

**Simulation and Optimization Tools to Study
Design Principles of Biological Networks**

by

Bambang Senoaji Adiwijaya

B.S., University of Wisconsin, Madison, WI (1995)
M.S., Carnegie Mellon University, Pittsburgh, PA (1996)

Submitted to the Biological Engineering Division
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2006

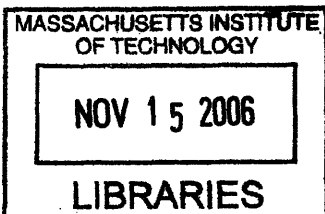
© Massachusetts Institute of Technology 2006. All rights reserved.

Author *B. Senoaji*
Biological Engineering Division
June 9, 2006

Certified by *Bruce Tidor*
Professor of Biological Engineering and Computer Science
Thesis Supervisor

Certified by *Paul I. Barton*
Professor of Chemical Engineering
Thesis Supervisor

Accepted by *Alan J. Grodzinsky*
Chairman, Department Committee on Graduate Students




ARCHIVES

Thesis Committee

- / / / /

Accepted by.....



Douglas A. Lauffenburger

Whitaker Professor of Biological Engineering, Chemical Engineering, and Biology

Chairman of Thesis Committee

Accepted by.....

Paul I. Barton

Professor of Chemical Engineering

Thesis Supervisor

Accepted by.....

Drew Endy

Assistant Professor of Biological Engineering

Thesis Committee Member

Accepted by.....

Peter Sorger

Professor of Biology and Biological Engineering

Thesis Committee Member

Accepted by.....

Bruce Tidor

Professor of Biological Engineering and Computer Science

Thesis Supervisor

To my wife Caroline,
to my daughter Frances,
and to my parents

Simulation and Optimization Tools to Study Design

Principles of Biological Networks

by

Bambang Senoaji Adiwijaya

Submitted to the Biological Engineering Division
on June 9, 2006, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

Recent studies have developed preliminary wiring diagrams for a number of important biological networks. However, the design principles governing the construction and operation of these networks remain mostly unknown. To discover design principles in these networks, we investigated and developed a set of computational tools described below. First, we looked into the application of optimization techniques to explore network topology, parameterization, or both, and to evaluate relative fitness of networks operational strategies. In particular, we studied the ability of an enzymatic cycle to produce dynamic properties such as responsiveness and transient noise filtering. We discovered that non-linearity of the enzymatic cycle allows more effective filtering of transient noise. Furthermore, we found that networks with multiple activation steps, despite being less responsive, are better in filtering transient noise. Second, we explored a method to construct compact models of signal transduction networks based on a protein-domain network representation. This method generates models whose number of species, in the worst case, scales quadratically to the number of protein-domain sites and modification states, a tremendous saving over the combinatorial scaling in the more standard mass-action kinetic models. The method was applied to examine determinants of specificity in yeast mating and filamentous growth pathways. In these pathways, a mass-action model was estimated to consist of more than 10^7 species and was too large to simulate; however, a simplified model consists of only 132 state variables and produced intuitive behavior. The resulting models were utilized to study the roles of a scaffold protein and of a shared binding domain to pathway functions.

Thesis Supervisor: Bruce Tidor

Title: Professor of Biological Engineering and Computer Science

Thesis Supervisor: Paul I. Barton

Title: Professor of Chemical Engineering

Acknowledgments

Preparation of this thesis has been a difficult but rewarding journey for me. First and foremost, I am indebted to my wife, Caroline H. Adiwijaya for her love and support.

I would like to thank Bruce Tidor, for his mentorship during my graduate studies and freedom to pursue a new direction of research. My deep appreciation also goes to Paul I. Barton, who has been a wonderful resource to obtain wealth of technical and computational discussions throughout the development of this thesis.

My gratitude goes to my thesis committee members, Douglas A. Lauffenburger, Drew Endy, and Peter Sorger. A special thanks to Doug, whose class project started me to work on the role of scaffold proteins in MAP kinase networks, and whose comments are deeply appreciated throughout the committee meetings; to Drew, who provided many constructive feedback during many thesis committee meeting; and to Peter, with whom I discussed the biological aspect of this thesis initially.

I enjoy and appreciate my friends in the Tidor Lab for stimulating discussions and precious feedbacks during numerous group meetings. In particular, to Joshua Apgar and Caitlin Bever, for numerous discussions and for being such wonderful officemates; A. Katharina Wilkins, Patricio Ramirez, and Jared Toettcher for multiple discussions of systems biology; Jay Bardhan, Michael Altman, David Green, Brian Joughin, Shaun Lippow, Kathryn Armstrong, Mala Radhakrishnan, Aurore Zyto, and Bracken King, for much feedback.

Furthermore, my special gratitude goes to members of MIT Process Systems Engineering Lab, in particular Cha Kun Lee, Alexander Mitsos, Adam Singer, John Tolsma, and Panayiotis Lemonidis, for your resourcefulness.

I would like to thank Ty Thompson, for thoughtful discussion and help on the biology of the yeast mating and filament growth pathways.

Last but not least, my deepest appreciation goes to my parents and family, for their visions to educate their children, patience with my long years of education, and many sacrifices to provide me with the best possible opportunities.

Contents

1	Introduction	23
1.1	Mitogen-Activated Protein (MAP) kinase network as a paradigmatic example of signal transduction networks.	23
1.2	Modeling as a tool to study design principles of MAP kinase networks.	26
1.3	Optimization tools to probe MAP kinase network models.	28
1.4	Thesis outline.	31
2	Biological Network Design Principles: Discovery through Dynamic Optimization	33
2.1	Abstract	33
2.2	Introduction	34
2.3	Framework and Methods	36
2.4	Results and Discussion	42
3	Differentiating signal from transient noise in non-linear signal transduction networks	55
3.1	Abstract	55
3.2	Introduction	56
3.3	Methods	60
3.4	Results	67
3.5	Discussion	77
4	Compact Modeling through Protein Domain Network Representa-	

tion Applied to Study Signaling Specificity	81
4.1 Introduction	82
4.2 Methods	84
4.3 Results	94
4.4 Discussion	110
4.5 Supplementary Materials	116
4.5.1 Models for System I	116
4.5.2 Models for System II	121
4.5.3 Lumping approach with linear inverse lumping applied to Sys- tem I	122
4.5.4 Proof for molar fractional expression to compute protein com- plexes in a network of larger sizes with independent binding .	123
4.5.5 Proof for the nonlinear reverse lumping	126
5 Conclusion and Future Directions	129
A Tables	133

List of Figures

- 1-1 Models of molecular interactions of multiple MAP kinase networks: yeast pheromone [54], yeast filamentous growth [4], yeast hyper-osmosis [116], mammalian ERK [17], and two instances of mammalian JNK [105]. The network superstructure, shown at the far right, includes the union of all of functional connections from the six specific network models to its left. Any of the specific models can be recovered by switching off proteins (nodes) and reactions (arcs) in the superstructure. 25
- 2-1 Alternative representations of the fundamental enzymatic modification and unmodification reactions. (a) Schematic of a one-step modification and unmodification reaction network; (b) alternative representation of the network given in (a), with emphasis on conservation of A ; (c) chemical reaction network representation of the network given in (a); (d) mass-action dynamic equations of the network given in (a); (e) schematic of a two-step modification and unmodification network; (f) schematic of a two-layer cascade with one-step modification and unmodification in each layer; (g) schematic of a three-layer cascade network with two-step modification and unmodification in each layer. 38

