarrow i} > 0$ denotes self-production or production of X_i from another unmeasured species while $k_{i \rightarrow i} < 0$ denotes self-decay. In the above kinetics, the kinetic rates $\{k_{i \rightarrow j}\}$ are the same across single cells (*i.e.*, a $k_{i \rightarrow j}$ does not depend on the cell index α). This is an approximation as some of these rates could represent nonlinear reaction such as enzymatic modifications where the rates could depend on the abundances of an enzyme which could vary from cell-to-cell. The ODEs in Eq. (1) can be represented in compact form by introducing a $p \times p$ matrix, \mathbf{M} , where $M_{ij}(i \neq j) = k_{j \rightarrow i}$, and $M_{ii} = k_{i \rightarrow i} - \sum_j k_{i \rightarrow j}$, *i.e.*,

$$\frac{dx_i^{(\alpha)}}{dt} = \sum_{j=1}^p M_{ij} x_j^{(\alpha)} \quad (2)$$

This representation of the mass-action kinetics offers a convenient closed-form solution (details in Materials and Methods section) for $\{x_i^{(\alpha)}(t_2)\}$ at given $\{x_i^{(\alpha)}(t_1)\}$, *i.e.*, $x_i^{(\alpha)}(t_2) = \sum_{j=1}^p [e^{\mathbf{M}(t_2-t_1)}]_{ij} x_j^{(\alpha)}(t_1)$.

Mass cytometry experiments are unable to track single cells over time as individual cells are destroyed at the time of measurement. However, it is still possible to follow the kinetics of average values, covariances and higher order moments of the protein abundances that are computed from the snapshot data.

$$\mu_i(t) = \frac{1}{N} \sum_{\alpha=1}^N x_i^{(\alpha)}(t) \quad (3)$$

$$J_{ij}(t) = \frac{1}{N} \sum_{\alpha=1}^N (x_i^{(\alpha)}(t) - \mu_i(t))(x_j^{(\alpha)}(t) - \mu_j(t)) \quad (4)$$

The ODEs in Eq. (2) can be solved to relate the average values $\{\mu_i\}$ and covariances $\{J_{ij}\}$ calculated at times t_1 and t_2 ,

$$\mu_i(t_2) = \sum_{j=1}^p [e^{\mathbf{M}(t_2-t_1)}]_{ij} \mu_j(t_1) \quad (5)$$

$$\mathbf{J}(t_2) = e^{\mathbf{M}(t_2-t_1)} \mathbf{J}(t_1) e^{\mathbf{M}^T(t_2-t_1)} \quad (6)$$

The bold form symbols in the above equations denote matrices. These above closed-form solutions make it efficient to compute predicted average values and covariances in the model compared to iteratively solving the ODEs in Eq. (2) numerically and then computing those variables using Eqs. (3) and (4). Matrix exponentials are calculated with the Padé approximation, and require significantly less computational time than standard numerical ODE solution techniques such as Runge-Kutta methods. Given the measured values of $\{\mu_i\}$ and $\{J_{ij}\}$ at time t_1 from cytometry experiments, we estimated the kinetic rates in \mathbf{M} that best fit cytometry data for $\{\mu_i\}$ and $\{J_{ij}\}$ at a later time t_2 . \mathbf{M} is estimated by minimizing a cost function (Eq. (7)) by employing a simulated annealing algorithm. The cost function is defined as,

$$\chi^2 = \sum_i \left(1 - \frac{\mu(t_2)_i^{(\text{predicted})}}{\mu(t_2)_i^{(\text{actual})}} \right)^2 + \sum_{i,j} \left(1 - \frac{J(t_2)_{ij}^{(\text{predicted})}}{J(t_2)_{ij}^{(\text{actual})}} \right)^2 + \sum_j (t_2 - t_1)^2 \left(\sum_i M_{ij} \right)^2 \quad (7)$$

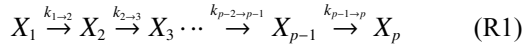
Where, $\mu(t_2)^{(\text{actual})}$ and $\mathbf{J}(t_2)^{(\text{actual})}$ are the means and covariances calculated from cytometry data at time t_2 using Eqs. (3) and (4), and, $\mu(t_2)^{(\text{predicted})}$ and $\mathbf{J}(t_2)^{(\text{predicted})}$ are the means and covariances predicted by the model following Eqs. (5) and (6) given $\mu(t_1)$ and $\mathbf{J}(t_1)$ calculated from cytometry data. This cost function has three terms, penalizing different aspects of the fit. The first term penalizes deviations from the observed averages. The second term penalizes deviations from the observed variances and covariances. The third term penalizes kinetics that do not abide by conservation of mass (assuming phospho-groups are transferred and no other changes in abundance occur). In other words, the third term penalizes $k_{i \rightarrow i} \neq 0$ from Eq. (1). If data come from a ground-truth first-order model which abides by conservation of mass, and the ground-truth \mathbf{M} is found, then $\chi^2 = 0$. A more detailed description of this method can be found in Mukherjee *et al.* 2017 [13].

RESULTS

McSNAC is capable of accurately estimating first-order reaction rates

Here we describe McSNAC's ability to (1) model data when the wiring of the ground-truth model is known but the values of the rate constants are unknown; (2) computationally scale as the number of proteins in the data is increased up to 40; and (3) separate models that do or do not include the ground-truth. The synthetic

cytometry data were generated for a coupled set of first-order reactions as shown in Reaction (R1) using Eq. (2), where the distributions of $\{x_1^{(a)}, \dots, x_p^{(a)}\}$ at $t = 0$ are chosen from log-normal distributions (see Materials and Methods for details).



The rates $\{k_{i \rightarrow j}\}$ were assigned from a uniform random distribution with a range of 0 to 1, and $\{x_1^{(a)}, \dots, x_p^{(a)}\}$ were computed for 2,500 single cells at two different time points. This constituted the synthetic cytometry data which were generated for a range of p between 8 to 40. We applied McSNAC to synthetic data following the architecture in (R1) to determine (1) how accurately parameters $\{k_{i \rightarrow j}\}$ are estimated; (2) how does the run time increase with increasing p . The comparison of the

McSNAC estimated parameters and their actual values shows high accuracy in parameter estimation (Fig. 1A–B). The run time of McSNAC on a single CPU (3.0GHz AMD 7302, compiled with *gfortran* with the *-O* optimization flag) ranged from minutes to hours where the case for $p=40$ took 4.8 hours to complete on average (Fig. 1C). Theoretical computational complexity is $O(n^3)$, and an empirical fit to the runtimes indicates that this is indeed observed. The complexity is primarily due to matrix multiplication, which means that improvements to runtime are possible with more advanced computational techniques. These results indicate that this technique is scalable with the large number of inputs afforded by CyTOF. It is important to note that exact runtimes vary depending on hyperparameter specification, such as number of temperature

