Title: Probability-based mechanisms in biological networks with parameter uncertainty

Authors: Oscar O. Ortega[1,†], Mustafa Ozen[2,†], Blake A. Wilson[2], James C. Pino[2], Michael W. Irvin[2], Geena V. Ildefonso[1], Shawn P. Garbett[3], Carlos F. Lopez[2,4,*]

Abstract:
Mathematical models of biomolecular networks are commonly used to study cellular processes; however, their usefulness to explain and predict dynamic behaviors is often questioned due to parameter uncertainty. Here, we employ a novel dynamic network analysis within a Bayesian statistic context to probabilistically identify dominant reaction paths in signaling networks. We introduce our approach in an extrinsic apoptosis model using fluorescence microscopy data from HeLa cells for parameter inference. While Bayesian optimization yields several parameter vectors with equally good-fit to experimental data, three execution modes emerge upon analysis of the reaction flux throughout the network using a discretization theory approach. Each mode exhibits different behaviors to *in silico* perturbations through different mechanisms, thus underscoring the need for thorough characterization of model parameters and associated dynamics. We then apply our approach on the JAK2/STAT5 pathway in colony-forming unit-erythroid (CFU-E) cells. Our analysis provides mechanistic explanation for survival responses of CFU-E cell population that was impossible by considering protein concentration dynamics in isolation.

Affiliations:
1. Chemical and Physical Biology Program, Vanderbilt University, Nashville, TN 37212
2. Dept. of Biochemistry, Vanderbilt University, Nashville, TN 37212
3. Dept. of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37203
4. Dept. of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203

[*] To whom correspondence should be addressed: Tel: +1 (615) 936-1311; Email: c.lopez@vanderbilt.edu
[†] These authors contributed equally to this work

Introduction

Many dynamic processes can be represented as networks of interconnected components enabling the study of their dynamics and signaling mechanisms[1–5]. In systems biology applications, these analyses typically entail building a network to represent intracellular biochemical reactions, either from prior knowledge or through network inference, developing a mathematical model of the network interactions, and subsequently calibrating the model to experimental data[6–8]. Although small networks have been studied with great success, larger networks present multiple challenges due to model parameter uncertainty, which in turn is rooted in a dearth of available data. This leads to a parameter optimization problem, whereby multiple parameter sets can be obtained that reproduce experimental measurements equally well and with no guidelines to choose one parameter set over another[9–11]. Therefore a practice has emerged where one or a few parameter vectors are chosen – often at random – to make mechanistic predictions, with varying degrees of success[12–14]. This has led to ongoing criticisms regarding the usefulness of large and complex mathematical models of cellular processes with multiple uncertain parameters.

Multiple approaches have been used to extract mechanistic insights from mathematical models of network-driven processes. These approaches range from simple tracking of one or multiple species in the network[14–18] to more complex applications, such as Information Theoretic concepts to estimate the channel capacity between nodes in a network[19–22]. The latter approaches have provided a quantitative paradigm to study input/output relationships in networks[20–24]. Seminal works in the field have shown that the maximum channel capacity of a biochemical process is context dependent and could require the cooperativity of multiple cells to achieve actual information transfer[21]. In recent work, our laboratory showed a non-linear dependence between channel capacity and reaction flux through a network[22]. Despite these advances to explore network-driven processes, a truly systems-level interpretation of non-equilibrium dynamics in a network and how these transient dynamics are impacted by parameter uncertainty remains a standing challenge in the field. Addressing this challenge might not only provide us a better resolution of the signal processing mechanisms in complex biological systems but also enable us constructing biologically relevant hypotheses to be tested in the presence of parameter uncertainty; hence, it is the focus of the present work.

Stoichiometric flux balance analysis (FBA) methods are perhaps the most widely used approaches to analyze reaction flux in a biochemical reaction network – typically at equilibrium[25–29]. These analyses allow computation of the steady-state flux distributions in biochemical networks by solving a constrained linear optimization[30] with the assumption that cells perform optimally with respect to a metabolic function of interest[27] such as growth or synthesis of a biomass[31], ATP production[32], and production of a specific metabolic product[33]. Although these methods are useful to analyze and interpret the metabolic capabilities of a system, prediction accuracy and the essence of the found flux distribution by FBA is vulnerable to the quality of the constraints implemented in the linear optimization and the model[34]. In addition, although FBA can provide dynamic flux profiles, it is limited due to the arbitrary choice of a single metabolic state at each time interval, which leads to only a fraction of the space of

dynamic flux profiles[27]. Moreover, solving the linear optimization problem becomes more challenging as the network/model size increases and the approach is largely unapplicable to non-equilibrium conditions.

To address the challenge to extract mechanistic insights from non-equilibrium network-driven processes in the context of uncertain model parameters, we introduce an approach allowing the analysis of dynamic execution patterns in biochemical networks by identifying detailed instantaneous fluxes throughout the system. In principle, the number of interactions and temporal dynamics present in a network make it difficult to identify emergent behaviors from myriad concurrent biochemical reactions. We therefore employ an approach inspired in Tropical Geometry and Ultradiscretization Theory that makes it possible to map continuous functions into piecewise linear meshes in the Ultradiscrete/Tropical space[35,36]. We test the hypothesis that the proposed dynamic analysis of biochemical networks enables us to identify dominant reaction fluxes that could in turn lead to emergent signal execution modes (i.e., unique dominant paths of flux propagation) and their dependence on model parameters. We reasoned that we could cast this analysis onto a Bayesian probability framework to assign statistical weight to the identified modes of signal execution, thus providing a novel statistical and mechanistic interpretation for network-driven processes.

The remainder of the article is organized as follows. We first show how Bayesian inference of model parameters can yield tens of thousands of parameter vectors that all reproduce an experimental data set equally well on a simplified extrinsic apoptosis model. We then introduce our method – inspired in Ultradiscretization Theory and Tropical Algebra – to define a dynamic signal execution fingerprint, which we then use to find whether execution modes can cluster despite different model parameters. Surprisingly, we find that despite thousands of parameter vectors reproducing the experimental data equally well, only a handful of execution modes emerge as possible signal processing mechanisms. We then demonstrate that choosing a single parameter vector offers a biased description of the signaling process, which could easily lead to misleading interpretations of execution mechanisms. Finally, we demonstrate our approach on the Janus kinase 2/signal transducer and activator of transcription 5 (JAK2/STAT5) signal transduction pathway in erythroid progenitor cells and identify the modes of signal execution that contribute to cellular survival responses upon stimulation with erythropoietin. Overall, our work introduces a probability-based methodology to explore network-driven processes and to understand mechanistic model predictions in the presence of parameter uncertainty while providing opportunity to gain novel insights into signal processing mechanisms in complex biological systems as elaborated in the subsequent sections.

Results

Bayesian parameter optimization yields indistinguishable protein concentration dynamics

To investigate the impact of parameter uncertainty on biochemical signal execution, we calibrated model of apoptosis to experimental data using a Bayesian parameter inference approach[37]. Apoptosis is a ubiquitous biological process in metazoans used as a mechanism to maintain cell numbers and overall organism homeostasis[38]. We explored calibration of various versions of the Extrinsic Apoptosis Reaction Model (EARMv2.0)[39] to ensure that inferred model parameter values would converge to a distribution and could therefore be identified. The abridged EARM (aEARM), depicted in Figure 1A, was the largest model we could build that would both preserve key biochemical interactions, and achieve convergence for all model parameters after model calibration. The abridged model preserves key biological features of apoptosis execution including signal initiation by TNF-Related Apoptosis Inducing Ligand (TRAIL), subsequent activation of initiator caspases (Caspase 8)[40], type 1 and type 2 activation of effector caspases (Caspase 3)[41], and completion of apoptosis execution by cleavage of Poly(ADP-ribose) polymerase (PARP)[42] . Overall, aEARM comprises 22 molecular species and 34 kinetic parameters (see Methods for details). We used PyDREAM – a python implementation of the MT-DREAM(ZS) algorithm[8] – to calibrate the model to previously published experimental data that comprises the concentration dynamics of truncated Bid (tBid) and cleaved PARP (cPARP) in HeLa cells[43]. As HeLa cells are Type-II cells, the simulated results are representative of signal processing and execution of Type-II cells treated with death-inducing ligands such as TRAIL[43].

Our Bayesian model calibration yielded 27,242 unique parameter vectors that fit the experimental data equally well (see Methods for details). We note that throughout the manuscript, a *parameter vector* refers to a set of positive real values, one value for each of the kinetic parameters defined in aEARM, used to run a simulation. A parameter distribution refers to the frequency of occurrence of different values from the same kinetic parameter that provide equally good-fit to the experimental data (Figure 1B). Each model parameter distribution converged by the Gelman-Rubin diagnostics[45] as shown in Supplemental Table 1 and Supplemental Figure 2. We obtained the marginal distributions from the sampled parameter vectors as shown in Figure 1C and Supplemental Figure 3. Given the Markov Chain Monte Carlo (MCMC)[46] aspect of our parameter inference method, we were able to obtain parameter vector probabilities as shown in Figure 1D. The probability distribution of parameter vectors exhibits characteristic exponential-like decay shape indicating that some parameters are more likely than others. Once we obtained a rigorously calibrated model, we then explored signal execution in aEARM from a probabilistic perspective.