2-2 The definition of the response-time (f_{response}) and output-amplitude ($[\mathcal{O}]_{\text{ss}}$) objectives. (a) A square input trajectory was used to stimulate each network; (b) an output trajectory after stimulation; rise- and decay-time objectives are related to the early and later grey regions, respectively. (c) the rise, decay, and response-time objective trajectories after stimulation; (d) the mathematical definition of rise, decay, and response-time objectives. 40

2-3 Trade-off curve of response-time and output-amplitude objectives, along with representative optimal networks. Legends on each optimal network: green, the rate is maximized to upper bound; red, the rate is minimized to lower bound; blue, the rate is optimized to a value between both bounds (the optimal value of $\log_{10}(\mu_i)$ is indicated); black-dotted, the rate is insensitive in the regions explored. Asterisk, when network is parameterized with the canonical MAP kinase values obtained from [28]. Strategy One network examples are enclosed in a red box, Strategy Two in a blue box, and Strategy Three in a green box. 43

2-4 One-dimensional parametric perturbation of response-time and amplitude objectives for four strategies. Legends: red, Strategy One (with amplitude of $10^{-4} \mu M$); blue, Strategy Two (with amplitude of $10^0 \mu M$); green, Strategy Three (with amplitude of $20 \mu M$); and black, the base strategy, with all “forward” reactions being maximized and all “opposing” reactions being minimized. The objective values are normalized to the optimal values of each respective strategy. The symbols represent the unperturbed parameter values. 46

2-5	Decomposition and sensitivity analysis to uncover mechanisms of optimal strategies. (a),(e), response (rise+decay) time optimal networks; (b),(f), rise-time optimal networks; (c),(g), decay-time optimal networks; (d),(h), sensitivities of rise, decay, and response-time objectives to rate constant multiplier μ_1 . (a-d), applied to Strategy One; (e-h), applied to Strategy Three strategy; Dotted line, rise-time; dashed line, decay-time; solid line, response-time.	48
2-6	State and sensitivity trajectories of two alternative strategies of a one-step network. (a) Trajectories for Strategy One; (b) trajectories for Fast Start.	50
2-7	Trade-off curve for response-time and output-amplitude objectives for variant networks. (a) the two-step network given in Figure 2-1e; (b) the two-layer network given in Figure 2-1f. Legends: solid lines, multiple strategies in each network; dotted line, trade-off curve for one-step network.	52
2-8	Response-time optimal kinase cascade networks at various amplitude constraints. (a) the optimal network for $[\mathcal{O}]_{ss} = 10^{-4}$; (b) the optimal network for $[\mathcal{O}]_{ss} = 10^{-3}$; (c) the optimal network for $[\mathcal{O}]_{ss} = 10^{-2}$; (d) the optimal network for $[\mathcal{O}]_{ss} = 10^{-1}$; (e) the optimal network for $[\mathcal{O}]_{ss} = 10^0$; (f) the optimal network for $[\mathcal{O}]_{ss} = 10^1$	53
3-1	Transient noise filtering problem formulation. (a) Illustration of the filtering function. (b) Objectives, which are amplitude, and Rapid but Delayed Response (RDR). (c) The schematic of networks being studied, which are: (i) a one-step enzymatic activation and deactivation reactions, and (ii) a two-step enzymatic activation-deactivation with the output of the first step acting as a substrate for the second step enzymatic reactions. All the topologies assume elementary (uni- and bi-molecular) reactions.	59

3-2 Response trajectories of RDR optimal linear and non-linear enzymatic cycle networks with various lengths. Black, the output trajectory of linear networks with length of 4 (dot-dashed), 10 (short-dotted), and 100 (long-dotted) steps; green, the one of a one-step enzymatic network; red, the one of a two-step enzymatic network. The optimal one-step and two-step enzymatic cycle networks produced a rapid but delayed response (RDR) the sharpness of whose transition can only be matched by a cascade of 10 and 100-step linear reactions in series. . 68

3-3 The relationship between RDR-optimality and transient noise filtering in a two-step enzymatic cycle and a 100-step linear network. (a) The output trajectories when the networks are subjected to inputs of pulse trains with varying duration. Top panel, the input trajectory of pulse trains with varying duration; lower panels, the output response for different network topology and parameterization. Black, input; red, RDR-optimized two-step enzymatic cycle with $t_d = 0.1$; blue, a cascade of 100-step linear reactions; green, RDR-optimized two-step enzymatic cycle with $t_d = 0$; magenta, literature-parameterized two-step enzymatic cycle. (b) The power spectra of a two-step enzymatic cycle and a 100-step linear networks with various parameterizations. Red, two-step RDR optimized enzymatic cycle with $t_d = 0.1$ (solid), $t_d = 1$ (short-dotted), and $t_d = 0.01$ (long-dotted); green, RDR-optimized two-step enzymatic cycle with $t_d = 0$; magenta, literature parameterized two-step enzymatic cycle; black dotted, two-step enzymatic cycle with literature-parameterized activation rates and RDR-optimized deactivation rates. blue, a cascade of 100-step linear reactions; 69

3-4 The dynamic behavior of an RDR optimal two-step cycle network. (a) Sensitivity of response trajectories to input amplitude (F_0). Red, baseline ($F_0 = 20 \mu M$); blue, increasing (solid) or decreasing (dotted) F_0 in $0.5 \mu M$ interval. (b) Sensitivity of response trajectories to reverse enzyme initial condition (R_0). Red, baseline ($R_0 = 16 \mu M$); blue, increasing (solid) or decreasing (dotted) R_0 in $0.1 \mu M$ interval. (c) The steady-state curve of the RDR-optimal two-step enzymatic network shows an all-or-none behavior. 74

3-5 Trade-offs in one- and two-step enzymatic cycle networks. (a) Trade-off curves between RDR and amplitude objectives. Blue, one-step enzymatic cycle network; red, two-step enzymatic cycle network. (b) Optimal networks at various amplitudes. Top, low amplitude ($f_{\text{amplitude}} = 10^{-4}$); middle, medium amplitude ($f_{\text{amplitude}} = 10^0$); bottom, high amplitude ($f_{\text{amplitude}} = 25$). Arrow colors indicate the location of optimal rate constant multipliers relative to bounds: red, at lower bound; green, at upper bound; blue, in-between bounds. The number next to blue arrows indicates the optimal value (in log-scale). Arrow thickness is proportional to the log scale of optimal value, normalized by respective lower and upper bounds. (c) Corresponding trajectories of output (red) and unbound reverse enzyme (blue) at low, medium, and high amplitudes. 75

4-1 Reaction schemes for System I. (a) The mass-action kinetics representation. The network consists of a scaffold protein S with two binding partners E and K ; Protein K may be modified to K_p when both E and K binds to S . Protein complexes are represented as $S(i, j)$, where i indicates states of the first domain of S (0, unbound; 1, bound to E) and j indicates states of the second domain of S (0, unbound; 1, bound to K ; 2, bound to K_p). Solid lines, binding fluxes; short dotted lines, modification fluxes. (b) The protein-domain network representation. D_i^{unb} , an unbound form domain variable i ; $D_{[i,i']}^{\text{bnd}}$, a bound form domain variable of i to its partner i' ; Solid lines, binding fluxes; short dotted lines, modification fluxes; dash-dotted lines, dependencies of a modification specie to bound-state protein domains. 97