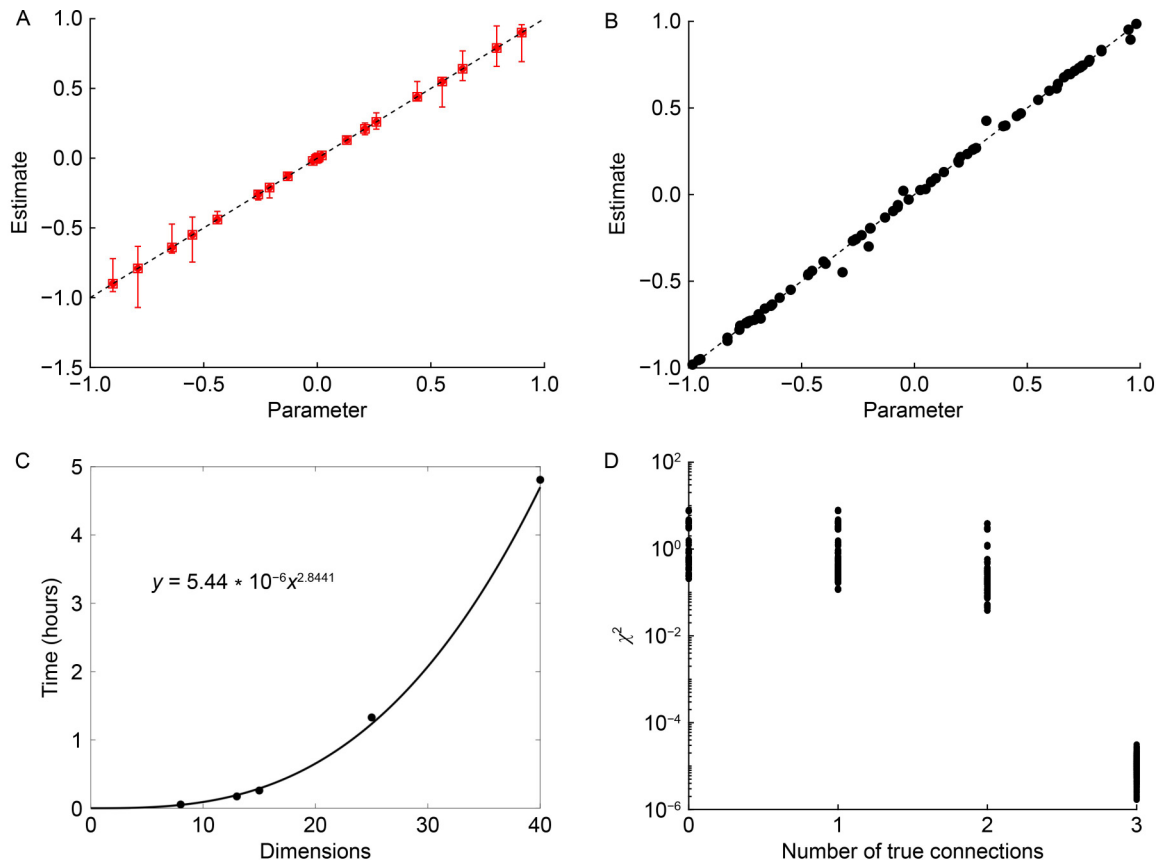


Figure 1. McSNAC is a scalable software capable of reconstructing first-order signaling dynamics. (A) Comparison of manually selected parameters of a 13-dimensional system and the estimated parameter values from McSNAC. Error bars denote confidence intervals on the parameter estimates. The dotted black line represents $y = x$, and perfectly estimated parameters will lie on this line. (B) Comparison of manually selected parameters of randomly selected parameters of a 40-dimensional system and the estimated value of each parameter by McSNAC. The dotted black line represents $y = x$, and perfectly estimated parameters will lie on this line. (C) The average time ($n = 3$) required to run McSNAC for the number of dimensions p shown for a system represented by Eq. (R1). (D) 2^{12} models were simulated, each one containing either none (0), some (1-2), or all (3) of the $k_{i \rightarrow j}$ connections in the ground-truth model. Models with all (3) of the true connections, comprising the ground-truth model, generate $\chi^2 \approx 0$, indicating a perfect fit.

cooling iterations and number of Monte Carlo samples per temperature step, and the exact data and model being fitted.

Next, we evaluated the ability of the cost function (Eq. (7)) to separate models that include or do not include the architecture of the ground-truth model. This is relevant for cases where interactions between protein species in a signaling network are not known and several potential network architectures can be hypothesized to describe the data. To test the above property, we generated synthetic cytometry data for a reaction scheme involving four protein species given by $p = 4$ in (R1). If we had no knowledge of the architecture of the ground-truth network that connected the proteins, these four proteins would offer a total of 12 possible $k_{i \rightarrow j}$'s, and 2^{12} unique network architectures (more details in Materials and Methods section). Applying McSNAC to each of the 2^{12} unique network architectures and determining the goodness-of-fit χ^2 shows that when a model contained the ground-truth architecture it generated close to zero values of χ^2 , whereas models that do not include the ground-truth architecture generated larger values of χ^2 (Fig. 1D). These results show that McSNAC is capable of reconstructing the ground-truth signaling dynamics from data resulting from a first-order system, regardless of dimensionality or *a priori* information of its wiring.

In the next sections we show results when McSNAC is applied for data sets that represent scenarios that are commonly present in cytometry datasets studying signaling kinetics. Signaling biochemical reactions contain reactions associated with propensities that are nonlinear functions of abundances of reacting protein species. Moreover, a large number ($\sim 1,000$ – $10,000$) of proteins and their chemically modified forms can be involved in signaling reactions, and CyTOF experiments are able to measure a small fraction of those protein species. Therefore, we tested the ability of the first-order reaction kinetics in McSNAC to approximate the signaling reaction kinetics in the above situations. The following cases are considered: (1) The ground-truth

model contains non-linear biochemical reactions, (2) Not all proteins participating in biochemical signaling reactions are measured. We find that McSNAC is able to describe the result of perturbations such as increase and decrease of specific protein abundances in most situations, however, McSNAC performs poorly when it comes to predicting protein abundances at a later time (Supplementary Fig. S1). In the next sections we provide further details regarding McSNAC's ability to predict outcomes of perturbations of signaling kinetics due to increase and decrease of specific protein abundances. Such perturbations are routinely performed using specific drugs such as Src family kinase inhibitor, Syk family kinase inhibitors, or small interfering RNA (siRNA).