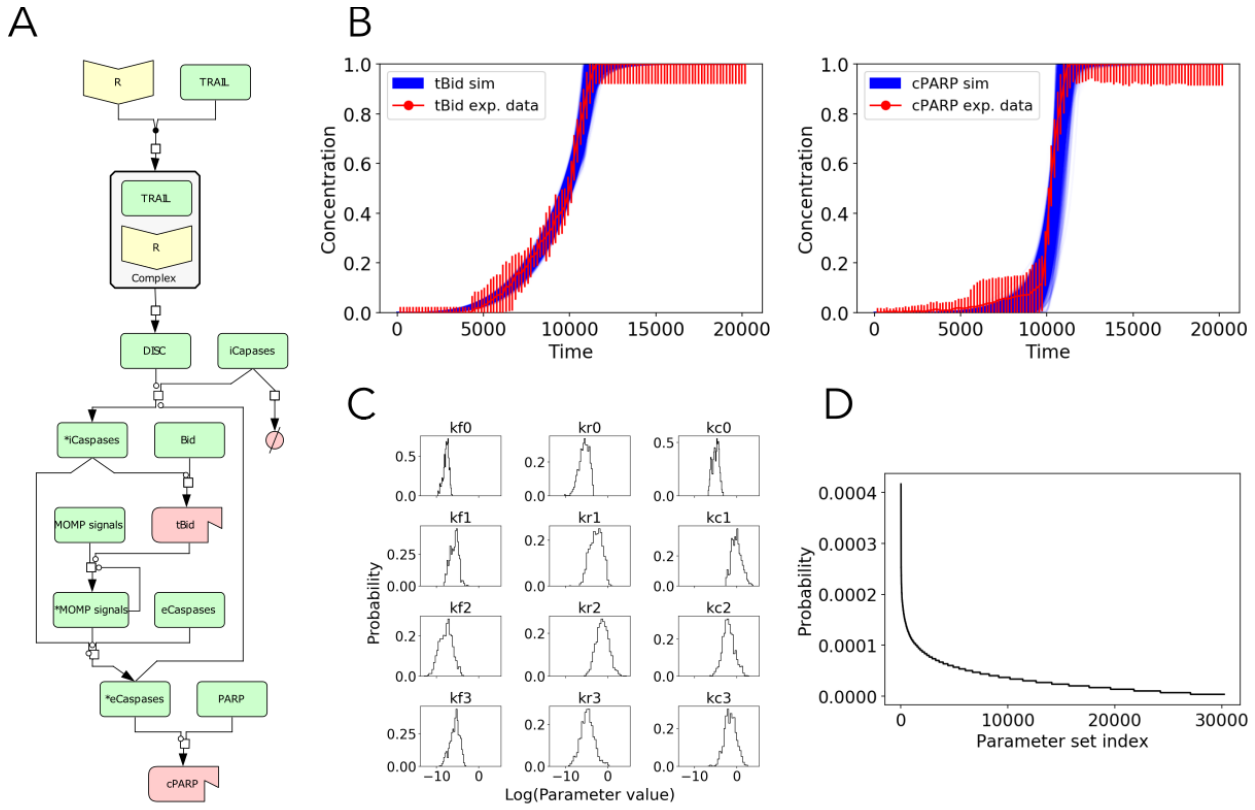

Figure 1. Abridged Extrinsic Apoptosis Reaction (aEARM) network and parameter calibration results. A) Reaction network using the Kitano convention. Yellow nodes are protein receptors, green nodes are generic proteins, and red nodes are truncated/cleaved proteins. B) Simulated trajectories of truncated Bid and Cleaved PARP calibrated to reproduce the experimental data. Red dots and bars indicate the mean and standard deviation of the experimental data and blue lines correspond to the simulated trajectories. C) Marginal probability distributions of the first 12 individual kinetic parameters that were recovered from the PyDREAM run by integrating out all other dimensions. Forward rates, reverse rate, and catalytic values were all found to be within biologically relevant ranges[44]. D) Probability of each of the unique parameter vectors sampled after burn-in in the PyDREAM calibration. To obtain the probability of each parameter set the number of visits to a specific parameter vector was normalized by the total number of visits.

A discretized flux-based analysis of signal execution in networks

As shown in Figure 1B, all parameter vectors obtained from Bayesian calibration yield protein concentration dynamics indistinguishable from the experimental trajectories of tBid and cPARP. Individual parameters from these vectors take widely different values as depicted by their distributions in Figure 1C. This uncertainty in the parameter values affects the reaction rates of the protein interactions generating different reaction flux patterns in the network during signal execution. We therefore studied the non-equilibrium flux of the reactions in the aEARM network to explore whether parameter uncertainty yields specific patterns of signal execution.

Analysis of flux dynamics during signal execution requires tracking the signal flow through a network at all simulation time points as multiple concurrent reaction rates consume or produce molecular species. For a particular species, we assumed that the reactions with the highest flux at any given time dominate the network signal execution and provide a proxy to observe the

effect of different parameter vectors in the network. Our aim was thus to identify the reaction rates with the highest flux throughout the whole network as simulations evolve over time. To analyze the non-equilibrium flux and find the dominant reaction paths during signal execution, we developed an algorithm inspired by Ultradiscretization Theory and Tropical Algebra as described in *Methods*[36,47]. Our approach enabled us to identify paths relevant for flux propagation in non-equilibrium states. We refer to these paths of flux propagation through the network as *execution modes* for the remainder of this manuscript.

We introduce the workflow for reaction flux discretization and execution mode identification as shown schematically in Figure 2. Signal discretization requires three steps. First, we identify a target node (Figure 2A) for which the signal flux will be tracked. Second, we calculate the reaction rates that produce or consume the target node, identify the largest reaction rate (x) and test whether it is dominant over other reactions (y) using the discretization operation $|\log_{10} x| - |\log_{10} y| > \rho$, where ρ is the order of magnitude difference necessary to consider dominance (see Methods section for details). Third, we identify the chemical species produced by the dominant reaction(s) and jump to that species, thus starting the process again from the first step, and thereby tracking the dominant signal fluxes through the whole network and obtaining a subnetwork. This dominant subnetwork is assigned a unique integer label as shown in Figure 2A. The procedure is repeated for all simulation time points. As a result, the dynamic nature of signal execution for a given parameter vector is abstracted to a sequence of labels that can be compared to other sequences using a suitable metric (Figure 2B). We call this sequence of labels obtained from a simulation as *dynamic fingerprint* because it is unique for a given signal processing event with a specific parameter set. A workflow of the algorithm is shown in Supplemental Figure 1.

Key execution modes emerge despite parameter uncertainty

To identify the dynamic execution patterns in aEARM in response to death ligand cues, we carried out our signal discretization analysis for the 27,242 unique parameters and obtained dynamic fingerprints for each parameter vector. We then asked whether there were similarities among dynamic fingerprints across parameter sets. To investigate this question, we quantified the distance between each dynamic fingerprint using the Longest Common Subsequence (LCS) metric[48]. We chose this metric due to its sensitivity to order differences in which successive subnetworks labels appear[49]. This metric thus assigns a larger distance to a pair of dynamics fingerprints that execute the signal differently. Calculation of the pairwise distance between all dynamic fingerprints resulted in a 27,242 by 27,242 distance matrix. This matrix enabled us to use an agglomerative clustering algorithm[50] to probe whether clusters of dynamic fingerprints would emerge. As shown in Figure 3A, we found that all 27,242 dynamic fingerprints could be classified into three clusters (Supplemental Table 2), which we denominate as “execution modes”. Given that each parameter vector has a defined probability (Figure 1D) and is associated with a dynamic fingerprint, we calculated the probabilities of signal execution through each mode as 42%, 36%, and 22% for Execution Mode 1 (EM1), Execution Mode 2 (EM2), and Execution Mode 3 (EM3), respectively. These three execution modes account for

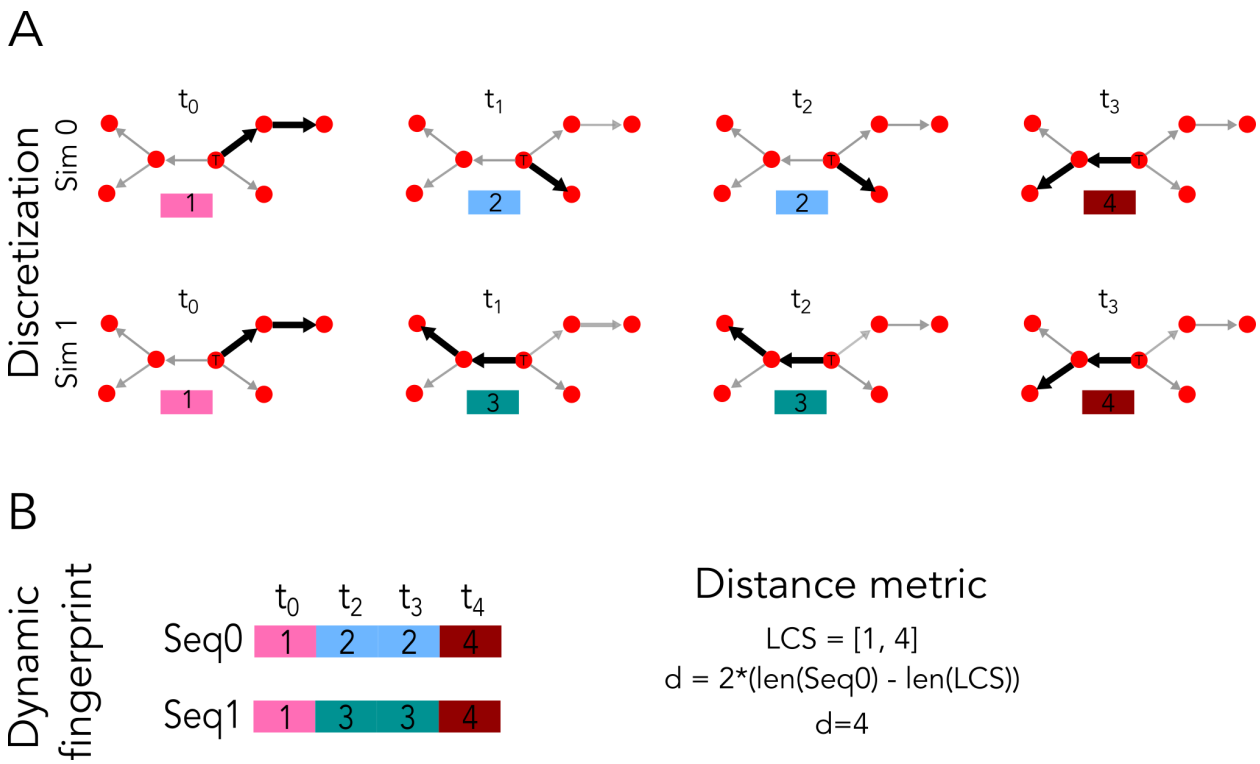

Figure 2. Discretization analysis workflow. A) First, the network of interaction is obtained from a model and a target node (labeled T) from where the signal is going to be tracked is defined. Red nodes are molecular species in a model, edges represent interactions between nodes, bolded edges are the dominant interactions. Next, at each time point of a simulation our analysis obtains a dominant subnetwork, bolded edges in the network, through which most of the signal is flowing and this subnetwork is assigned a label. Sim 0 and Sim 1, simulations ran with different parameter sets, exhibit different dominant subnetworks. B) As each subnetwork is assigned a label, we can get a sequence of labels per simulation that can be compared to other simulations with the Longest Common Subsequence metric and obtain a distance matrix. This distance matrix can be used with clustering algorithms to obtain groups with similar modes of signal execution.

all the parameter vectors inferred from the explored parameter space and no vectors were found that did not belong to either of these clusters. We note that these execution modes comprise three of the eight possible subnetworks for signal flow.