4-2 Accuracy quantification of four approaches (Approaches 1-4) to implement DN representation applied to System I. Accuracy is measured by integrated squared difference of a trajectory of an approach to the one of Approach 0 (f^{accuracy}); (a-c), accuracy of a state (K_p) as a function of a network parameter; (f-h), accuracy of a parametric sensitivity ($\frac{\partial K_p}{\partial k_{\text{on}}^{\text{SK}}}$) as a function of a network parameter; (d), trajectories of state K_p at the canonical parameter values; (e), trajectories of state K_p when $k_{\text{off}}^{\text{SK}}$ is 10^{-3} times its canonical value; (i,j), trajectories of parametric sensitivity $\frac{\partial K_p}{\partial k_{\text{on}}^{\text{SK}}}$ at parameters defined in (d) and (e). Legends: dotted black, Approach 0; red, Approach 1; blue, Approach 2; black, Approach 3; green, Approach 4. 99

4-3 The histogram of f^{accuracy} of Approach 1 and 2 applied to System II, for randomized parameter variations in six-orders of magnitude ranges. Legends: white, Approach 1; black, approach 2. 102

4-4	Comparing parametric behavior of Approach 1 to the ones of Approach 0. Both approaches are applied to a system similar to System I with added binding cooperativity (System II) and to a system of dimerizing scaffold proteins with four kinase partners (System III). (a) Contour plot of a responsiveness function (f^{response}) as a function of a cooperative factor (β) and a dissociation rate constant ($k_{\text{off}}^{\text{SK}}$) applied to System II. Top, Approach 1; middle, Approach 0; bottom, both; Legends: blue, Approach 1; red, Approach 0; (b) Contour plot similar to (a) as a function of cooperative factor (β) and an initial condition (S_0). (c) Contour plot for System III as a function of an initial condition (S_0) and a modification rate constant ($k_{\text{cat}}^{\text{SK}2}$). (d) Contour plot similar to (d) as a function of an initial condition (S_0) and a dissociation rate constant ($k_{\text{off}}^{\text{SK}3\text{p}}$).	103
4-5	Accuracy of Approach 1 compared to species elimination method applied to System III. Black, Approach 1 (with 31 state variables); red, reduced MA model with 25/194 species eliminated; yellow, reduced MA model with 100/194 species eliminated; white, reduced MA model with 163/194 species eliminated.	104
4-6	The scope of yeast Mating (M) and Filamentous Growth (FG) pathways studied in this paper. The network consists of a G-protein receptor (Ste4), a cascade of kinases that can be activated in the presence and absence of a scaffold protein (Ste5), two MAP kinases (Fus3 and Kss1) for each pathway, and two transcription factors for each (Ste12 and Tec1). The input to the network is the amount of activated Ste4 and it is shared between both pathways, and the outputs are defined as time-integrated of activated Ste12 homo-dimers binding to Mating gene for M pathway, and hetero-dimer Ste12-Tec1 binding to Filamentous Growth gene for FG pathway.	106

- 4-7 The reaction scheme in domain network representation of the yeast Mating (M) and Filamentous Growth (FG) pathways. Legends: triangle, a protein domain; open circle, an inactive modification state; solid circle, an active modification state; blue line, a binding interaction; red line, a modification reaction; black dotted line, a dependency interaction between bound domains and modification species; &, combination of multiple dependency interactions to compute a modification specie. 107
- 4-8 Yeast Mating (M) and Filamentous Growth (FG) signal specificity as a function of selected protein expressions and pathways input. (a) M and FG responses as a function of Ste5 for a parameterization on a non-dimerizing Ste5 model (System IV); (b) similar as (a) on a dimerizing Ste5 model (System V); (c) similar as (a) but with added cooperative binding between activated Ste7, Ste5, and Fus3, implemented by multiplying relevant dissociation constants by 0.3 fold of values in (a); (d) M and FG responses as a function of Ste7 in System IV. Legends: blue, M response; red, FG response. α^* , normalized input; $Ste5_0$, normalized Ste5 initial condition. Top, M response only; middle, FG response only; bottom, an overlay of both M and FG responses. 108
- 4-9 Perturbation studies to examine the emergent behavior of the Fus3 common domain. (a) Mating response for System IV; (b) activated Fus3; (c) unbound Fus3 common domain; (d) activated Ste12; (e) a complex of Msg5 and Fus3; (f-j) similar to (a-e) for a network with separate binding domain between Ste7 and Fus3. 110

List of Tables

3.1	Perturbation of network structure and initial conditions to examine the mechanism of optimal Rapid but Delayed Response (RDR). LLN: the Length of a Linear reaction Network connected in series producing an equivalent f_{RDR}/t_d value.	72
4.1	Test systems being used to validate and apply protein-domain network representation of dynamic models. Numbers in parentheses indicate the number of fluxes and additional variables implemented for numerical stability. An asterisk in the number indicates an estimate.	95
4.2	Equations applied to System I. Equation 4.11, Mass Action (MA) kinetics models; Equation 4.12, Domain Network (DN) models; Equation 4.13, definition of the DN lumping; Equation 4.14–4.17, four approaches to implement $g_{\hat{m}}$ within the DN approach.	114
4.3	Experimental phenotypic behaviors being used to parameterize models of yeast mating and filamentous growth pathways.	115
4.4	Nonlinear constraint formulation to represent the phenotypic behaviors listed in Table 4.3. M, mating response; FG, filamentous growth response; over, overexpression; Δ , deletion mutation; Fus3-K42R, mutational to Fus3 with reduced activity; Kss1 ^{nobind} , mutation to non-binding Kss1.	115
4.5	Canonical parameter values for System I, II and III	128
A.1	Canonical parameter values in a unit network of one-step activation-deactivation reactions.	133

A.2	Differential equations governing a one-step enzymatic activation-deactivation reaction network	134
A.3	Differential equation governing a two-step enzymatic activation-deactivation reaction network	135

Chapter 1

Introduction

This thesis aims to develop and apply computational tools to uncover design principles in signal transduction networks. We applied the developed tools to the mitogen-activated protein (MAP) kinase networks, a paradigmatic example of eukaryote signal transduction networks, however, the tools are also applicable to the study of other networks. First, we discuss the outstanding questions in MAP kinase networks. An overview of available computational tools to study biological networks follows. And finally, we summarize the thesis outline.

1.1 Mitogen-Activated Protein (MAP) kinase network as a paradigmatic example of signal transduction networks.

A MAP kinase network is a network of interacting protein kinases activated by mitogen and functions to transduce signals from activated membrane surface receptors to transcription factors in the nucleus. The network typically consists of three well conserved protein kinases, each may be phosphorylated upon activation, and in turn, acts as an enzyme for the activation of the next kinase, in essence forming a cascade of phosphorylation reactions. Many detailed reviews about the network's mechanism have been written elsewhere [47, 116, 54, 105, 118]. Further signal mod-

ulation is achieved through phosphatases, enzymes that dephosphorylate activated kinases [95, 81, 127], and through scaffold proteins, proteins with multiple binding domain sites and are thought to be important for imparting specificity across multiple networks [53, 4, 40, 61, 134, 14, 43, 13]. Extensive studies has been done previously on MAP kinase network to uncover many of its detailed molecular mechanism and reaction schemes, however, to construct dynamic models of this network, some of its reaction schemes and many of its kinetic parameters remain incomplete.

Despite the lack of a complete and correct specification, previous works have been able to suggest design principles derived from the MAP kinase network structure, some of which are reviewed here. First, Goldbeter and Koshland [6] suggested that a zero-order sensitivity allows a network of enzymatic activation and deactivation, inspired by a MAP kinase network, to filter small amplitude noises from large amplitude signals by producing a sigmoidal steady-state input-output curve. Huang and Ferrell [28] later showed that this filtering may be enhanced with existence of multiple phosphorylation sites. Furthermore, Ferrell and Machleder [73] found that *Xenopus* oocytes produced an all-or-none activation behavior consistent with the previous sigmoidal input-output curve, and, because the observed data produced a sigmoidal curve whose steepness could not be explained by multiple phosphorylation sites alone, they proposed an additional positive feedback to enhance the sigmoidalness further. Moreover, Gunawardena [60] extent this study to the limit of a large number of phosphorylation sites and demonstrated that multiple phosphorylation sites may create a good thresholding function but produce a poor switch.