McSNAC is capable of approximating simple non-linear systems

We test the ability of McSNAC to capture kinetics in a ground-truth model describing a second-order reaction in which two reactants combine to form a product. The candidate models probed by McSNAC are composed of first-order reactions. We generated *in silico* data formed by ground-truth model GT1 given by Reaction (R2) at two different time points and fitted the candidate models (C1 to C7) shown in Table 1 to estimate reaction rates.



The McSNAC candidate models with the best fit M values describe few abundances at t_2 reasonably well, but over- or under- shot values for certain species. For example, candidate model C1 ($B \rightarrow C$) matches mean abundances of B but overestimates the mean abundance of C for model GT1 at t_2 (details in the Supplementary Material). This behavior is expected as McSNAC uses first-order reactions to approximate second-order reactions in the ground-truth model. Next, we evaluated how the candidate models in McSNAC perform for predicting outcomes of *in silico* perturbation experiments where increased abundances of species (*e.g.*, B) at

Table 1 List of candidates first-order models for describing a unidirectional second-order reaction $A + B \rightarrow C$ in ground-truth model GT1

Candidate models	Directionality included	Prediction of perturbation outcome for changing B abundance
$C1: B \rightarrow C$	Y	$\Delta\mu_B(t_1, t_2)$ and $\Delta\mu_C(t_1, t_2)$ are qualitatively accurate.
$C2: A \rightarrow C, B \rightarrow C$	Y	Same as above, however, does not capture the behavior in $\Delta\mu_A(t_1, t_2)$.
$C3: B \leftrightarrow C$	Y	Qualitatively accurate for $\Delta\mu_B(t_1, t_2)$ but not for $\Delta\mu_C(t_1, t_2)$.
$C4: A \leftrightarrow C, B \rightarrow C$	Y	Same as above
$C5: A \leftrightarrow C, B \leftrightarrow C$	Y	Same as above
$C6: C \rightarrow B$	N	Inaccurate for both $\Delta\mu_B(t_1, t_2)$ and $\Delta\mu_C(t_1, t_2)$ ($\cong 0$)
$C7: C \rightarrow A, C \rightarrow B$	N	Same as above

t_1 and assessed the ability of the candidate models to predict changes in the abundances of all the species at a later time t_2 . We defined a variable $\Delta\mu_X(t_1, t_2)$ describing changes in mean abundances of species X between two successive times t_1 and t_2 ($>t_1$) for quantifying comparisons between the ground-truth and candidate models in the perturbation experiments. $\Delta\mu_X(t_1, t_2)$ is given by

$$\Delta\mu_X(t_1, t_2) = \mu_X(t_2) - \mu_X(t_1) \quad (8)$$

$\Delta\mu_X(t_1, t_2) > 0$ (or < 0) indicates net production (or consumption) in the mean abundance of X as the kinetics progress from time t_1 to t_2 . To illustrate how $\Delta\mu_X(t_1, t_2)$ can be used to analyze outcomes of a perturbation experiment, if $\Delta\mu_C(t_1, t_2) > 0$ in the ground-truth model and also in a candidate model (e.g., C1) when abundance of B is increased at time t_1 , it implies that candidate model C1 is able to qualitatively predict the outcome of the perturbation. The quantitative difference between the ground-truth model and the prediction made by the candidate model can be given by the squared distance between $\Delta\mu_X(t_1, t_2)$ obtained from the ground-truth and a candidate model. First, we investigated results from increasing or decreasing abundances of B at time t_1 in the ground-truth model GT1 and the candidate models (Table 1, Fig. 2A and Supplementary Fig. S2). The candidate models C1 and C2 correctly predict the net production of species B and C or $\Delta\mu_B(t_1, t_2) > 0$ and $\Delta\mu_C(t_1, t_2) > 0$ as in GT1. In addition, as abundance of B is increased (or decreased) in the perturbation experiment the productions of B and C are increased (or decreased); this monotonic behavior is also correctly predicted by candidate models C1 and

C2. However, the values of $\Delta\mu_B(t_1, t_2)$ and $\Delta\mu_C(t_1, t_2)$ in the models C1 and C2 are different than that in the ground-truth model GT1 indicating that the candidate models are able to correctly predict outcomes of perturbations qualitatively but not quantitatively (Table 1 and Fig. 2A). The models C3 and C7 are unable to correctly predict the above qualitative changes correctly (Table 1 and Fig. 2A). However, none of the candidate models are able to capture the decrease (or increase) in $\Delta\mu_A(t_1, t_2)$ as abundance of B is increased (or decreased) in GT1 (Fig. 2B) as the first-order reactions are unable to capture the dependencies of consumption of A and B given by the second-order reaction in GT1. Thus, as long as the directionality of the reaction, i.e., irreversible production of C from B and A is correctly captured in the candidate models they are able to qualitatively predict outcomes of the perturbation for most of the species.

To extend the exploration of approximating simple nonlinear dynamics, we simulated data from a similar nonlinear system shown by Reaction (R3).



Because the directionality of this model depends on the relative abundance of reactants and products, we formed three sets of data: (1) reactants are more abundant than product (rightward directionality); (2) product is more abundant than reactants (leftward directionality); (3) reactants and product are approximately equal in magnitude (almost no directionality). We modeled each set of data formed by this nonlinear system with the approximations in candidate models