The dominant flux subnetwork for each execution mode is shown schematically in Figure 3B. The highlighted paths represent the dominant reaction fluxes, i.e., these fluxes are within an order of magnitude of the largest reaction at each node for the given parameter set and simulation time point. As shown in Figure 3B (yellow pathway), EM1 comprises events from initial death-ligand binding to the receptor, through formation of the Death Inducing Signaling Complex (DISC), and subsequent activation of initiator Caspase. The initiator Caspase then truncates and activates Bid, which in turn activates MOMP, a species that abstracts mitochondrial outer membrane pore (MOMP) formation. As highlighted in Figure 3B (EM1), activated MOMP is dominantly used to both activate more MOMP, through the positive feedback loop, and activate the effector Caspase (eCaspase). Note that although the species MOMP is represented as a single node in the model (Figure 1A), it indeed represents a more complex pathway through initiator Caspases, Bid, Bax, Smac, XIAP, and apoptosome which eventually leads to eCaspase activation[14]. Therefore, the direct interaction between MOMP and eCaspase in the aEARM model is a simplification of the indirect interaction elaborated above.

The flux through the network in EM2 is similar to that of Mode 1 but the execution path differs at MOMP regulation. As highlighted in blue in Figure 3B (EM2), activated MOMP is largely consumed in the positive feedback loop to activate more MOMP. The signal flux downstream of activated MOMP is at least an order of magnitude less than the highlighted route for the parameters in EM2. Therefore, eCaspase activation and apoptosis execution takes place due to a smaller reaction flux in the network relative to the MOMP-level activity in EM2. For those parameters belonging to EM3, signal execution seems to flow largely toward PARP cleavage, with less MOMP-level regulation. Our results therefore show that despite uncertainties in inferred model parameters due to limited available data, the modes of signal execution are identifiable. Identifying a limited number of execution modes highlights the need to thoroughly characterize the model parameter space, given experimental constraints, to understand and make inferences about execution mechanisms. We note that using a single vector of parameters would lead to incomplete model prediction as no one single parameter vector captures the rich dynamics exhibited by all the statistically inferred parameter vectors.

To further understand the impact of each execution mode on MOMP regulation, we examined the relative concentration of activated MOMP (MOMP*) and the binding complexes in which it participates. We calculated the percentages of MOMP*, inactive MOMP bound to MOMP* (MOMP-MOMP*), and eCaspase bound to MOMP* (eC-MOMP*). As shown in Figure 3C, the relative abundance of these species over time exhibits different concentration patterns in each execution mode. In EM1, the relative abundance of eC-MOMP* is ~20%, indicating that the signal flow through this reaction is lower than in EM3 but still important in the overall dynamics. In EM2, 85 % of MOMP* is bound to inactive MOMP at all time points before cell death. This can be explained by a high MOMP activation rate due to Bid and the MOMP* positive feedback loop autoactivation. In contrast to EM2, the MOMP-MOMP* abundance in EM3 decreased to ~35%, while eC-MOMP* is increased to ~50%. This increase in eC-MOMP* abundance, indicates that the binding rate of MOMP* to eC is larger than the binding rate of MOMP to MOMP*. We note that the initial concentration of inactive MOMP in the model is an order of magnitude larger than that of eC. Thus, this result is in stark contrast with the result from EM2 where the reaction rates exhibit different relative values.

Overall, the three execution modes arise from the several parameters sets with equally well experimental data fitness. Although the difference between the dominant pathways in the observed execution modes might seem visually insignificant (i.e., a few interactions), they are indeed considerably different when considering the coarse-grained aspect of the aEARM model, as each execution mode is elaborated above. In addition, we would like to note that it does not mean that the nondominant pathways or interactions are not important or do not have any effect in the model responses, but rather that they are not the dominant interaction during the execution of the network process. For instance, eC is not included in the dominant pathway of the second execution mode (Figure 3B) while it exists in the other two modes. However, perturbing eC results in different responses in all execution modes compared to wildtype, which is thoroughly explained in the next section.

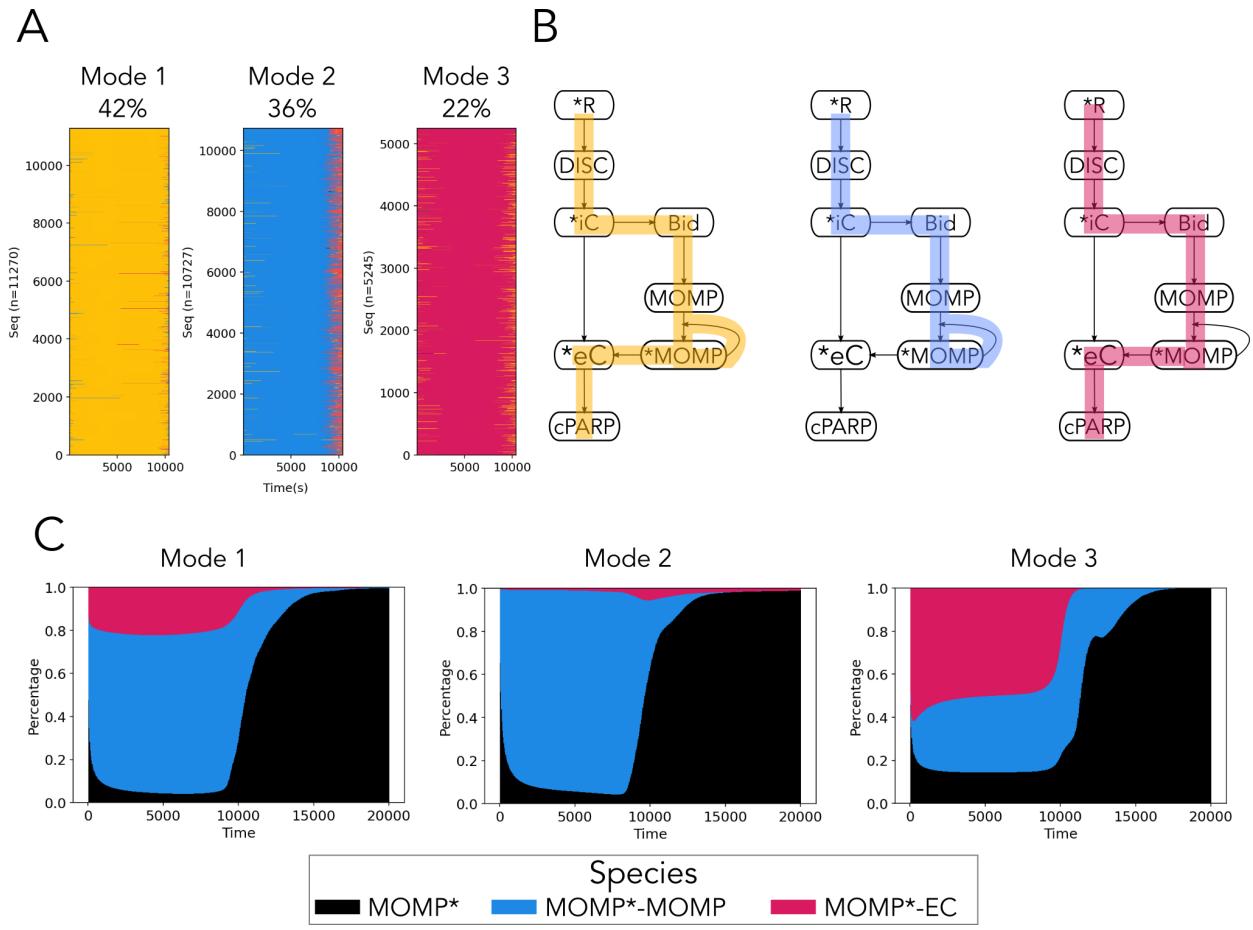

Figure 3. Modes of signal execution in aEARM. A) Dynamic fingerprints organized by the clusters they belong to. Each cluster plot is composed of horizontal lines that correspond to dynamic fingerprints, i.e. sequences of dominant subnetworks, and each subnetwork is assigned a different color. B) Signal execution modes as defined by the most common subnetwork in each cluster. The complete aEARM network is shown in black, and the dominant subnetworks for Mode 1, 2, and 3 are highlighted in yellow, blue, and red, respectively. C) Effect of the different signal execution modes in the relative concentration of activated MOMP and its associated complexes. For each cluster, we calculated the temporal relative concentration of MOMP, MOM*P-MOMP and MOMP*-eC point by obtaining their individual average concentrations and dividing it by the sum of their concentrations. This visualization provides insights about the usage MOMP* in each cluster.*

Signal execution modes respond differently to eCaspase perturbation

We then asked whether *in silico* experiments could help us understand differences in signal execution that could lead to experimentally testable hypotheses. We therefore carried out *in silico* knockdown experiments of eCaspase, as its activation is essential for the final steps of apoptosis execution[51]. In addition, eCaspase inhibitors are readily available for laboratory use[52,53]. We tested the hypothesis that each execution mode would exhibit different execution mechanisms when eCaspase is knocked down by 50%. To explore the impact of eCaspase knockdown for each execution mode, we compared the concentration dynamics for MOMP and cPARP given by wild type (WT) and eCaspase knockdown conditions.