Previous computational studies of MAP kinase networks [28, 8, 103, 15] have not considered detailed molecular mechanisms known to the networks. These mechanisms include subtle network structural differences observed in multiple MAP kinase pathways highlighted by different roles of scaffold proteins (summarized in Figure 1-1), mutual phosphorylation of MAP kinase and MAP kinase kinase [125], and sharing of common-domain binding sites among phosphatases, kinases, and MAPK activated protein kinases (MAPK-APKs) [123]. A recent study suggested further that the Ste5 scaffold protein in yeast mating pathway plays an active role in the mating response

modulation via allosteric interactions with its kinase partners [106]. Previous computational studies partially addressed the scaffold-dependent reactions, but they did not examine detailed differences in the roles of scaffold proteins across multiple MAP kinase pathways. These studies include (1) Sternberg and co-workers, who suggested biphasic activation arising from a binding model of a generic scaffold protein to multiple kinases [8] and (2) Rapoport and co-workers, who suggested that prolonged signal duration results if dephosphorylation occurs only outside of scaffold (versus when it occurs everywhere) [103]. These two studies did not investigate the specific behavior of scaffolds in different MAP kinase networks, thus they could not distinguish the significance of specific scaffold roles. One limitation to study networks with scaffolds in details is the combinatorially large number of state variables being needed to construct a mass-action kinetic model. For example, a yeast Ste5 scaffold protein homodimer with three kinase partners contain up to 25,666 binding and modification states [101]. The large number of species to be considered has made a construction of network models with scaffold proteins difficult.

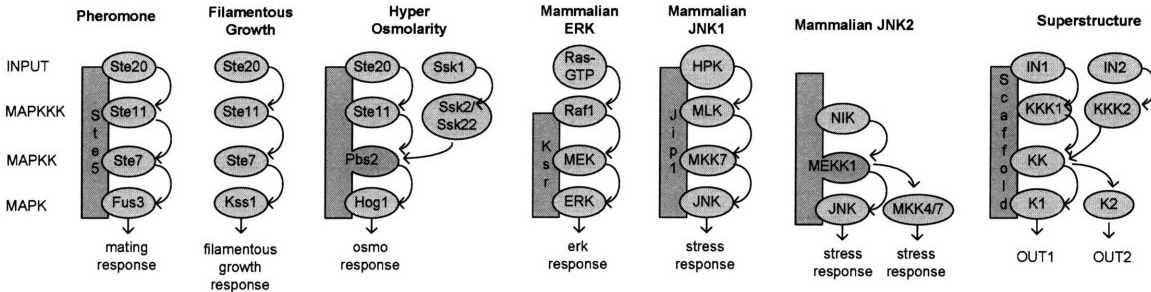


Figure 1-1: Models of molecular interactions of multiple MAP kinase networks: yeast pheromone [54], yeast filamentous growth [4], yeast hyper-osmosis [116], mammalian ERK [17], and two instances of mammalian JNK [105]. The network superstructure, shown at the far right, includes the union of all of functional connections from the six specific network models to its left. Any of the specific models can be recovered by switching off proteins (nodes) and reactions (arcs) in the superstructure.

1.2 Modeling as a tool to study design principles of MAP kinase networks.

Modeling serves as a tool to study signal transduction networks. Mathematical models in general may be used to organize our knowledge of a biological system compactly, develop further understanding of a system's behaviors, generate predictions, and guide intelligently decisions of effective systematic perturbations. Applied to signal transduction networks, modeling may capture a complex systemic behavior from separated knowledge of network pieces and allow further exploration about the significance of some interactions in a network. Different levels of modeling, reviewed in [119], have been introduced, each requires different type of data and may represent different kinds of behaviors. In this thesis, we focus on models that represent dynamic behavior based on detailed mechanistic understanding of a system and its biochemical reaction schemes.

Many models has been developed to describe dynamic behavior of networks in mechanistically detailed systems [1, 15, 31, 55, 22, 128, 94, 69]. These models can be viewed as consisting of two components: molecular mechanisms that portray the qualitative role of players, and dynamic equations that convey the quantitative relationships among the players. Said another way, these models can be depicted as a network, with state variables as nodes, and allowed transitions among the state variables as arcs. In this network representation, the qualitative molecular relationships are expressed by the selection of nodes and arcs for a specific biochemical network, and the quantitation resides in dynamic equations describing the transitions (arcs) that relate changes in the state variables (nodes) and contain adjustable kinetic parameters.

The standard method to model a mechanistically detailed system is the mass-action kinetics, where a system is represented as a network, with nodes being binding and modification states of protein complex species, and arcs being binding and modification reactions among species. The mass-action kinetic method follows simple rules to construct models and has been well documented and applied to study biological

networks [15, 31, 55, 22, 128, 94, 69]; however, this method suffers from two major limitations: first, as a deterministic dynamic method, it cannot represent stochastic behavior in the limit of small amount of reacting components, and second, when applied to a network with scaffold proteins, the method requires a number of nodes and arcs that may be large and scales combinatorially to the number of protein domain sites and modification states. The first limitation has been addressed by modeling with stochastic formulation of reaction systems [35], while methods to address the second limitation will be discussed next.

Various strategies have been proposed to address the large number of species and reactions in mass-action kinetic models. One strategy is to provide a simpler user interface, where modelers only specify interaction of protein domains and lumped parameters, and species and reactions are generated automatically. Two groups have followed this strategy: Blinov and groups [88] developed a software that generate combinatorial species and reactions before simulation, and Lok and groups [76] developed a stochastic differential equation software that only generate species on-the-fly during a simulation. Another strategy is to reduce the size of the dynamic model by applying various model reduction methods (reviewed in [91]) while maximizing accuracy of the reduced models. This strategy is not ideal applied to signal transduction networks for two reasons. First, these methods require construction of the fully enumerated mass action kinetic model (at least initially), which may be intractable for a large network containing scaffold proteins. Second, these methods are accurate for a parameterization, however, they may potentially be inaccurate at different parameterizations. To address the second issue, a formulation based on semi-infinite optimization may search for reduced models that are accurate across parameter ranges, however, this formulation is difficult to solve.

To study and simulate biological networks, it is convenient for the current discussion to consider two separate, but related, classes of variables describing a biological network models: discrete network structure and continuous kinetic parameters. This decomposition allows systematic treatment of biochemical models and appropriate computational tools being applied to the models. Furthermore, this decomposition

follows the level of knowledge in biological systems; in many of the previous models of MAP kinase networks described in Section 1.1, network structures have been well-studied, however, kinetic parameters, such as rate constants and initial conditions, are not as readily available, and thus, simplifying assumptions had to be made. For example, most of the above models utilized Michaelis-Menten kinetics to represent enzymatic reactions (the Michaelis-Menten kinetic requires one less kinetic parameter for each enzymatic reaction), or Huang and Ferrell [28] assumed that dephosphorylation kinetics to be equal to the phosphorylation kinetics because of lack of data to regress the dephosphorylation rate constants. These simplifying assumptions of the parameterizations may potentially produced networks with different behaviors. Furthermore, because the effect of parameterizations and network structures to the overall network behaviors are hard to decouple, such simplifying parameterizations may lead to incorrect conclusions about behaviors of the underlying network structures. Thus, to analyze network models correctly, new methods must be developed to probe the contribution of discrete network structures separately from continuous parameterizations.