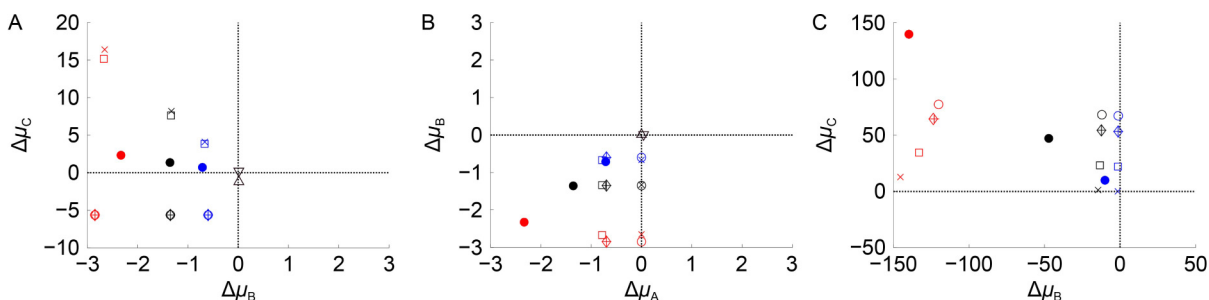


Figure 2. McSNAC is capable of predicting the effects of perturbations in simple nonlinear systems. (A) Predictions of the change in averages through time in B and C resulting from perturbations of B for the ground-truth model indicated by Reaction GT1, as predicted by the approximations in Models C1-C7. Large, filled circles represent the ground-truth, and markers for candidate predictions are as follows. C1: $B \rightarrow C$ (X's); C2: $A \rightarrow C, B \rightarrow C$ (Squares); C3: $B \rightarrow C$ (Open Circles); C4: $A \leftrightarrow C, B \rightarrow C$ (Diamonds); C5: $A \leftrightarrow C, B \leftrightarrow C$ (+s); C6: $C \rightarrow B$ (upward-facing triangles); and C7: $C \rightarrow A, C \rightarrow B$ (downward-facing triangles). Black points are estimations without a perturbation, red points are estimations in which B was perturbed upwards two-fold, and blue points are estimations in which B was perturbed downwards two-fold. $\Delta\mu_B$ and $\Delta\mu_C$ are defined as the change in averages through time, shown in Eq. (8). (B) Predictions of the change in averages through time in A and B resulting from perturbations of B for the ground-truth model indicated by Reaction GT1. Markers and perturbation information are the same as in (A). (C) Predictions of the change in averages through time resulting from perturbations of B for the ground-truth model indicated by Reaction GT2, as predicted by the approximations in Models C1-C5. Markers are the same as in (A), but perturbations of B are 10-fold.

McSNAC is capable of predicting effects of perturbations in complex non-linear signaling networks

Here we applied McSNAC to test its ability to predict effects of changes in species abundances for a signaling model describing membrane proximal signaling events in an NK cell. The signaling reaction contains multiple binding-unbinding and enzymatic reactions—all of which are composed of coupled second-order reactions. The NK cell signaling model, which is the ground-truth model here, is described in [Fig. 3A](#). The model contains

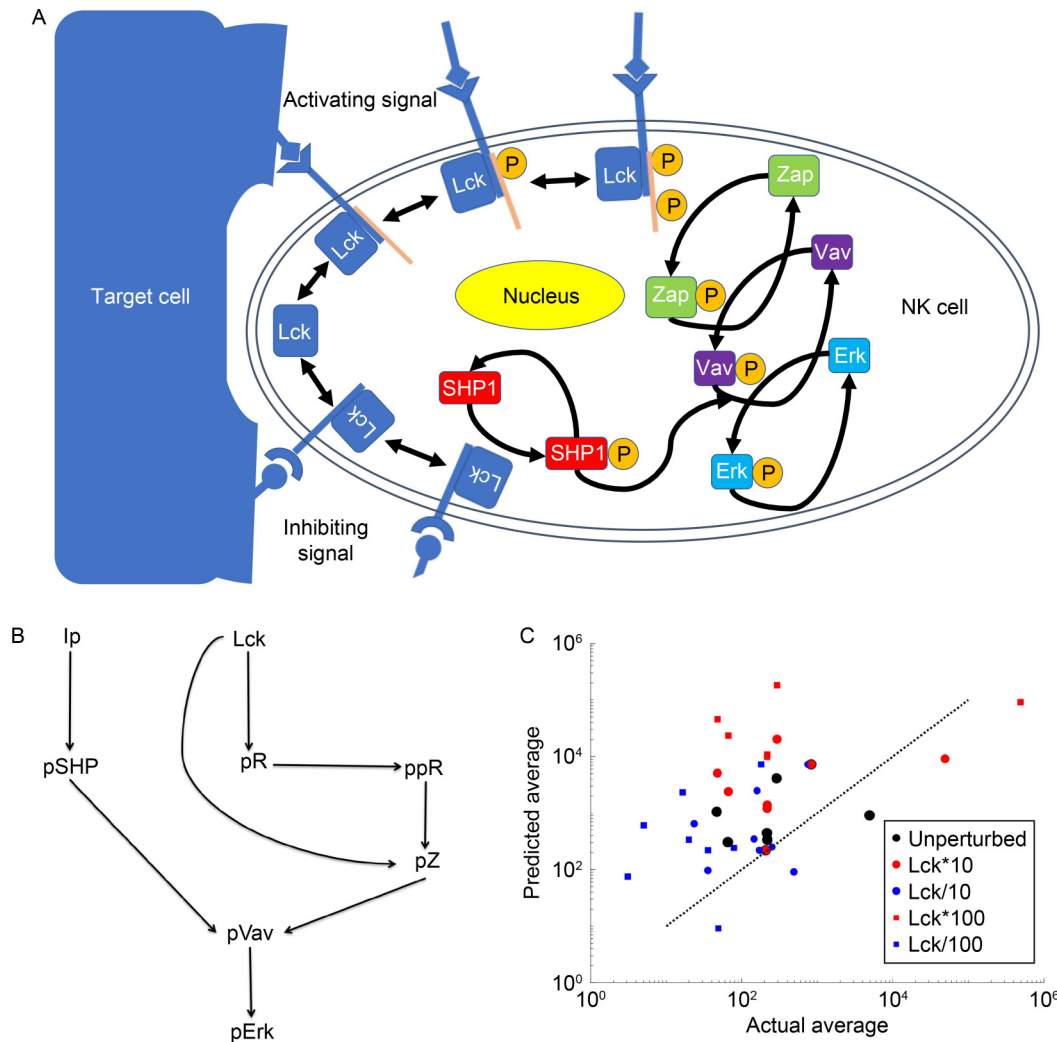


Figure 3. McSNAC is capable of predicting the effects of perturbations in complicated nonlinear systems. (A) A representation of the ground-truth system used for the *in silico* nonlinear signaling network simulation of Erk signaling in NK cells [22,23]. NK cells recognize ligands corresponding to activating and inhibiting signals on target cells. These signals result in binding of various proteins to the receptor, activation (phosphorylation) of proteins, and a cascade of phosphorylation to proteins in the cytosol of the NK cell. The proteins Erk and Vav become phosphorylated, which in turn mediate lysis of target cells. (B) A first-order network representation of (A), passed to McSNAC for estimation of kinetic parameters. (C) The results of perturbations performed on activated Lck. Actual averages of each species resulting from Lck perturbation in (A) are compared with predicted ones from the first-order approximation in (B).