310 For each execution mode we plotted the cPARP concentration trajectories and obtained the
311 time of death (ToD) for each simulated cell as described in Methods. As shown in Figure 4A, the
312 ToD in EM1 exhibits a modest decrease of 14.96 s, but also presents a larger standard deviation
313 of 702 s. For EM2, the ToD increases from 10351 $\pm$ 132 s (WT) to 10809 $\pm$ 226 s when eCaspase
314 is knocked down (Δt = 458 s). In contrast, EM3 eCaspase knockdown leads to a decreased ToD
315 from 10261 $\pm$ 83 in WT to 9507 $\pm$ 516 s in the knockdown (Δt = -754 $\pm$ 523 s). These results
316 therefore show that each execution mode can exhibit considerably different – and a times
317 juxtaposed – responses to the same perturbation.
318
319 We then probed the effect of the eCaspase knockdown on the reaction rates associated to
320 MOMP* (a node where the signal bifurcates): MOMP* binding to MOMP, and MOMP* binding
321 to eC. Specifically, we focused on the reaction rate peak and the time to reach peak of the
322 reaction rate throughout the simulation, as shown in Figure 4B and Supplemental Figure 2. The
323 peaks of the MOMP*+MOMP binding reaction (Figure 4B upper row) appear unchanged across
324 all execution modes, yet the time to reach the peaks vary significantly. Compared to the WT
325 simulations, the median time to peaks is 6.14%, 0.36%, and 11.76% faster for modes 1, 2, and 3,
326 respectively. Concurrently, the peaks of the MOMP*+eC binding reaction (Figure 4B lower row)
327 reduce approximately 50% as expected by the 50% reduction of the available eC, and the
328 median time to peaks are 6.77%, 0.4%, and 14.48% faster for modes 1, 2, and 3, respectively. In
329 combination, for mode 1, the relative change of the MOMP and eC reaction peaks have large
330 interquartile ranges IQR= -10.37% to –1.01% and IQR=-1.13% to –11.39%, respectively, which
331 explains the variability in the time to cell death. For mode 2, the time to the peak of MOMP and
332 eC reactions change marginally and given that the eC peak is 50% of the WT condition, this
333 leads to longer times to accumulate the necessary number of eC molecules for cells to commit
334 to apoptosis. Finally, for mode 3, the median time to reach the MOMP reaction peak and the eC
335 reaction peak are 11.76% and 14.48% faster than in the WT condition, respectively. This causes
336 faster activation of MOMP and eC which leads to earlier apoptosis in cells. To summarize,
337 although the biochemical signal flows differently in each execution mode, the protein
338 concentration dynamics exhibit similar outcomes (Figure 4A Wild Type). However, when a
339 perturbation is made to the network, the outcome can vary significantly, as shown for each
340 execution mode. Considering that each execution mode was obtained from several parameter
341 vectors with equally well data fitness, such a difference is theoretically noteworthy and might
342 be an indicator of more significant difference in other systems. Moreover, these observations
343 highlight the importance of choices of model parameter values as different parameter vectors
344 may result in different outcome. Therefore, care should be taken while analyzing data-driven
345 network models in the presence of parameter uncertainty.
346
347 In addition to perturbation analysis of the execution modes, we showed how the execution
348 mode uncertainty can be reduced through parameter measurements using machine learning
349 techniques in the Supplemental Information (Section 1). Furthermore, we studied the signal
350 execution modes of the extended apoptosis model (EARMv2.0)[39] to show feasibility of the
351 proposed framework on a larger model with higher parameter uncertainty, whose details are
352 also provided in the Supplemental Information (Section 2). In the next section, we apply our
353 framework to the Janus kinase 2/signal transducer and activator of transcription 5 (JAK2/STAT5)

signal transduction pathway in erythroid progenitor cells and provide novel mechanistic insights into their survival responses when the cells are exposed to different erythropoietin (Epo) levels.

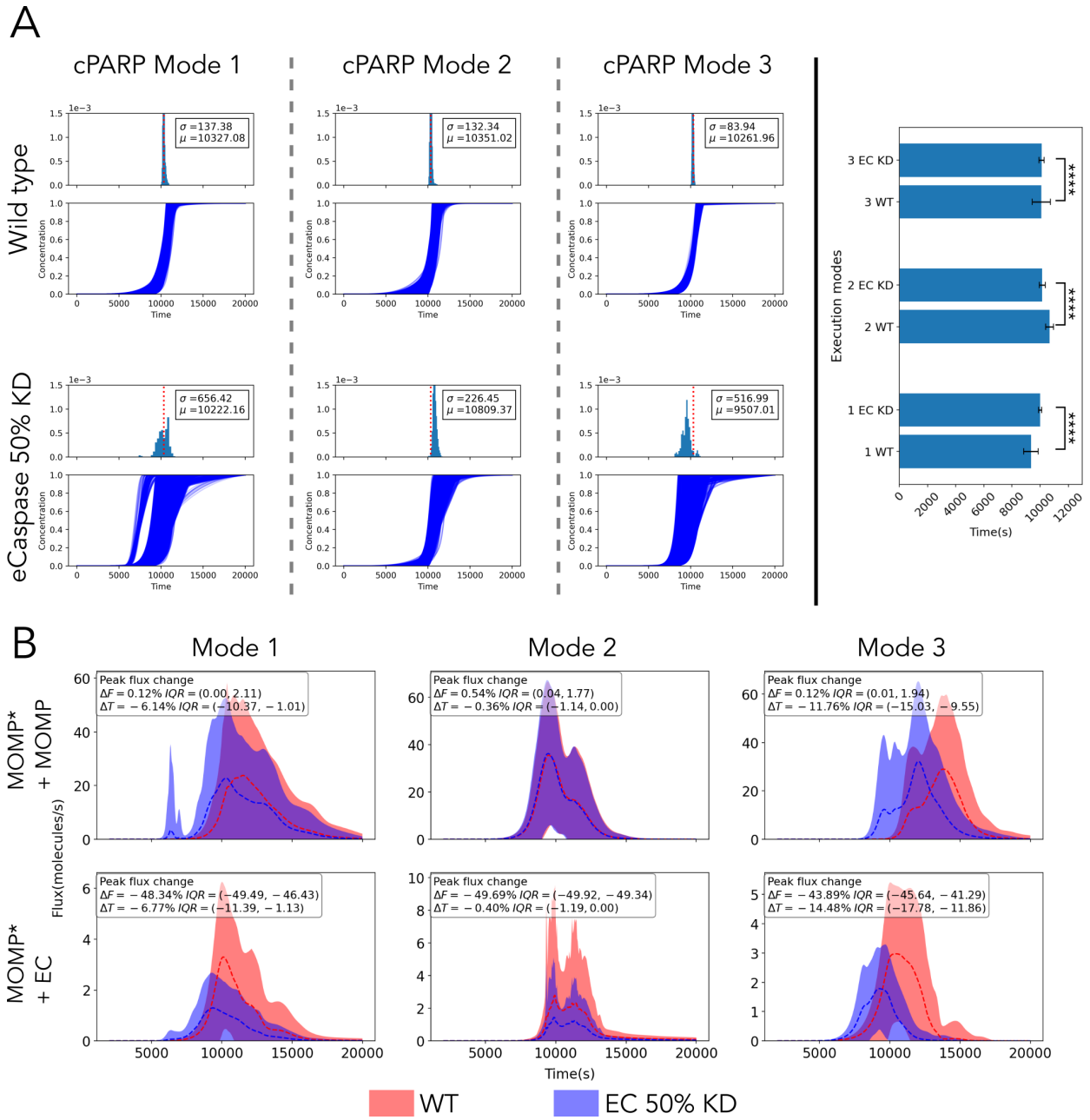

Figure 4. Time of death responses are markedly different for the same perturbation. A) Cleaved PARP (cPARP) protein concentration trajectories for the "wild type" case (top row) grouped by Execution Modes. Mode 1, 2, 3 have 11270, 10727, and 5245 trajectories, respectively. Inset includes the average time to death and the standard deviation calculated from all trajectories in each execution mode. PARP cleavage exhibits a markedly different trajectory pattern (bottom row) after eCaspase is knocked down by 50%. B) MOMP + MOMP and MOMP* + eC reaction rate trajectories. Dashed lines correspond to the mean of all reaction rates trajectories in an execution mode and the shadows represent the standard deviations. Trajectories from the "wild type" condition are colored in red and trajectories from the 50% effector caspase KD are colored in blue, and show key differences in their dynamics. Insets include the median percentage change in the reaction rate peak (ΔF) and the time to reach that peak (ΔT) in the eC KD condition relative to the wild type condition. The interquartile range is included as a measure of the variation in the ΔF and ΔT changes.*