1.3 Optimization tools to probe MAP kinase network models.

Optimization is a natural tool to support the modeling process, in particular, to address uncertainty in both network structures and parameterizations. One method to address uncertainty in variables is to assign the variables as optimization variables that are searched for some objectives, be either fitting to experimental data, or others. In the most general formulation when applied to biological network models, the problem falls into a class of problem called a mixed-integer dynamic optimization, which consists of a mixture of two types of optimization variables: discrete (integer) network structure and continuous kinetic parameters, and which is constrained by underlying non-linear dynamic equations.

Previous works to optimize networks can be classified based on the treatment of these two types of variables:

1. Fixed network structures with variable kinetic parameters [22, 128, 48]. In these studies, a fixed network structure was manually selected before optimizing kinetic parameters against experimental data. The choice of network structure in these studies was based on chemical intuition and experimental data, but a systematic study of alternative network topology was not made.
2. Fixed kinetic parameters with manually varied network structures [103]. For example, Rapoport and co-workers [103] compared dynamic simulations of alternative network structure. The work only examined pairwise comparisons of hypothetical structure through manual derivation and simulation. A larger set of topologies could not be studied efficiently; for this system, combinatorial effects might have been overlooked.

Both treatments currently lack availability of global optimization techniques, which not only allows better parameterization in a fixed network structure, but also permits mathematical guarantee when comparing alternative network structures. The guarantee would allow examination of combinatorial number of alternative networks systematically by simultaneous optimization of both network structures and kinetic parameters.

Global optimization techniques play a significant role in optimizing a mixed-integer problem. Global optimization of kinetic parameters would allow us to obtain mathematical guarantees that lead to correct comparisons of network structures. This approach allows us to probe theoretical limits that a given network structure may function and uncover its optimal operational strategy. Without a rigorous global optimization tool, one may mistakenly dismiss a good network structure because of its erroneous parameterizations. In a parallel example of parameter estimation of a physical chemistry reaction scheme [11], local optimization may produce misleading conclusion and erroneous elimination of a reaction scheme, which otherwise is a global optimal solution. Rigorous global optimization techniques to solve dynamic

optimization problem are recently available [132, 18], where both techniques relied on a spatial branch-and-bound technique to prune the continuous search space and derived rigorous objective bounds as a function of parameter ranges. Unfortunately, when applied to MAP kinase network problems, these rigorous techniques have not yet produced bounds with strong elimination power (data not shown). Improved global optimization tools would be beneficial to correct analysis of biological network models.

Optimization of biological networks have broad applications. A traditional application of optimization is in estimating the dynamic parameters from models, which has been applied widely in signal transduction networks [128, 22] although none has rigorous global optimal guarantee of the results. Another area of application is to use optimization to understand operational strategies utilized by a biochemical network, for example, how GTPase may work at different operational regimes [113], or how the network may implement a just-in-time principle in transcription programs of metabolic pathways [9]. The later application demonstrates that we may uncover design principles in biological networks despite having only limited amount of information to regress parameters in biochemical networks.

Evolution of natural networks such as MAP kinase networks may be driven by not only one, but multiple functions. Multiple functions are experienced by a network simultaneously, and thus, may compete for the limited resources available in a natural systems. To uncover realistic understanding design principles in biological networks, multi-objective optimization techniques (reviewed in [98]) should be applied. These techniques would allow probing the trade-offs behaviors experienced by a natural system and be utilized as a framework to compare alternative network structures. However, we are not aware of any works that have applied multi-objective optimization techniques to study design principles of biological networks.

The application of optimization in biological networks may benefit from parallel experience in other systems. First, the decomposition of discrete network structure and continuous kinetic parameters are similar to the problems of side-chain optimization in protein design, where one optimize the discrete amino-acid and continuous

rotamer configurations for either binding affinity or specificity [79, 37, 23]. The need to optimize simultaneously both the discrete and continuous variables is circumvented in the protein design problem by discretizing the continuous space and solving only the discrete optimization, which can be solved efficiently because of bounding heuristics available specific to the problem. Optimizing simultaneously the discrete and continuous variables in biological network models is more difficult to solve because of the lack of efficient bounding procedure. Second, post-optimality analysis is required to extract knowledge from optimization results. A raw optima may possess some variables whose optimal values are insensitive locally, and thus, one would need a method to take accounts these insensitivities. For a quadratic programming problem with convex space, as in protein design charge optimization [80], optimal results may be analyzed by utilizing a Singular Value Decomposition to extract only sensitive dimensions. Similar post-optimality analysis methods for a non-convex problem would be beneficial for biological network models.

1.4 Thesis outline.

In Chapter 2, we developed an optimization-based approach to study design principles in an enzymatic cycle, a sub-network of MAP kinase networks. The enzymatic cycle is probed for a dynamic property, its responsiveness to a stimulus, and for amplification of its output. We developed methods to probe optimal operational strategies that may be utilized by a given network structure based on local optimization with multiple starts to improve solutions towards global optimality. To extract knowledge from the optimal results, we developed methods based on sensitivity analysis at optima and demonstrated utilization of the methods to obtain further insights into the design principles of a network model. Furthermore, to uncover trade-offs experienced by the networks, we applied a multi-objective optimization technique based on parametric optimization with one objective constrained at a value and parameterized its value. Using the trade-off behavior as a framework, we compared multiple network structures to understand functional advantages of a given network construction.

In Chapter 3, we probed MAP kinase networks further for another dynamic property: the ability to filter transient noise. This objective was examined by probing networks for their ability to produce a Rapid but Delayed Response (RDR). We relied on the methods developed in Chapter 2 to look into this other objective and we found that these non-linear networks have tremendous advantages over linear networks with multiple steps. Furthermore, to promote this function, we pinpointed the important role of the phosphatase. Moreover, we established the relationship between this RDR property and the resulting input-output steady-state curve compared to previous studies [6, 28, 73, 60]. Finally, we examined the trade-offs between the ability to produce a sharp RDR and output amplification, another objective, and found that networks with multiple activation steps have advantages for producing a sharper RDR across wider ranges of amplitude.

In Chapter 4, we developed a method to represent dynamics of signal transduction networks compactly, while avoiding the combinatorial explosion in the number of species of the mass-action kinetic modeling method. We examined alternative implementation approaches to produce compact models that accurately represent corresponding mass-action kinetic models. Furthermore, the performance of this approach to model reduction methods are compared when the parameters are varied in a range. Finally, the method is applied to Mating and Filament Growth MAP kinase pathways to uncover significance of detailed mechanisms to the pathway functions, in particular, to reveal the role of Ste5, a scaffold protein, and of a MAPK common domain.

Chapter 2

Biological Network Design Principles: Discovery through Dynamic Optimization

2.1 Abstract

An important challenge in systems biology is the inherent complexity of biological network models, which complicates the task of relating network structure to function and of understanding the conceptual design principles by which a given network operates. Here we investigate an approach to analyze the relationship between a network structure and its functions using the framework of optimization. A common feature found in a variety of biochemical networks involves the opposition of a pair of enzymatic catalyzed chemical modification reactions such as phosphorylation/dephosphorylation or methylation/demethylation. The modification pair frequently adjusts chemical properties, such as activating and deactivating further enzyme function. We applied optimization methodology to study an modification and unmodification network unit commonly found in signal transduction systems, and we explored the use of this methodology to discover design principles. The results demonstrate that different sets of rate constants used to parameterize the same network topology represent different compromises made in the resulting network operating characteristics. Moreover, the same topology can be used to encode different strategies for achieving performance goals. The ability to adopt multiple strategies may lead to significantly improved performance across a range of conditions through rate modulation or evolutionary processes. The optimization framework explored here is a feasible approach to support discovery of design principles in biological networks.