18 molecule types, 112 total species (molecules and complexes of bound molecules), and 53 reactions, and is simulated using BioNetGen, a software package that specializes in simulating biological reaction networks [16,17]. We set up a candidate model, where we assumed a subset of protein species are measured in cytometry experiments, and these proteins react via first-order reactions as shown in Fig. 3B. Because this model only explores a subset of the total number of species and omits measurements such as un-phosphorylated proteins, it is reflective of a CyTOF experiment studying signaling kinetics where many proteins are unmeasured. Omission of unmeasured proteins does not impact a simpler ground-truth linear system (Supplementary Fig. S4), so we continue to this more complex nonlinear ground-truth simulation. First, we simulated kinetics of signaling protein abundances in single cells in the ground-truth model using BioNetGen over a time interval $t = 0$ s to 60 s. The details regarding rate constants and initial distributions of protein abundances are given in Materials and Methods section and the supplemental BioNetGen file. The mean abundances and covariances calculated for the protein species in the candidate model were then fitted at two time points ($t_1 = 0$ s and $t_2 = 30$ s) to estimate rates in the matrix M . The best fit value of M estimated mean abundances well for several species, however, over- and under- estimated for 4 out of 8 species (Fig. 3C). This behavior is expected as McSNAC used first-order reactions to model a ground-truth model containing nonlinear reactions. Next, we evaluated the ability of the candidate model to describe outcomes of perturbations of total abundances for specific protein species. We increased and decreased the total amount of the Src kinase Lck at time $t_1 = 0$ in the ground-truth model and the candidate model with the best fit M and compared mean abundances at a later time $t_2 = 30$ s (Fig. 3C). Lck is a kinase that regulates early time signaling events such as receptor tyrosine phosphorylation and directly and indirectly influence abundances of the species considered in the candidate model. We found that mean protein abundances at $t_2 = 30$ s in the candidate model showed a statistically significant positive correlation ($= 0.33$) with that in the ground-truth model (Fig. 3C), thus, the predictions for Lck perturbation in McSNAC are able to qualitatively capture the changes in mean abundances in the ground-truth model.

Above we constructed the candidate linear model by intuition, therefore, we investigated if such a candidate model can be constructed using a *fully connected* linear model where any pair of species interact with each other. We reasoned based on our results (Fig. 1D) for the linear model that estimated rates over a user defined threshold in the fully connected model can point to a

candidate model. We found that when data are generated from a simple nonlinear model (GT1), parameter estimations in the fully connected model resembled a linear candidate model C2 in Table 2 (Fig. 4A–B). However, for data generated from a more complex nonlinear model in Fig. 3A, reactions in fully connected network associated with larger values of estimated values showed similarities and dissimilarities with our candidate model in Fig. 3B (Fig. 4C–D). The fully connected network for our *in silico* data did correctly reconstruct many of the existing signaling steps, such as the signal from the inhibitory receptor to pSHP, from the single-phosphorylated activating receptor to the double-phosphorylated activating receptor, and from pZap70 to pVav1. However, there were some notable deviations. For instance, in the fully connected network there is a large kinetic rate for the reaction $pZap \rightarrow pErk$, with a small rate for $pVav \rightarrow pErk$, which skips a signaling reaction (*e.g.*, pVav induced phosphorylation of Erk) in the ground-truth model (Fig. 4C–D). Additionally, the inhibitory receptor is found to phosphorylate ITAMs associated with activating receptor, another step which is not realistic. These unrealistic estimations could arise due to the inability of the linear network to correctly capture co-dependencies between the species in a complex non-linear ground-truth model. Therefore, we recommend using the linear representation as a generation tool.

Graphical User Interface

McSNAC is designed with ease-of-use in mind. It can be run by dragging a file to a terminal and following on-screen prompts to enter information about the desired signaling network to model (Fig. 5). The output of the software is an interactive display which shows the flux arising from each connection (calculated by $k_{i \rightarrow j} \mu_i(t_1)$). Additionally, a .csv file is generated to save the results for later viewing. This intuitive graphical user interface (GUI) and output should make McSNAC approachable for biologists not accustomed to running computational software from the command line.

DISCUSSION

McSNAC is a scalable software package for developing signaling kinetic models based on first-order biochemical reactions using time-stamped CyTOF data. The estimation of kinetic rates in McSNAC accounts for cell-cell variations of protein abundances found in cytometry data and separates changes in protein phosphorylation generated due to tonic or basal signaling and signaling induced by receptor stimulation. Our previous work showed that the framework based on first-order

A			
	A	B	C
A	-0.01	0	0
B	0	-0.02	0
C	5E-05	0.004	0

B			
	A	B	C
A	-0.02	1E-06	1E-07
B	2E-09	-0.07	2E-10
C	0.01	0.058	-0

C								
	ppR	pR	Lck	pZap	pSHp	Ip	pVav	pErk
ppR	-0.13	0.312	0	0	0	0	0	0
pR	0	0.091	0.007	0	0	0	0	0
Lck	0	0	0.136	0	0	0	0	0
pZap	0.198	0	0.222	-0.82	0	0	0	0
pSHp	0	0	0	0	0.193	0.037	0	0
Ip	0	0	0	0	0	0.086	0	0
pVav	0	0	0	0.897	0.032	0	0.027	0
pErk	0	0	0	0	0	0	0.038	0

D								
	ppR	pR	Lck	pZap	pSHp	Ip	pVav	pErk
ppR	0.056	1E-06	3E-05	1E-06	0.008	0.575	0.052	9E-07
pR	0.192	-0.05	6E-10	8E-08	1E-05	4E-06	6E-08	2E-07
Lck	2E-08	4E-09	0.492	2E-07	8E-08	6E-07	9E-08	9E-08
pZap	3E-06	1E-06	1E-08	-0.15	0.002	7E-07	0.077	3E-06
pSHp	0.012	0.187	3E-09	3E-06	0.146	0.207	5E-07	3E-07
Ip	5E-09	9E-09	8E-04	3E-10	2E-09	-0.74	8E-10	4E-09
pVav	0.033	4E-07	6E-10	0.188	0.021	1E-05	0.089	0.146
pErk	4E-08	9E-08	2E-04	0.067	2E-08	9E-07	1E-07	-0.01

Figure 4. McSNAC displays mixed results when modeling non-linear networks without a priori network proposal. (A–B) Shows elements of the M matrix when data generated from the nonlinear network GT1 are modeled with a linear network given by (A) model C2 in Table 1, and with (B) a *fully connected* network. (C–D) Estimated elements of the M matrix when data generated from the *in silico* network in Fig. 3A are modeled with (C) the proposed linear network in Fig. 3B, and with (D) a *fully connected* network. The shades of red show the magnitude of the value of each estimated element.

reaction kinetics is able to describe synergy between cytokine treatment and receptor stimulation for CyTOF data obtained for primary human NK cells. The McSNAC software is based on this framework and implements a GUI to add features such as protein species selection and specification of a reaction of interest as desired by users.