368 **Identifying execution modes of the JAK2/STAT5 signal transduction network with respect to**
369 **increasing erythropoietin**
370
371 To show how the proposed dynamic network analysis pipeline can provide mechanistic insights
372 into existing biological systems, we applied it to a model of the Janus kinase 2/signal transducer
373 and activator of transcription 5 (JAK2/STAT5) signal transduction pathway that captures cellular
374 population response[54,55]. This pathway has an active role in the survival of erythroid progenitor
375 cells[56], which are highly sensitive to erythropoietin (Epo) at the colony forming unit erythroid
376 (CFU-E) stage[57]. To ensure a robust physiological production of erythrocytes, these cells
377 demonstrate a gradual input-output relationship when exposed to Epo[58]. The details of the
378 JAK2/STAT5 signaling network population-average model and its calibration, the data used in
379 the model calibration, and the uncertain parameter vectors providing equally well data fitness
380 were obtained from Adlung et al.[55].
381
382 The JAK2/STAT5 signaling network is presented in Figure 5A. The signal transduction, which
383 eventually induces gene expression[59] and CFU-E survival[54], is initiated by binding of Epo to its
384 receptor EpoR. This is followed by the activation of receptor-associated JAK2 by the formed
385 Epo-EpoR complex. Active JAK2 phosphorylates both the cytoplasmic tail of EpoR on multiple
386 tyrosine residues[60] and STAT5 that is recruited into the receptor complex[61]. Then,
387 phosphorylated STAT5 (pSTAT5) translocases to the nucleus and induces the anti-apoptotic
388 genes for the control of cell survival[62]. STAT5 also targets other genes including cytokine-
389 inducible SH2-domain-containing protein (Cish)[63] and suppressor of cytokine signaling 3
390 (SOCSC3)[64], which attenuates signal transduction. Finally, signal transduction is terminated by
391 dephosphorylated JAK2[60] caused by recruitment of SH2-domain-containing protein tyrosine
392 phosphatase 1 (SHP1) to the activated receptor via its SH2 domain[65].
393
394 Although the fraction of surviving cells increases in a graded level, population mean
395 fluorescence intensity of the total STAT5 – the key integrator of Epo-induced survival signal
396 transduction in CFU-E cells – does not change with respect to increasing Epo doses[54,55]. Upon
397 exposing the cells to Epo = 5 U/ml, the total STAT5 remains nearly constant, as shown in Figure
398 5B. On the other hand, mean fluorescence intensity of the pSTAT5 in the population gradually
399 increases in cytoplasm and nucleus over time that correlates with survival responses after Epo
400 stimulation (Figure 5B). Similarly, CFU-E cells demonstrate a rapid increase in the level of
401 pSTAT5 in cytoplasm relative to total STAT5 (Figure 5C). Furthermore, the overall pSTAT5 levels
402 increase as Epo level increases (Figure 5D), which is correlated with the increasing cell
403 survival[55]. Despite the model reproducing the experimental data and identifying
404 aforementioned key regulators of CFU-E cell survival responses with respect to Epo doses[55], a
405 mechanistic understanding of these networked processes that give rise to this behavior was
406 lacking. Revealing the underlying mechanism of this behavior and understanding of signal
407 processing mechanism of CFU-E cells in the presence of parameter uncertainty would enable us
408 to control their population level responses through their signal execution modes. This could
409 pave the way for model-driven hypotheses generation to control cellular states and state
410 transitions. Therefore, here, we employ this system with the provided parameter distributions
411 (Figure 5E) and apply our proposed dynamic network analysis framework to unveil the

underlying mechanism behind this behavior by identifying the JAK2/STAT5 signaling network execution modes. Moreover, we analyze how the number of modes varies with respect to the increasing Epo levels and demonstrate the time-dependent dominant pathways that play role in the increase of pSTAT5 levels (Figure 5) and hence the cell survival responses.

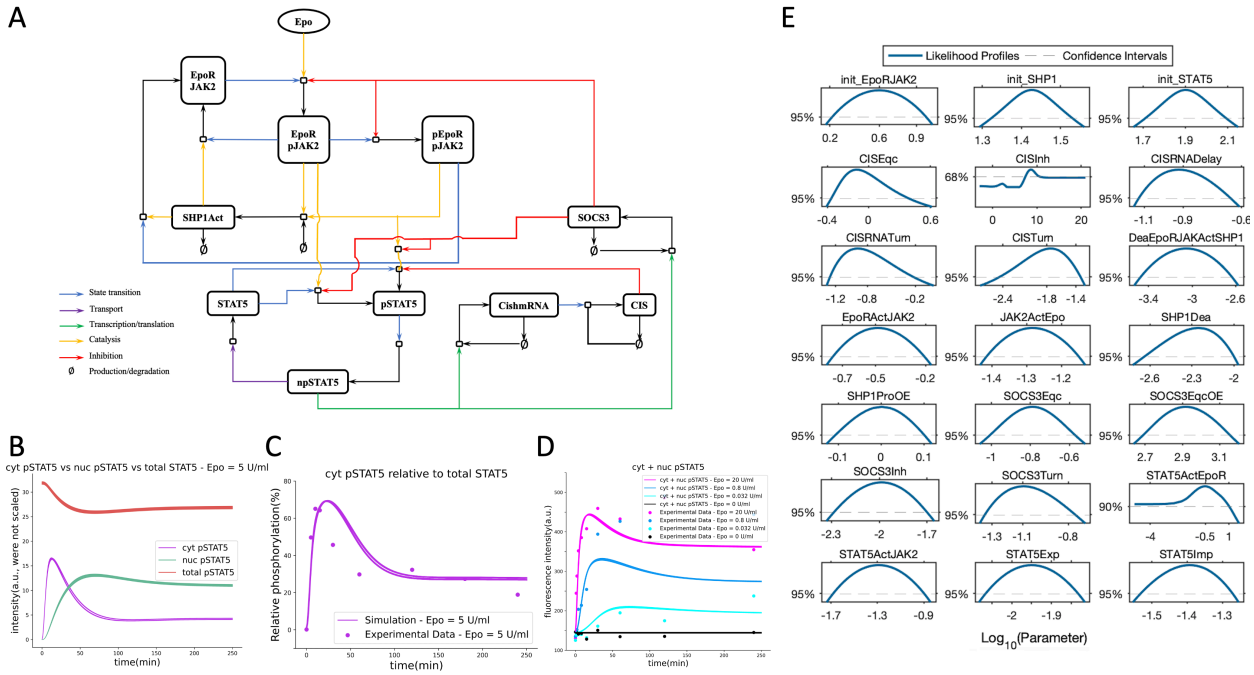

Figure 5. JAK2/STAT5 population-average model, model parameters, and STAT5 responses that were reproduced using the calibrated model provided by Adlung et al.[55]. A) Graphical representation of the JAK2/STAT5 signaling network model. B) Comparison of average total STAT5, pSTAT5 in cytoplasm (cyt), and pSTAT5 in nucleus (nuc) responses of CFU-E cell population over time when exposed to Epo = 5 U/ml. C) Simulated pSTAT5 (cyt) relative to total STAT5 versus the experimental measurements. D) Overall pSTAT5 responses over time with respect to increasing Epo doses versus corresponding experimental data. Note: The experimental data and the model parameters were provided by Adlung et al.[55]. 1907 parameter sets were simulated in total and plotted all together. As seen in B) and C), all parameter sets are fitting the data equally well and their trajectories are mostly overlapping. E) Likelihood profiles of the 21 dynamic model parameters[55].

The JAK2/STAT5 signaling network execution modes and the dominant pathways over time vary after stimulating with different Epo doses, which correlate with the cell survival rates

To apply the proposed dynamic network analysis framework to the JAK2/STAT5 signaling network, we first implemented the model in Python using the PySB modeling tool[39], which was further used to reproduce the network responses (Figure 5B-D). The model consists of 21 dynamic parameters as well as several offset and scaling parameters, in which the parameters have uncertainty with wide likelihood-based confidence intervals (Figure 5E) and equally good fitness to experimental data[55] (Figure 5C-D). Among the provided parameter distributions, we picked the best-fitting 1907 parameter vectors and used them to analyze the network mechanistically.

First, we identified pSTAT5 as the target node because we are interested in the mechanism that causes the increase of pSTAT5 in both cytoplasm and nucleus which is correlated with the cell

442 survival responses. For illustration purposes, we pruned the JAK2/STAT5 signaling network by
443 removing inhibitory interactions and created the production network highlighting only positive
444 reactions (Figure 6A). Next, we computed the reaction rates that produce pSTAT5 and tracked
445 the signal flux to create the dominant subnetwork at each time point. Then, we labeled the
446 identified unique dominant subnetworks by a unique integer and obtained the dynamic
447 fingerprint of the network. Lastly, we repeated these steps for each parameter vector. After
448 computing similarities between each obtained sequence using the LCS metric and clustering
449 them, we observed the different signal execution modes of the network for each Epo dose (see
450 Methods section for more details on each step).
451
452 Interestingly, different Epo doses lead to different execution modes (Figure 6C-G) with various
453 dominant pathways over time (Figure 6B). When Epo doses are low (e.g., Epo = 0, 0.032, and
454 0.8 U/ml), the signal is executed in a single mode, but the modes are slightly different (Figure
455 6C-E). More precisely, when Epo = 0 U/ml (Figure 6C), the only dominant pathway that can
456 induce pSTAT5 production over time is the feedback loop between pSTAT5, npSTAT5, and
457 STAT5 (Figure 6B-1), which depends on the initial STAT5 abundance only. When Epo = 0.032
458 U/ml (Figure 6D), three different pathways that become dominant at different time intervals
459 emerge. To elaborate, at t = 1 min after Epo was applied, JAK2 is activated by the Epo-EpoR
460 complex and eventually phosphorylated receptor complexes (i.e., EpoRpJAK2 and pEpoRpJAK2)
461 are formed. Then, upon involvement of STAT5 with pEpoRpJAK2, initial levels of pSTAT5 are
462 observed (Figure 6B-2). Between t = 2 – 15 min, a cytosolic protein, SHP1, is recruited into the
463 active receptor complexes and dephosphorylates JAK2 to terminate signal transduction (Figure
464 6B-3). Between t = 16 and 250 mins, the signal from medium to cytoplasm and nucleus
465 attenuates and the only pSTAT5 production source becomes the STAT5 feedback pathway
466 (Figure 6B-1). Similarly, when the Epo dose is increased to 0.8 U/ml from 0.032 U/ml, the
467 network executes the signal in a single mode with slightly different dominant pathways (Figure
468 6E). This increase in Epo level reduces the duration of the activity of signal transduction
469 terminator pathway from 14 to 9 mins and additionally activates one of the direct pSTAT5
470 producer pathways (Epo–EpoRpJAK2–pSTAT5; Figure 6B-4) which contributes to the production
471 of more pSTAT5 for 7 mins. However, this activation does not last long due to weak Epo dose
472 and again the STAT5 feedback loop becomes the only dominant pathway producing pSTAT5
473 between t = 18 and 250 min.
474
475 Increasing the Epo level further increases the number of signal execution modes and changes
476 both order and duration of the dominant pathways (Figure 6F-G). When Epo = 5 U/ml, the
477 dominant signal flux begins with the activation of the direct pSTAT5 producing pathway (Figure
478 6B-4) at t = 1 min and switches to the other pSTAT5 producing path through the
479 phosphorylated receptor complex pEpoRpJAK2 (Figure 6B-2) between t = 2 and t = 34 min in
480 the first mode. In the meantime, although the signal transduction terminator pathway through
481 SHP1 becomes dominant for 4 mins, it is not enough to prevent pSTAT5 producing pathways
482 from becoming dominant again because of the strong Epo signaling. First, the STAT5 feedback
483 pathway becomes dominant between t = 35 and t = 39 min and then it is followed by the major
484 dominance of the direct pSTAT5 producing pathway (Figure 6B-4) from t = 40 mins until the end
485 of the simulation, maintaining high levels of pSTAT5. The same scenario occurs in the second