Keywords: biological networks, design principles, optimization, systems biology

2.2 Introduction

The development of quantitative models describing biological networks for a number of interesting systems is being undertaken [120, 56, 1, 27, 69, 128, 22, 7, 15, 28, 94, 84, 30, 86, 121, 16, 78, 104, 41]. These models aim to capture the underlying structure, dynamics, and detailed mechanisms of their experimental counterparts in a manner that recapitulates known behaviors, provides a means for understanding that behavior, and also predicts new behavior. The detail, accuracy, and number of systems for which models are available is expected to grow for the foreseeable future. These models may potentially be used to generate hypotheses, understand design principles, create synthetic components, and produce effective therapeutic strategies. An important challenge is the inherent complexity of biological network models, which complicates the task of relating network structure to function and of understanding the conceptual design principles by which a given network operates.

Here we investigate one class of approaches for analyzing the relationship between network structure and functional behavior. We are concerned with dynamic properties, which may be particularly important for biological systems, although the same framework can address questions of steady state behavior, which are generally simpler.

The overall approach involves applying optimization techniques to identify the best combinations of model parameters to achieve canonical functional characteristics. In this manner, we study the relationship between model parameters (generally rate constants) and function. This complements, but differs dramatically from, approaches in which systematic variation is applied to the parameters and the resultant change to behavior is monitored [15, 107, 62, 34, 102]. Here, essentially by manipulat-

ing the desired functional behavior (generally properties of the network output) and monitoring the resultant required parameters — i.e., the inverse of more standard variational approaches — we can learn about the relationship between function and parameterization and may be in a position to discover new design principles.

The approach taken here utilizes a simulation model for the biological network, a search space, and one or more objective functions. Additionally, sets of constraints are employed in some applications. In the current work, we used mass-action kinetic models with state variables corresponding to biomolecule concentrations, parameters corresponding to kinetic rate constants, and initial conditions defining total biomolecule concentrations and their distribution at the start of the simulation. The models used simulate fundamental steps of signal transduction cascades in biological systems. Mass-action kinetic models have been popular and successful for simulating biological networks, although there are certain limitations and approximations inherent in their use, some of which can be avoided with other modeling methods [35, 1]. The search space corresponds to ranges of individual parameters that are considered during optimization. Here each rate constant was varied continuously from its canonical value through a range from three orders of magnitude smaller to three orders of magnitude larger ($k_i = \mu_i k_{i,0}$ for $\mu_i \in [10^{-3}, 10^3]$). We are also investigating variation of network topology in addition to parameterization, but the combined optimization of continuous and discrete variables is beyond the scope of the current report. The objective function describes the desired functional behavior of the network. Here we have focused on one-input–one-output networks and for illustrative purposes have examined the objectives of response-time and signal amplification. These functional characteristics correspond to measures of how quickly and accurately changes in output follow changes to the input signal.

To examine the effect of competing objectives, we implemented a technique of multiple-objective optimization, with one function as the objective function and an-

other enforced as a constraint. Enforcement as a constraint produces a rigorous method to map trade-off curves for multi-objective optimization, although other techniques are also available and may be more efficient [98]. In particular, by enforcing a constraint on signal amplification, we have clarified how different strategies for minimizing response-time are encoded in a single network topology. Interestingly, the strategy that is more optimal depends on the desired level of signal amplification. That is, the results suggest that a fixed topology network consisting of linked enzyme-catalyzed reactions can adopt multiple strategies through varying only rate parameters over a relatively modest range. The ability to adopt multiple strategies may lead to significantly improved performance under changing conditions on evolutionarily accessible time scales. Also, the use of chemical modification or binding modulators can customize the same generic network to operate with different strategies under different conditions.

2.3 Framework and Methods

Models. The fundamental network unit examined here consists of the enzyme catalyzed chemical modification of a protein molecule and the unmodification catalyzed by a different enzyme. This basic unit is found repeatedly in multiple configurations throughout a wide variety of biological pathways, including methylation/demethylation reactions integral to bacterial chemotaxis [92] and phosphorylation/dephosphorylation reactions of MAP kinase signaling cascades [6], as just two specific examples of a very general motif. Here the generic nature of network architecture constructed from this basic unit will be emphasized by the use of general nomenclature and symbols, but the examples are motivated by current models of MAP kinase signaling cascades; model details, including canonical parameters, were taken from that class of networks.

Figure 2-1 illustrates the network topologies examined here. The fundamental unit involves the modification of protein A by enzyme F (for forward modification) to give activated protein A' ; protein A' is also unmodified by enzyme R (reverse modification = unmodification) to give protein A (Figure 2-1a). For many applications it is useful to consider the activity of F as network input and the activity of A' as network output, although A and R activity are also important and will be treated as secondary inputs here. The enzymatic forward and reverse modification mechanism was treated as a single-substrate Michaelis–Menten mechanism (Figure 2-1b) with a coupled ODE model, as is typically done in many models of biological signal transduction; this is an approximation because it neglects the source and sink for the modification reactions (such as binding of ATP and release of ADP, as well as any intervening chemistry), but is thought to be reasonably accurate when concentrations of neglected species remain constant (Figure 2-1b, c, and d). This fundamental network unit can be coupled to copies of itself (or to other units) to create a variety of topologies such as the two-step modification typical of an individual layer of MAP kinase pathways (Figure 2-1e), in which the output of the first unit acts as a secondary input (substrate) to the next unit, which then produces the final output [28]. In this example, the same input, F , is applied in parallel to both units, as is R . Another topology studied here is the simplified two-layer network shown in Figure 2-1f, in which the output of the first unit serves as the input (forward modification enzyme) to the second, which is a form of series connection. Repeating the fundamental unit six times with appropriate interconnections reproduces common models of MAP kinase cascades that use no scaffolding interactions (Figure 2-1g).

Canonical parameter values were obtained from the work of Huang et al. [28] (Table A.1). Here wide ranges of rate constant parameter values were systematically examined spanning three orders of magnitude to either side of the canonical value ($k_i = \mu_i k_{0,i}$ for $\mu_i \in [10^{-3}, 10^3]$).