Application of McSNAC for a set of *in silico* experiments showed the framework based on first-order reactions is able to accurately estimate rates when the proposed reaction network contains the wiring of the ground-truth model consisting of first-order reactions. Candidate networks that do not contain the wiring of the ground-truth model produce substantially large values of the cost function. Our *in silico* tests showed that when ground-truth models are nonlinear second-order biochemical reactions, McSNAC is able to generate qualitatively correct predictions for perturbations of protein abundances when directionality of reactions are correctly accounted for in McSNAC. We also found these agreements with qualitative predictions hold when a subset of signaling species is measured and modeled in McSNAC for synthetic cytometry experiments. Perturbations of cell signaling are commonly carried out to analyze signaling networks. Therefore, McSNAC can be a valuable tool for screening potential models that can be hypothesized for describing cytometry datasets.

An important limitation of the first-order approximation is its inability to capture non-monotonic behavior. In many true signaling networks, activity is differentially regulated at various time intervals. For instance, several

proteins (e.g., pAkt) display non-monotonic response—such behaviors can arise as protein degradation initiated by early signaling events can decrease abundances of phosphorylated forms of proteins. The first-order framework is unable to capture such non-monotonic temporal changes. However, one can divide the non-monotonic kinetics into time intervals where the kinetics is monotonic and use McSNAC to model each of these intervals separately. The rate parameters obtained from those piece-wise models could provide insight regarding change of specific signaling reactions with time. McSNAC is also not able to fit many species abundances well, and generate correct predictions for protein abundances at future times when the underlying model contains non-linear biochemical reactions. Conceptually, McSNAC is capable of fitting data at more than two timepoints. Incorporating more data is likely to improve parameter estimation for a ground-truth linear signaling network [18,19], but more realistic non-linear signaling networks suffer as the first-order approximation will change at varying times. Improving on these limitations will require including non-linear reactions into the framework, a task that will be computationally demanding, as matrix exponentiation must be abandoned for numerical ODE solutions.

The GUI and code base will be updated to reflect features desired by users. Some possible future directions would be to implement a method to automatically report confidence intervals on parameter estimates, implementing the ability to run on high-performance

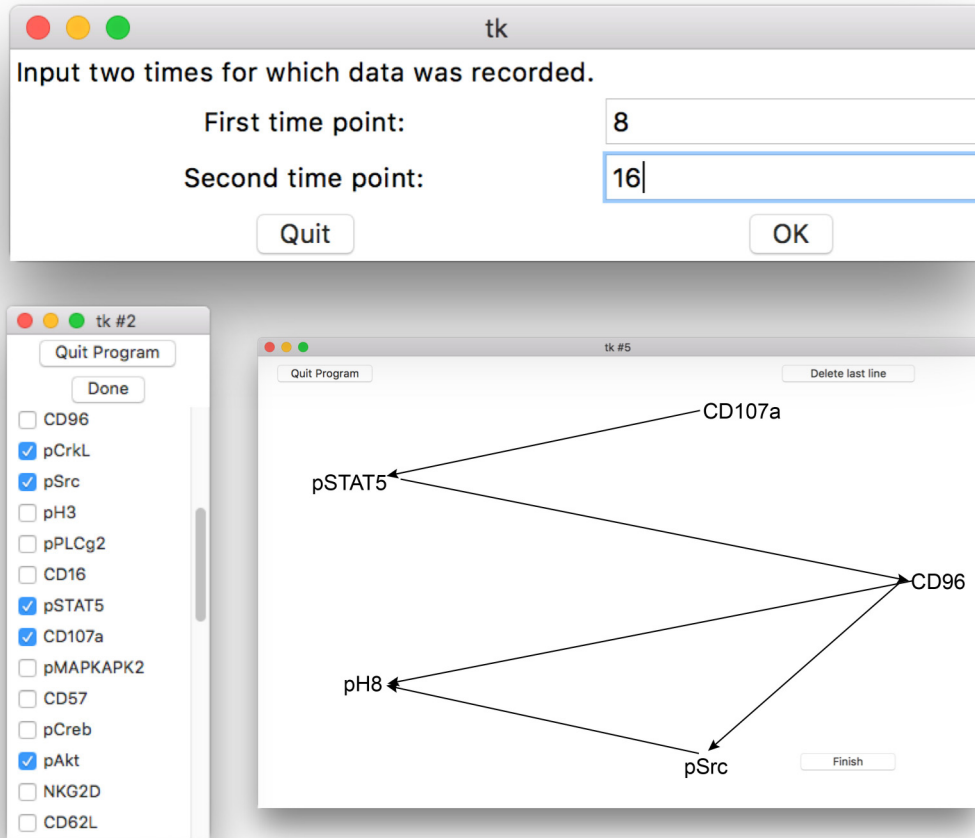


Figure 5. McSNAC is designed for ease-of-use with an intuitive graphical user interface. The software starts by having a user drag a file to the terminal and then following on-screen prompts to indicate the files they would like to analyze, the time difference between data collection, and the proteins to include in the network. Selected proteins are displayed in a circle and the user can draw arrows connecting proteins to indicate the proposed model network. Upon completion, the network is displayed to the user with flux calculations for each connection, which are displayed by hovering over the arrow of interest.

computing clusters, or adapting it to other data types other than .fcs files. The software is designed exclusively for OS X operating systems, but the recent addition of the Windows PowerShell may make it feasible to adapt it for Windows as well. The software is intended to be open-source, and we encourage users to request features they'd like to see and/or implement them themselves. The simulated annealing scheme is written in Fortran, and the GUI and parent script are written in Python. The software can be found on github (dweth/mcsnac), and code edits can be made there as well.

MATERIALS AND METHODS

Solution of Eq. (2)

In matrix notation where $x = \{x_i\}$ is a column vector, the

ODE in Eq. (2) can be rewritten as

$$\frac{dx}{dt} = Mx$$

If S denotes the similarity transformation matrix that diagonalizes M to a diagonal matrix D then,

$$D = SMS^{-1}.$$

D contains eigenvalues $\{\lambda^{(i)}\}$ of M as diagonal elements. Defining $y = Dx$, the ODE in x can be rewritten as

$$\frac{dy}{dt} = SMS^{-1}y = Dy.$$

Since D is a diagonal matrix, the above equation represents a set of uncoupled equations in y_i , where $y = \{y_i\}$. The equation in y_i is given by, $dy_i/dt = \lambda^{(i)}y_i$; therefore, $y_i(t) = y_i(0)\exp(\lambda^{(i)}t)$ or $y(t) = \exp(Dt)y(0)$. Since, $x = S^{-1}y$, $x(t)$ is given by, $x(t) = S^{-1}y(t) =$

$$\mathbf{S}^{-1} \exp(\mathbf{D}t) \mathbf{S} \mathbf{S}^{-1} y(0) = [\mathbf{S}^{-1} \exp(\mathbf{D}t) \mathbf{S}] x(0) = \exp(\mathbf{S}^{-1}(\mathbf{D}t) \mathbf{S}) x(0) = \exp(\mathbf{M}t) x(0).$$

Generation of single cell *in silico* data for first-order and nonlinear reaction models

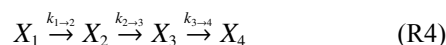
First-order *in silico* networks were simulated with MATLAB. For the results shown in Fig. 1A–B, 2,500 cells were assigned initial ($t_1 = 0$) values of n protein abundances drawn from lognormal distributions and the single cell kinetics were simulated using the closed form solution of Eq. (2). Averages and covariances at t_2 ($= 30$ s) were calculated by the software and used for fitting. For the results shown in Fig. 1C, means and covariances at t_2 used for fitting were simply calculated from the closed-form solutions given in Eqs. (5)–(6).