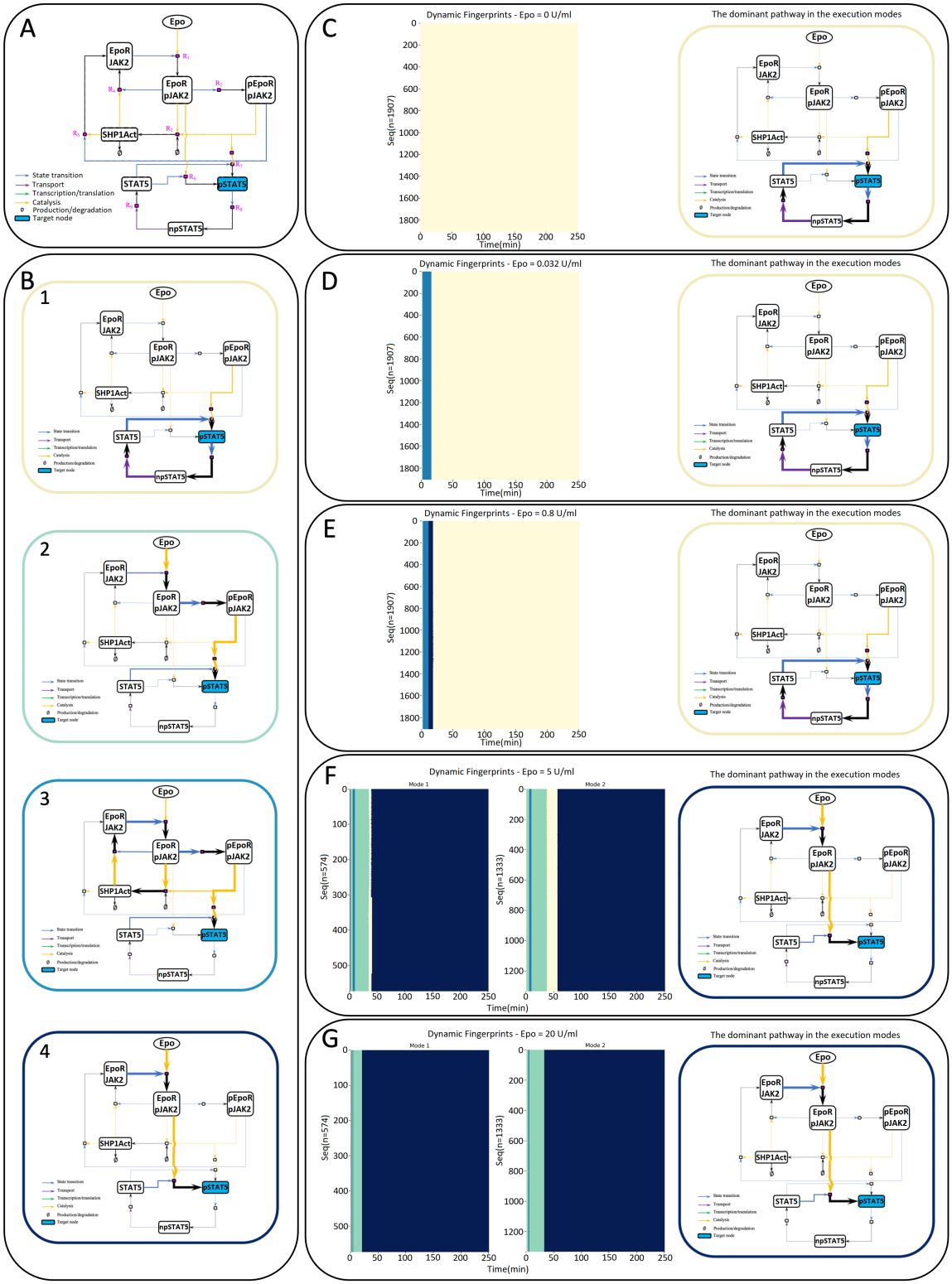

Figure 6. Execution modes of JAK2/STAT5 signaling network for each Epo dose when pSTAT5 is the target node. A) The pruned JAK2/STAT5 signaling network. The inhibitory interactions were removed for demonstration purposes to show the positive reactions only. B) Dominant pathways over time framed with the corresponding colors in the identified execution modes in C-G. C-G) Dynamic fingerprints of the network for each parameter set (each row corresponds to a fingerprint sequence for a unique parameter set) when Epo = 0 U/ml, 0.032 U/ml, 0.8 U/ml, 5 U/ml, and 20 U/ml with their most common dominant pathways, respectively.

492 execution mode as well. The only difference between the two modes is the duration of the
493 dominance of the STAT5 feedback pathway. More precisely, the STAT5 feedback pathway is
494 dominant about 4 mins in mode 1 whereas it is dominant about 20 mins in the second mode,
495 which does not result in an observable difference in the abundance of pSTAT5. When Epo is
496 applied at the level of 20 U/ml (Figure 6G), both execution modes have the same dominant
497 pathways that directly produce cytoplasmic pSTAT5 (Figure 6B-2 and 4). Only the duration of
498 the dominance of the pathway through pEPoRpJAK2 receptor complex lasts longer in the
499 second mode, which results in barely observable, long lasting pSTAT5 peak levels.
500
501 The change in the signal execution modes and the temporal dominant pathways with respect to
502 increasing Epo doses may mechanistically imply the increase in the cell survival percentages.
503 When the Epo level ranges from 0 to 20 U/ml, the percentage of surviving cells gradually
504 increases from 6.9% to 82%, inferring to a correlation with the high responses of pSTAT5 (Figure
505 5D)[55]. As elaborated above, increasing the Epo level results in quantitatively and qualitatively
506 different signal execution modes. Comparing all signal execution modes in Figure 6C-G indicates
507 that when Epo levels are low, the main dominant pathway contributing to the pSTAT5
508 production is the STAT5 feedback pathway in which the production rate mostly depends on the
509 initial abundance of STAT5 and is weakly affected by the Epo-induced signaling, resulting in
510 lower cell survival rates (Figure 6C-E). On the other hand, when the applied Epo dose becomes
511 higher, the cells mechanistically switch to another state. More precisely, the main dominant
512 pathways over time become the direct pSTAT5 producing pathways through Epo receptor
513 complexes and phosphorylated JAK2 (Figure 6F-G), which results in a more robust pSTAT5
514 production and eventually causes higher cell survival rates. The relationship between varying
515 Epo doses, the signal execution modes, and the resulting pSTAT5 trajectories are further
516 demonstrated in Supplemental Figure 10.
517
518 In summary, upon applying the proposed framework to the JAK2/STAT5 signal transduction
519 pathway model of Adlung et al.[55], we reveal novel, detailed explanations on how different
520 levels of Epo signaling are mechanistically processed by the CFU-E cells and concluded with a
521 population level cellular decision, i.e., survival or death, by identifying its switch-like signal
522 execution modes as presented in Figure 6. We hypothesize that these Epo-dependent
523 execution modes might be further used to precisely control CFU-E cells population level
524 decisions by specifically targeting the associated dominant pathways through experiments.
525 Furthermore, this analysis show that the proposed approach might be used to reveal broad
526 information about signal flow in the other complex systems and perhaps to elicit the ways of
527 reprogramming the cells and control their fates through their unique signal execution modes.
528
529 **Discussion**
530
531 It has been long recognized that model parameter optimization to experimental data is key to
532 investigate the dynamical properties that control cell behavior[66]. Unfortunately, parameter
533 optimization usually yields large parameter uncertainties due to a general lack of quantitative
534 data as well as model identifiability[67]. Even a complete set of time course data is insufficient to
535 constraint most rate parameters[9]. In this work, we examined the effects of parameter