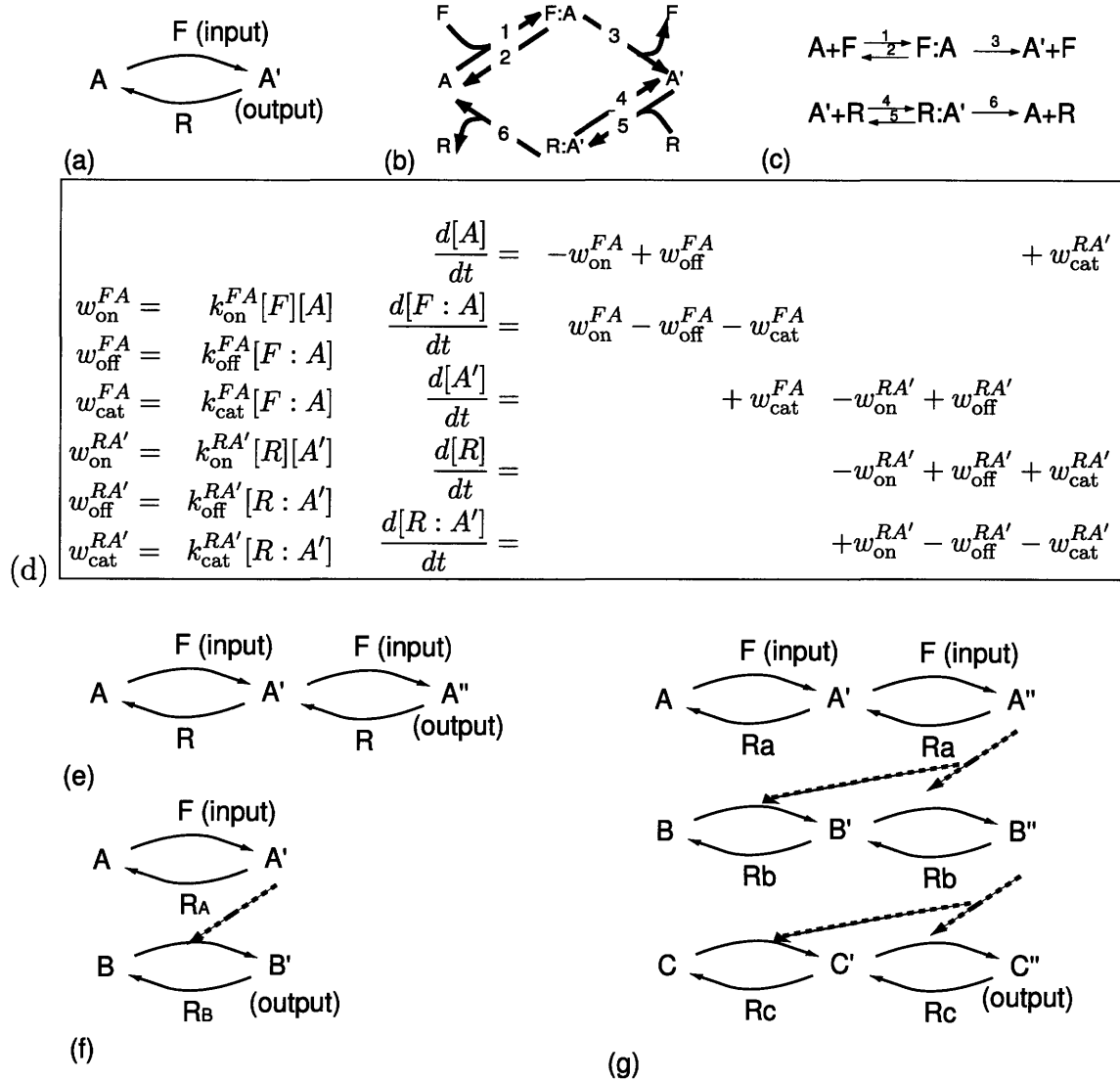


Figure 2-1: Alternative representations of the fundamental enzymatic modification and unmodification reactions. (a) Schematic of a one-step modification and unmodification reaction network; (b) alternative representation of the network given in (a), with emphasis on conservation of A; (c) chemical reaction network representation of the network given in (a); (d) mass-action dynamic equations of the network given in (a); (e) schematic of a two-step modification and unmodification network; (f) schematic of a two-layer cascade with one-step modification and unmodification in each layer; (g) schematic of a three-layer cascade network with two-step modification and unmodification in each layer.

Optimization functions. Parameter values were computed to optimize functional characteristics of the network topologies examined. Two target functions were used individually in different optimization calculations, which are defined by the following

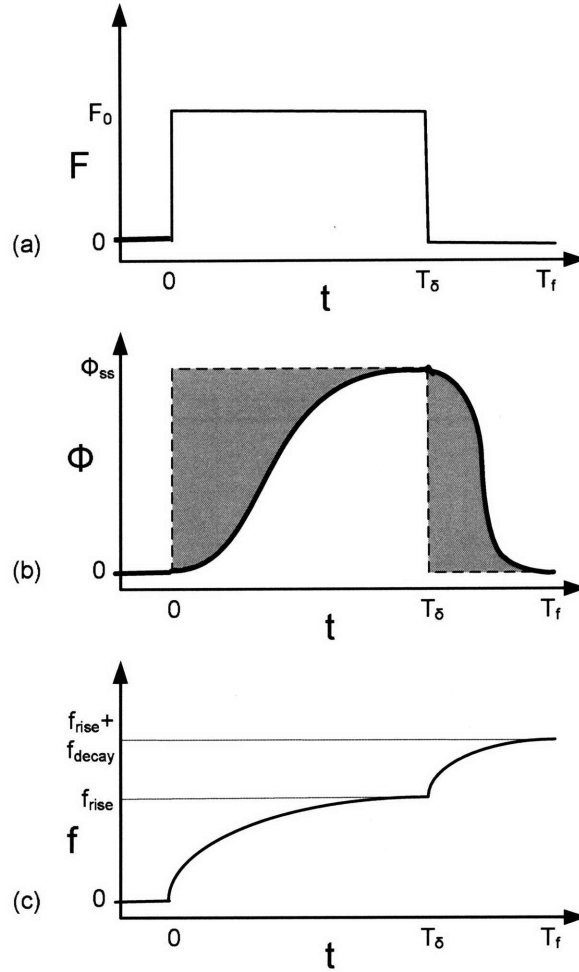
equations:

$$f_{\text{response}} = f_{\text{rise}} + f_{\text{decay}} = \int_0^{T_\delta} \left(\frac{[\mathcal{O}] - [\mathcal{O}]_{\text{ss}}}{[\mathcal{O}]_{\text{ss}}} \right)^2 dt + \int_{T_\delta}^{T_f} \left(\frac{[\mathcal{O}]}{[\mathcal{O}]_{\text{ss}}} \right)^2 dt \quad (2.1)$$

$$[\mathcal{O}]_{\text{ss}} = [\mathcal{O}](t)|_{t=T_\delta} \quad (2.2)$$

where $[\mathcal{O}](t)$ is the activity (equivalent to concentration here) of the network output (in Figure 2-1), which is $[A']$ for parts a–d, $[A'']$ for part e, $[B']$ for part f, and $[C'']$ for part g; $[\mathcal{O}]_{\text{ss}}$ is the steady state value of the output while a constant input is applied. Each objective function was studied in the context of the response to a square-wave input pulse of width T_δ and amplitude F_0 (see Figure 2-2). The response-time objective, f_{response} , measures how quickly and accurately step changes to network input are reflected in the output. T_f is a time sufficiently long that the network has returned to its unmodified state. The value of the objective evaluated for a given network with a given input pulse is zero if the output tracks the input perfectly (pure amplification) and accumulates an increasingly positive value for increasing discrepancies. Normalization by the steady state output achieved with constant input provides comparable objective for networks of varying attenuation. The response-time objective is the sum of two parts, f_{rise} that quantifies the fidelity of the network in replicating the *on* or *rise* portion, and f_{decay} that is defined similarly for the *off* or *decay* portion.