Nonlinear reaction networks in ground-truth models GT1, GT2, and the NK cell signaling model (Fig. 3A) were simulated using the software BioNetGen. Protein abundances in 250 single cells at $t_1 = 0$ were drawn from lognormal distributions. The rate constants and the parameters for the lognormal distributions are provided in BioNetGen scripts shown in the Supplementary Material. Mean abundances and covariances at a later time t_2 ($= 30$ s) were calculated in McSNAC from the synthetic single cell data. Perturbations of abundances of A, B, and C for Fig. 2A–B, and Supplementary Fig. S2 were generated by multiplying the initial concentration of the given protein by 2 or 1/2, while Fig. 2C and Supplementary Fig. S3 were generated by multiplying the initial concentration of the given protein by 10 or 1/10.

Initial abundances for the 8 measured phosphoproteins, their 7 unmeasured unphosphorylated counterparts (activating receptor is measured as both phosphorylated and double-phosphorylated), activating and inhibiting ligand, and a pErk phosphatase were sampled from lognormal distributions for the BioNetGen simulation of the NK cell signaling model. These initial values for all 250 cells can be found in the Supplementary Materials (x1_data.txt). For Lck perturbation experiments, the initial abundance of phosphorylated and unphosphorylated Lck across all cells for both was multiplied by the constant indicated in Fig. 3C.

Generation of first-order ground-truth model and candidate models

For Fig. 1D, we constructed an *in silico* dataset based on the reaction network in Reaction (R4), a representation of Reaction (R1) with $p = 4$.



We constructed a Boolean reaction matrix \mathbf{A} where presence (or absence) of a reaction $j \rightarrow i$ is indicated by 1 (or 0) at the ij th element. The ground-truth model given by Reaction (R4) is given by the matrix $\mathbf{A}_{\text{ground-truth}}$ below:

$$\mathbf{A}_{\text{ground-truth}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Since there are 12 off diagonal elements in the above matrix, and we construct 2^{12} candidate models where the each of the 12 off diagonal elements are set to 0 or 1. The candidate models that contain non-zero values in \mathbf{A} corresponding to $k_{1 \rightarrow 2}$, $k_{2 \rightarrow 3}$, and $k_{3 \rightarrow 4}$ contain the ground-truth model in Reaction (R4). The “number of true connections” is given by

$$n_{\text{true_connections}} = \sum_{i,j} ([\mathbf{A}_{\text{ground-truth}}]_{ij} \text{AND} [\mathbf{A}_{\text{proposed}}]_{ij})$$

The variable $n_{\text{true_connections}} = 3$ whenever a candidate model includes the ground-truth reaction in (R4).

Estimation of parameter confidence intervals

We estimate confidence intervals (CIs) for the elements of the \mathbf{M} matrix following a Profile Likelihood based approach [20]. First, we estimate the \mathbf{M} matrix corresponding to the minimum of the cost function in Eq. (7) using simulated annealing. Then we use this best fit value of the \mathbf{M} matrix to generate a large number of samples (e.g., 100,000) of matrix \mathbf{M}^* following $M_{ij}^* = M_{ij} \times b^{a_{ij}}$, where b is a scalar (e.g., 2), and a_{ij} is a random number between -1 and 1 drawn from a uniform distribution $U(-1,1)$. For each sample of \mathbf{M}^* we calculate a variable $d = \chi^2(\{\mathbf{M}^*\}) - \chi^2(\{\mathbf{M}\})$, describing the difference in the cost function χ^2 defined in Eq. (8) between the best fit \mathbf{M} and the sampled \mathbf{M}^* . Now, we chose a specific element (e.g., k, l) of the \mathbf{M} matrix and create a list of the a_{kl} values that were used to generate M_{kl}^* , and bin the a_{kl} values between -1 to $+1$. Next, d values that correspond to the sampled $\{\mathbf{M}_{kl}^*\}$ and specific bins of a_{kl} are collected, thus, each bin of a_{kl} would contain multiple values of d . We determine the smallest value of d for each bin in the a_{kl} space. Then, starting from the bin that contains $a_{kl} = 0$ and advancing to the bins of $a_{kl} > 0$ (or $a_{kl} < 0$) we determine the first bin that generates a minimum $d \geq 2.71$, this first bin corresponds to the upper (or lower) bound of M_{kl} .

It is important to note that this method suffers the curse of dimensionality. As the number of dimensions increases, so too must the number of samples to account for coverage of all bins in all dimensions. We report confidence intervals for our 13-dimension parameter

benchmarking (Fig. 1A), but not the 40-dimensional one (Fig. 1B), because we do not have the computational capability to form that many samples. This is a limitation of our confidence-interval estimation technique in high dimensions.

Simulated annealing hyperparameters

Simulated annealing (SA) is a global optimization algorithm that searches for an optimum value of a cost function, χ^2 . The optimization is based on the Metropolis algorithm for simulating a thermodynamic system at temperature T where system variables (*e.g.*, positions of atoms) associated with an energy E are distributed in a Boltzmann distribution ($\propto \exp(-E/(k_B T))$). k_B is the Boltzmann constant. Slowly cooling or annealing the system takes the system to lower energy states (*e.g.*, crystals) from higher energy random configuration of system variables (*e.g.*, gaseous states). This approach can be adapted for parameter estimation where the ‘energy’ $E(=\chi^2)$ represents the cost function which is a function of the parameter values. In SA the temperature parameter ($k_B T$) is initiated at larger values (*e.g.*, comparable to the largest values of E) and is decreased slowly to low values. As the temperature parameter is decreased gradually the system is supposed to relax to parameter configuration corresponding to the global minimum of E . However, the landscape of E can contain multiple minima and attaining the global minimum is not guaranteed in such situations. Several improvements on the basic SA algorithm have been proposed to deal with such scenarios. At each value of the temperature parameter, the Metropolis algorithm is run for sufficiently large number of Monte Carlo trials to make sure equilibrium Boltzmann distributions are realized in the simulations. At high temperatures, large increases in cost function can be accepted, while at low temperatures, only insignificant increases will be accepted. Simulated annealing gradually lowers the temperature parameter according to a cooling rate until a ‘global’ minimum is reached. For more information on simulated annealing, refer to [21].