536 uncertainty on signal execution through a biochemical network with an approach inspired in
537 Tropical Geometry and Ultradiscretization Theory. Despite the many parameter vectors which
538 reproduce the experimental protein dynamics, we found that the signal flow in a network was
539 constrained to only a few modes of execution. Our analysis further shows that within a
540 Bayesian calibration scheme, it is possible to assign probabilities to each execution mode, thus
541 greatly improving our understanding of signal dynamics. Therefore, the probabilistic approach
542 introduced in this work could open a novel perspective to understand network-driven
543 processes from a statistical perspective.
544
545 To introduce the proposed framework, we first studied a simplified extrinsic apoptosis reaction
546 model (aEARM) that comprises key biological properties of apoptosis and allows us to analyze
547 all aspects of our method thanks to convergence of all model parameters after parameter
548 calibration. Despite the found 27,242 unique parameter vectors with equally good data fitness
549 after model calibration with Bayesian methods (Figure 1B), only three distinct execution modes
550 emerged as signal processing mechanisms (Figure 3). To understand the functional differences
551 between the execution modes, we performed *in silico* perturbation experiments. Interestingly,
552 once we knockdown the effector caspase (eCaspase) by 50%, each execution mode exhibited
553 considerably different time of death (ToD) responses compared to wild type (Figure 4). This
554 analysis shows that parameter uncertainty as a result of model calibration can be mapped to
555 signal execution modes that respond differently to perturbations, thus demonstrating that
556 using a single best fit parameter vector is insufficient to understand signal dynamics in complex
557 models.
558
559 In addition to the aEARM, we studied the signal execution modes of the extended apoptosis
560 model (EARMv2.0)[39], that provides more detailed resolution to apoptosis mechanism, to show
561 feasibility of the proposed framework on a larger model with higher parameter uncertainty,
562 whose details and results are provided in the Supplemental Information. While both EARM
563 models represent the same mechanism and exhibit significant overlap in terms of capturing the
564 key features of apoptosis, their resolution and hence the underlying system of differential
565 equations are different. Nevertheless, once we compare the highly probable execution modes
566 observed for both models, we see that there exists a functional overlap between the underlying
567 dominant pathways although they visually seem different (Supplemental Information, Section
568 2). Such functional commonality in the observed execution modes may imply that the execution
569 modes are intrinsic features of the underlying system and independent of the specific way in
570 which the model is implemented.
571
572 Next, we applied the proposed framework to JAK2/STAT5 signal transduction network (Figure
573 5A) in erythroid progenitor cells to reveal mechanistic explanation for their survival responses
574 in population level, which is correlated with the phosphorylated STAT5 (pSTAT5) abundance
575 with respect to increasing erythropoietin (Epo) signaling[55]. Interestingly, lower Epo doses
576 resulted in a single execution mode whereas high Epo doses resulted in two execution modes,
577 with different dominant pathways over time (Figure 6). Analysis of the execution modes
578 revealed that when the Epo dose is low, the majorly dominant pSTAT5 producing subnetwork is
579 the STAT5 feedback pathway in which pSTAT5 production depends on the initial STAT5 level

580 and is weakly affected by Epo signaling (Figure 6C-E), causing lower survival rates. Conversely,
581 when Epo dose is high, the cells switch to another phase and pSTAT5 production occurs through
582 Epo receptor complexes and phosphorylated JAK2 (Figure 6F-G), that results in robust pSTAT5
583 production and hence higher cell survival rates.
584
585 Although our approach provided novel insights about signal execution in important biological
586 networks, it has certain limitations. Our analysis assumes that reactions with high flux are the
587 most important for signal processing in a network. However, this may not always be the case
588 for other networks or for networks with temporal changes in model topologies[68]. Although our
589 approach is computationally expensive, particularly as models increase in size, requiring
590 hundreds of thousands of parameter samples to reach a convergence criterion, we believe this
591 is a relatively small price to pay in contrast to the number of experiments that would be
592 necessary to attain the same level of mechanistic knowledge about a network-driven process.
593
594 Overall, the proposed framework is advantageous in the analysis of data-driven complex
595 network models by providing broad information about signal flow and how the system of
596 interest mechanistically executes the signals in the presence of parameter uncertainty. In
597 addition, the knowledge of the dominant pathways contributing to the production (or
598 consumption) of a target protein over time could guide the researchers through experiments
599 and allow easy identification of the other experimental targets to test hypotheses about the
600 target protein and the pathway itself. Also, identification of signal execution modes might
601 provide guidance to precisely control the cell behaviors, cellular states, state transitions, and
602 perhaps their fates when they are exposed to an input signal. Furthermore, the information
603 about signal flow could be used to study drug-induced network rewiring processes[69], provide
604 mechanistic explanations to drug responses, and predict sequential combinations of drugs that
605 could better modulate a response signal in biochemical networks.

Acknowledgements
The authors would like to thank Erin Shockley, Alexander Lubbock, Leonard Harris, Vito Quaranta, and Eric Deeds for insightful conversations and critical feedback on this work. This work was supported by the following funding sources: OOO was supported by Vanderbilt International Students Program; CFL was supported by the National Science Foundation (NSF) [MCB 1411482] and NSF CAREER Award [MCB 1942255]; and the National Institutes of Health (NIH) [U54-CA217450 and U01-CA215845].

Availability
All the code to reproduce the figures that contain model calibration, modes of signal execution, visualizations, and hypothesis exploration, is open source and can be found as Jupyter Notebooks in this GitHub repository: https://github.com/LoLab-VU/pydyno. These shareable and reusable notebooks contain all the source code and markup text that explains the rationale for each step in the analysis.

Author Contributions
O.O.O. developed the methods, performed the simulations and computations, and wrote the manuscript.
M.O. developed the methods, performed the simulations and computations, and wrote the manuscript.
B.A.W. contributed to and discussed the research topics and edited the manuscript.
J.C.P. contributed to and discussed the research topics and edited the manuscript.
M.W.I. contributed to and discussed the research topics and edited the manuscript.
G.V.I. contributed to and discussed the research topics and edited the manuscript.
S.P.G. contributed to and discussed the research topics and edited the manuscript.
C.F.L. conceived the ideas and concepts, developed the methods, and wrote the manuscript.

Competing Interest
The authors declare that no conflict of interests exist.

635 **References**

636

637 1. Jordan, J. D., Landau, E. M. & Iyengar, R. Signaling networks: The origins of cellular
638 multitasking. *Cell* **103**, 193–200 (2000).

639 2. Weerts, H. H. M., Van den Hof, P. M. J. & Dankers, A. G. Identifiability of linear dynamic
640 networks. *Automatica* **89**, 247–258 (2018).

641 3. Bonneau, R. Learning biological networks: From modules to dynamics. *Nat. Chem. Biol.* **4**,
642 658–664 (2008).

643 4. Janes, K. A., Reinhardt, H. C. & Yaffe, M. B. Cytokine-Induced Signaling Networks
644 Prioritize Dynamic Range over Signal Strength. *Cell* **135**, 343–354 (2008).

645 5. Barzel, B. & Barabási, A. L. Universality in network dynamics. *Nature Phys.* **9**, 673–681
646 (2013).

647 6. Jaqaman, K. & Danuser, G. Linking data to models: Data regression. *Nat. Rev. Mol. Cell*
648 *Biol.* **7**, 813–819 (2006).

649 7. Raue, A., Kreutz, C., Maiwald, T., Klingmuller, U. & Timmer, J. Addressing parameter
650 identifiability by model-based experimentation. *IET Syst. Biol.* **5**, 120–130 (2011).

651 8. Shockley, E. M., Vrugt, J. A. & Lopez, C. F. PyDREAM: High-dimensional parameter
652 inference for biological models in python. *Bioinformatics* **34**, 695–697 (2018).

653 9. Eydgahi, H. *et al.* Properties of cell death models calibrated and compared using Bayesian
654 approaches. *Mol. Syst. Biol.* **9**, 644 (2013).

655 10. Gutenkunst, R. N. *et al.* Universally sloppy parameter sensitivities in systems biology
656 models. *PLoS Comput. Biol.* **3**, 1871–1878 (2007).

657 11. Mitra, E. D. & Hlavacek, W. S. Parameter estimation and uncertainty quantification for
658 systems biology models. *Current Opinion in Systems Biology* **18**, 9–18 (2019).

659 12. Janes, K. A. *et al.* Cell signaling: A systems model of signaling identifies a molecular basis
660 set for cytokine-induced apoptosis. *Science (80-.).* **310**, 1646–1653 (2005).

661 13. Becker, V. *et al.* Covering a broad dynamic range: Information processing at the
662 erythropoietin receptor. *Science (80-.).* **328**, 1404–1408 (2010).

663 14. Albeck, J. G., Burke, J. M., Spencer, S. L., Lauffenburger, D. A. & Sorger, P. K. Modeling a
664 Snap-Action, Variable-Delay Switch Controlling Extrinsic Cell Death. *PLoS Biol.* **6**, e299
665 (2008).

666 15. Münzner, U., Klipp, E. & Krantz, M. A comprehensive, mechanistically detailed, and
667 executable model of the cell division cycle in Saccharomyces cerevisiae. *Nat Commun* **10**,
668 1308 (2019).

669 16. Rukhlenko, O. S. *et al.* Dissecting RAF inhibitor resistance by structure-based modeling
670 reveals ways to overcome oncogenic RAS signaling. *Cell Syst.* **7**, 161–179 (2018).

671 17. Erickson, K. E., Rukhlenko, O. S., Shahinuzzaman, M., Slavkova, K. P., Lin, Y. T., Suderman,

672 R., *et al.* Modeling cell line-specific recruitment of signaling proteins to the insulin-like
673 growth factor 1 receptor. *PLoS Comput Biol* **15**, e1006706 (2019).

674 18. Zhang, Q. *et al.* NF-κB dynamics discriminate between TNF doses in single cells. *Cell Syst.*
675 **5**, 638–645 (2017).

676 19. Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **27**, 379–423
677 (1948).

678 20. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information
679 Transduction Capacity of Noisy Biochemical Signaling Networks. *Science (80-.).* **334**, 354–
680 358 (2011).

681 21. Suderman, R., Bachman, J. A., Smith, A., Sorger, P. K. & Deeds, E. J. Fundamental trade-
682 offs between information flow in single cells and cellular populations. *Proc. Natl. Acad.*
683 *Sci.* **114**, 5755–5760 (2017).

684 22. Shockley, E. M., Rouzer, C. A., Marnett, L. J., Deeds, E. J. & Lopez, C. F. Signal integration
685 and information transfer in an allosterically regulated network. *npj Syst. Biol. Appl.* **5**,
686 (2019).

687 23. Levchenko, A. & Nemenman, I. Cellular noise and information transmission. *Curr. Opin.*
688 *Biotechnol.* **28C**, 156–164 (2014).

689 24. Brennan, M. D., Cheong, R. & Levchenko, A. Systems biology. How information theory
690 handles cell signaling and uncertainty. *Science* **338**, 334–5 (2012).

691 25. Varma, A. & Palsson, B. O. Stoichiometric flux balance models quantitatively predict
692 growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Applied
693 and environmental microbiology 60, 3724–3731 (1994).

694 26. Lee, J. M., Gianchandani, E. P. & Papin, J. A. Flux balance analysis in the era of
695 metabolomics. Brief. Bioinform. 7, 140–150 (2006).

696 27. Raman, K. & Chandra, N. Flux balance analysis of biological systems: applications and
697 challenges. Brief. Bioinform. 10, 435–449 (2009).

698 28. Gianchandani, E. P., Chavali, A. K. & Papin, J. A. The application of flux balance analysis in
699 systems biology. WIREs Syst. Biol. Med. 2, 372–382 (2010).

700 29. Anand, S., Mukherjee, K. & Padmanabhan, P. An insight to flux-balance analysis for
701 biochemical networks. Biotechnology and Genetic Engineering Reviews 36, 32–55 (2020).

702 30. Price, N. D. *et al*. Genome-scale microbial in silico models: the constraints-based
703 approach. *Trends Biotechnol.* **21**, 162–169 (2003).

704 31. Shastri A. A. & Morgan J. A. Flux balance analysis of photoautotrophic metabolism.
705 *Biotechnol. Prog.* **21**, 1617–1626 (2005).

706 32. Ramakrishna R. *et al*. Flux-balance analysis of mitochondrial energy metabolism:
707 consequences of systemic stoichiometric constraints. *Am J. Physiol. Regul. Integr. Comp.*
708 *Physiol.* **280**, R695–R704 (2001).

709 33. Izallalen M. *et al*. Geobacter sulfurreducens strain engineered for increased rates of
710 respiration. *Metab. Eng.* **10**, 267–275 (2008).

711 34. Kauffman, K. J., Prakash, P. & Edwards, J. S. Advances in flux balance analysis. *Curr. Opin.*
712 *Biotechnol.* **14**, 491–496 (2003).

713 35. Kato, T., Tsujimoto, S. & Zuk, A. Spectral Analysis of Transition Operators, Automata
714 Groups and Translation in BBS. *Commun. Math. Phys.* **350**, 205–229 (2017).

715 36. Noel, V., Grigoriev, D., Vakulenko, S. & Radulescu, O. Tropical geometries and dynamics
716 of biochemical networks. Application to hybrid cell cycle models. *Elect. Notes in Theo.*
717 *Comp. Sci.* **284**, 75–91 (2012).

718 37. Elmore, S. Apoptosis: A Review of Programmed Cell Death. *Toxicol. Pathol.* **35**, 495–516
719 (2007).

720 38. Koonin, E. & Aravind, L. Origin and evolution of eukaryotic apoptosis: the bacterial
721 connection. *Cell Death Differ.* **9**, 394–404 (2002).

722 39. Lopez, C. F., Muhlich, J. L., Bachman, J. A. & Sorger, P. K. Programming biological models
723 in Python using PySB. *Mol. Syst. Biol.* **9**, 1–19 (2013).

724 40. Kantari, C. & Walczak, H. Caspase-8 and Bid: Caught in the act between death receptors
725 and mitochondria. *Biochim. Biophys. Acta - Mol. Cell Res.* **1813**, 558–563 (2011).

726 41. Özören, N. & El-Deiry, W. S. Defining characteristics of types I and II apoptotic cells in
727 response to TRAIL. *Neoplasia* **4**, 551–557 (2002).

728 42. Kaufmann, S. H., Desnoyers, S., Ottaviano, Y., Davidson, N. E. & Poirier, G. G. Specific
729 Proteolytic Cleavage of Poly(ADP-ribose) Polymerase: An Early Marker of Chemotherapy-
730 induced Apoptosis. *Cancer Res.* **53**, 3976–3985 (1993).

731 43. Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M. & Sorger, P. K. Non-genetic origins of
732 cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**, 428–432 (2009).

733 44. Zhou, H. X. Rate theories for biologists. *Q. Rev. Biophys.* **43**, 219–293 (2010).

734 45. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences.
735 *Stat. Sci.* **7**, 457–472 (1992).

736 46. Chiband, S. & Greenberg, E. Understanding the Metropolis-Hastings Algorithm. **49**, 327–
737 335 (2008).

738 47. Kato, T., Tsujimoto, S. & Zuk, A. Spectral Analysis of Transition Operators, Automata
739 Groups and Translation in BBS. *Commun. Math. Phys.* **350**, 205–229 (2017).

740 48. Bergroth, L., Hakonen, H. & Raita, T. A survey of longest common subsequence
741 algorithms. *Proc. - 7th Int. Symp. String Process. Inf. Retrieval, SPIRE 2000* 39–48 (2000).

742 49. Studer, M. & Ritschard, G. What matters in differences between life trajectories: a
743 comparative review of sequence dissimilarity measures. *J. R. Stat. Soc.* A **179**, 481–511
744 (2016).

745 50. Rokach, L. & Maimon, O. Clustering methods. *Data Min. Knowl. Discov. Handb.* 321–352

746 (2005). doi:10.1007/0-387-25465-X_15

747 51. Mehal, W. Z., Inayat, I. & Flavell, R. A. Caspases 3 and 7: Key Mediators of Mitochondrial
748 Events of Apoptosis. *Science (80-.).* 847–851 (2006).

749 52. Solania, A., González-Paéz, G. E. & Wolan, D. W. Selective and Rapid Cell-Permeable
750 Inhibitor of Human Caspase-3. *ACS Chem. Biol.* **14**, 2463–2470 (2019).

751 53. Perry, D. K. *et al.* Zinc is a potent inhibitor of the apoptotic protease, caspase-3: A novel
752 target for zinc in the inhibition of apoptosis. *J. Biol. Chem.* **272**, 18530–18533 (1997).

753 54. Bachmann, J. *et al*. Division of labor by dual feedback regulators controls JAK2/STAT5
754 signaling over broad ligand range. *Mol. Syst. Biol.* **7**, 516 (2011).

755 55. Adlung, L. *et al*. Cell-to-cell variability in JAK2/STAT5 pathway components and
756 cytoplasmic volumes defines survival threshold in erythroid progenitor cells. *Cell Reports*
757 **36** (2021).

758 56. Socolovsky, M. *et al*. Ineffective erythropoiesis in Stat5a$^{-/-}$5b$^{-/-}$ mice due to decreased
759 survival of early erythroblasts. *Blood* **98**, 3261–3273 (2001).

760 57. Nijhof, W. & Wierenga, P. K. Isolation and characterization of the erythroid progenitor
761 cell: CFU-E. *J. Cell Biol.* **96**, 386–392 (1983).

762 58. Koulnis, M., Porpiglia, E., Hidalgo, D. & Socolovsky, M. Erythropoiesis: from molecular
763 pathways to system properties. *Adv. Exp. Med. Biol.*, In: Corey, S., Kimmel, M. & Leonard
764 J. (eds) A Systems Biology Approach to Blood. *Advances in Exp. Med. and Biol.*, 844.
765 Springer, New York, NY (2014).

766 59. Swameye, I. *et al*. Identification of nucleocytoplasmic cycling as a remote sensor in
767 cellular signaling by databased modeling. *Proc. Natl. Acad. Sci. USA* **100**, 1028–1033
768 (2003).

769 60. Klingmüller, U. *et al*. Multiple tyrosine residues in the cytosolic domain of the
770 erythropoietin receptor promote activation of STAT5. *Proc. Natl. Acad. Sci. USA* **93**,
771 8324–8328 (1996).

772 61. Gouilleux, F. *et al*. Prolactin, growth hormone, erythropoietin and granulocyte-
773 macrophage colony stimulating factor induce MGF-Stat5 DNA binding activity. *EMBO J.*
774 **14**, 2005–2013 (1995).

775 62. Socolovsky, M. *et al*. Fetal anemia and apoptosis of red cell progenitors in
776 Stat5a$^{-/-}$5b$^{-/-}$ mice. *Cell* **98**, 181–191 (1999).

777 63. Yoshimura, A. *et al*. A novel cytokine-inducible gene CIS encodes an SH2-containing
778 protein that binds to tyrosine-phosphorylated interleukin 3 and erythropoietin receptors.
779 *EMBO J.* **14**, 2816–2826 (1995).

780 64. Sasaki, A. *et al*. CIS3/SOCS-3 suppresses erythropoietin (EPO) signaling by binding the
781 EPO receptor and JAK2. *J. Biol. Chem.* **275**, 29338–29347 (2000).

782 65. Neel, B. G., Gu, H. & Pao, L. The 'Shp'ing news: SH2 domain-containing tyrosine
783 phosphatases in cell signaling. *Trends Biochem. Sci.* **28,** 284–293 (2003).

784 66. Read, M. N., Alden, K., Timmis, J. & Andrews, P. S. Strategies for calibrating models of
785 biology. *Brief. Bioinform.* **21**, 24–35 (2018).

786 67. Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J. A. & Blom, J. G. Systems biology:
787 Parameter estimation for biochemical models. *FEBS J.* **276**, 886–902 (2009).

788 68. Klinke, D. J. Signal transduction networks in cancer: Quantitative parameters influence
789 network topology. *Cancer Res.* **70**, 1773–1782 (2010).

790 69. Lee, M. J. *et al.* Sequential application of anticancer drugs enhances cell death by
791 rewiring apoptotic signaling networks. *Cell* **149**, 780–794 (2012).

792