All simulations reported in the manuscript used the same hyperparameters (cooling rate, number of Monte Carlo samples, *etc.*) found in the Supplementary Material script Simulated_Annealing.f, with the exception of the number of iterations of the temperature cooling loop. Table 2 below shows the number of steps used for each figure.

SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at <https://doi.org/10.15302/J-QB-022-0308>.

Table 2 Number of cooling steps for simulated annealing for results shown in each figure

Figure number	Number of cooling steps in simulated annealing
1A, 1B	2,000
1C	3,000
2A (same as S2)	2,000
2B (same as S3)	2,000
3C (same as S1)	5,000
S4	5,000

SOFTWARE AVAILABILITY

The software McSNAC can be found on github ([dweth/mcsnac](https://github.com/dweth/mcsnac)).

AUTHOR CONTRIBUTIONS

D.W. built the parent script and GUI for McSNAC, built *in silico* data, created MATLAB codes, and ran the simulations described in the text. S.M. wrote the simulated annealing Fortran code and NK cell signaling BioNetGen codes. J.D. and D.W. planned the experiments and analyzed the results. All authors contributed to writing and editing the manuscript.

ACKNOWLEDGEMENTS

This work is supported by the NIH awards R01-AI 143740 and R01-AI 146581 to J.D. We would like to thank Bill Stewart for his help in developing the profile likelihood confidence interval estimator.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Darren Wethington, Sayak Mukherjee and Jayajit Das declare that they have no competing interests.

This article does not contain any studies with human participants or animals performed by any of the authors.

OPEN ACCESS

This article is licensed by the CC By under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

1. Kholodenko, B. N. (2006) Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.*, 7, 165–176
2. Murphy, K. and C. Weaver. (2016) *Janeway’s immunobiology*.

- New York: Garland science
3. Chakraborty, A. K. and Das, J. (2010) Pairing computation with experimentation: a powerful coupling for understanding T cell signalling. *Nat. Rev. Immunol.*, 10, 59–71
 4. Das, J. and Lanier, L. L. (2019) Data analysis to modeling to building theory in NK cell biology and beyond: how can computational modeling contribute? *J. Leukoc. Biol.*, 105, 1305–1317
 5. Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., *et al.* (2014) A draft map of the human proteome. *Nature*, 509, 575–581
 6. Salman, H., Brenner, N., Tung, C. K., Elyahu, N., Stolovicki, E., Moore, L., Libchaber, A. and Braun, E. (2012) Universal protein fluctuations in populations of microorganisms. *Phys. Rev. Lett.*, 108, 238105
 7. Swain, P. S., Elowitz, M. B. and Siggia, E. D. (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 99, 12795–12800
 8. Spitzer, M. H. and Nolan, G. P. (2016) Mass cytometry: single cells, many features. *Cell*, 165, 780–791
 9. Bandura, D. R., Baranov, V. I., Ornatsky, O. I., Antonov, A., Kinach, R., Lou, X., Pavlov, S., Vorobiev, S., Dick, J. E. and Tanner, S. D. (2009) Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.*, 81, 6813–6822
 10. Das, J. and Jayaprakash, C. (2018) Systems immunology: an introduction to modeling methods for scientists. Cleveland: CRC Press
 11. Goldstein, B., Faeder, J. R. and Hlavacek, W. S. (2004) Mathematical and computational models of immune-receptor signalling. *Nat. Rev. Immunol.*, 4, 445–456
 12. Veglia, F., Perego, M. and Gabrilovich, D. (2018) Myeloid-derived suppressor cells coming of age. *Nat. Immunol.*, 19, 108–119
 13. Mukherjee, S., Jensen, H., Stewart, W., Stewart, D., Ray, W. C., Chen, S. Y., Nolan, G. P., Lanier, L. L. and Das, J. (2017) *In silico* modeling identifies CD45 as a regulator of IL-2 synergy in the NKG2D-mediated activation of immature human NK cells. *Sci. Signal.*, 10, eaai9062
 14. Yuan, B., Shen, C., Luna, A., Korkut, A., Marks, D. S., Ingraham, J. and Sander, C. (2021) CellBox: interpretable machine learning for perturbation biology with application to the design of cancer combination therapy. *Cell Syst.*, 12, 128–140.e4
 15. Mukherjee, S., Stewart, D., Stewart, W., Lanier, L. L. and Das, J. (2017) Connecting the dots across time: reconstruction of single-cell signaling trajectories using time-stamped data. *R. Soc. Open Sci.*, 4, 170811
 16. Faeder, J. R., Blinov, M. L. and Hlavacek, W. S. (2009) Rule-based modeling of biochemical systems with BioNetGen. *methods. Mol. Biol.*, 500, 113–167
 17. Harris, L. A., Hogg, J. S., Tapia, J. J., Sekar, J. A., Gupta, S., Korsunsky, I., Arora, A., Barua, D., Sheehan, R. P. and Faeder, J. R. (2016) BioNetGen 2. 2: advances in rule-based modeling. *Bioinformatics*, 32, 3366–3368
 18. John, W., Stewart W. CL, Ciriya J. and Jayajit D. (2022) Generalized Method of Moments improves parameter estimation in biochemical signaling models of time-stamped single-cell snapshot data. *bioRxiv*, p. 2022.03.17.484491
 19. Lück, A. and Wolf, V. (2016) Generalized method of moments for estimating parameters of stochastic reaction networks. *BMC Syst. Biol.*, 10, 98
 20. Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U. and Timmer, J. (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25, 1923–1929
 21. Press, W. H. (1992) Numerical recipes in C: the art of scientific computing. 2nd ed. New York: Cambridge University Press. xxvi, 994 p
 22. Das, J. (2010) Activation or tolerance of natural killer cells is modulated by ligand quality in a nonmonotonic manner. *Biophys. J.*, 99, 2028–2037
 23. Jiang, K., Zhong, B., Gilvary, D. L., Corliss, B. C., Hong-Geller, E., Wei, S. and Djeu, J. Y. (2000) Pivotal role of phosphoinositide-3 kinase in regulation of cytotoxicity in natural killer cells. *Nat. Immunol.*, 1, 419–425