Simulation and optimization. Simulations were performed with the DSL48S program [131] within the ABACUSS package (version 2) [67]. The problem formulation for the work reported here involves differential algebraic equations (DAEs) that are both stiff and sparse. The differential equations are fundamental to the kinetic chemical model of the biochemistry; the algebraic equations enter through a flux model formulation, which is not necessary but is more efficient because it reduces the need



$$(d) \quad f_{\text{response}} = f_{\text{rise}} + f_{\text{decay}} = \int_0^{T_\delta} \left(\frac{[\mathcal{O}] - [\mathcal{O}]_{\text{ss}}}{[\mathcal{O}]_{\text{ss}}} \right)^2 dt + \int_{T_\delta}^{T_f} \left(\frac{[\mathcal{O}]}{[\mathcal{O}]_{\text{ss}}} \right)^2 dt$$

Figure 2-2: The definition of the response-time (f_{response}) and output-amplitude ($[\mathcal{O}]_{\text{ss}}$) objectives. (a) A square input trajectory was used to stimulate each network; (b) an output trajectory after stimulation; rise- and decay-time objectives are related to the early and later grey regions, respectively. (c) the rise, decay, and response-time objective trajectories after stimulation; (d) the mathematical definition of rise, decay, and response-time objectives.

for repeat computations and can produce a better sparsity pattern [33]. A stiff solver is necessary for efficient computation of at least some of simulations made because parameter ranges spanning six orders of magnitude were involved, creating the opportunity for vastly different time scales within one simulation. The system of equations is sparse because the time derivative of each activity or concentration (state variables)

depends on only a small number of the other state variables. The DSL48S program uses a *staggered corrector* method [131] to simultaneously solve differential algebraic equations, a large-scale sparse linear algebra package MA48 [58] to make efficient use of sparsity, and a Backward Difference Formula (BDF) [75] to deal with stiffness. The solution for the set of rate constant multipliers (μ_i variables) that minimize the objective function for a given differential equation model, set of initial conditions, and constraints are solved with the *sequential modular* algorithm [74], where the system dynamics (including objective function and constraints, as well as their derivatives with respect to the optimization variables) are solved for given sets of optimization variables, and this information is used iteratively to carry out nonlinear optimization in the space of the optimization variables to minimize the objective function subject to the constraints. The *sequential modular* algorithm may be more efficient for stiff system than other alternative methods [66]. Ideally the optimization problem with embedded dynamics could be solved with a deterministic and guaranteed procedure. We are working on methodology to achieve that goal [12]. For the current work, local optimization of variables was carried out from a large number of random starting seeds. The local optimization was carried out with the nonlinear programming (NLP) solver NPSOL, which implements a successive quadratic programming algorithm [99]. NPSOL terminates each local optimization when a Karush–Kuhn–Tucker (KKT) [57, 90] point is reached, to within a numerical tolerance of $10^{-6} \mu M$. It should be noted that in an infinite (unbounded) space of optimization variables and with no constraints, a local minimum is identified as a point with zero gradient. For the work reported here, optimization was carried out in a space of optimization variables consisting of the logarithms of the rate constant multipliers (bounded by -3 and $+3$). Values of $T_\delta = 10$ s and $T_f = 110$ s were used. Simulations were initiated from an equilibrated point in the absent of F (Table A.1) at time $t = 0$. Local optimization from a large number of random starting points was used; the best optimum was taken

as the global optimum. As a check on convergence, for the six-variable problem, when ten times as many runs were performed, no new lowest local optima were observed. Problems with six (twelve) parameter variables were repeated from 1,000 (20,000) sets of starting parameters using random initial seeds on the logarithmic scale, and the lowest local optima were observed 10 (3–5) times. Typically each optimization run in a multi-start set required 3 s (75 s) for a six-(twelve-) variable problem using a single 2.8 GHz Intel Pentium III Xeon processor.

2.4 Results and Discussion

The fundamental network unit consisting of a pair of Michaelis–Menten enzymatic forward and reverse modification reactions (Figure 2-1a-d) was studied using optimization methodology. The response-time objective function f_{response} was minimized while all six rate-constant multipliers were permitted to vary subject to a series of constraints applied to the steady state amplitude of the network output ($[\mathcal{O}]_{\text{ss}}$; $10^{-6} \leq [\mathcal{O}]_{\text{ss}} \leq 34$; Figure 2-2). Total concentrations of $[A]_0$ and $[R]_0$ were fixed (Table A.1) and the same input square pulse of $[F]$ (pulse height of $[F]_0 = 20 \mu\text{M}$; pulse width of $T_\delta = 10 \text{ s}$) was applied for each optimization. The results are shown in Figure 2-3. The characteristics of the trade-off between f_{response} and $[\mathcal{O}]_{\text{ss}}$, along with representative global optimal networks for each value of the $[\mathcal{O}]_{\text{ss}}$ constraint are illustrated schematically in the Figure. In each optimal network, a red (green) reaction arrow indicates that the corresponding rate constant was driven toward its minimum (maximum) value through optimization; a blue reaction arrow optimized to an intermediate value; a black dashed arrow indicates that the objective was not very sensitive to the corresponding rate constant.

Interestingly, the patterns of rate constants attained over the set clustered into three distinct groups that can be thought of as strategies for achieving the minimum

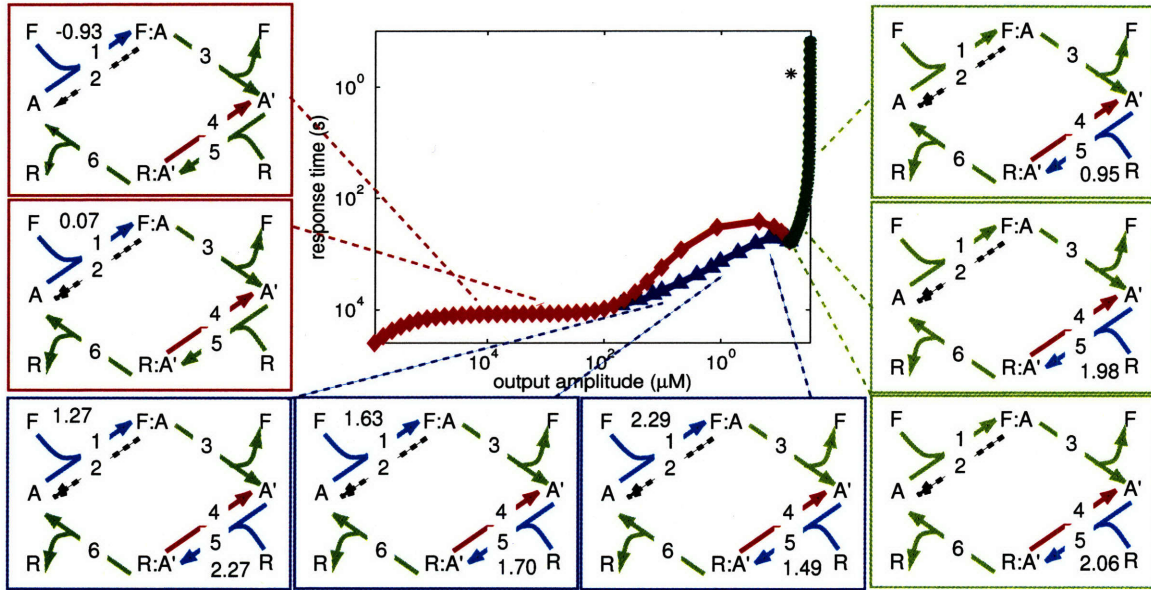


Figure 2-3: Trade-off curve of response-time and output-amplitude objectives, along with representative optimal networks. Legends on each optimal network: green, the rate is maximized to upper bound; red, the rate is minimized to lower bound; blue, the rate is optimized to a value between both bounds (the optimal value of $\log_{10}(\mu_i)$ is indicated); black-dotted, the rate is insensitive in the regions explored. Asterisk, when network is parameterized with the canonical MAP kinase values obtained from [28]. Strategy One network examples are enclosed in a red box, Strategy Two in a blue box, and Strategy Three in a green box.

response-time objective. Strategy One, involving an intermediate value of μ_1 and high value of μ_4 , was optimal at low values of the output-amplitude constraint; Strategy Three, involving a high value of μ_1 and an intermediate value of μ_4 , was optimal for high amplitude constraint values; and Strategy Two, combining aspects of the other two optima, was optimal for intermediate amplitude constraint values. Other characteristics were common to all three strategies, including small values of μ_5 , insensitivity to μ_2 , and maximal values of other rate constants.

To determine the advantages attained by a network that always optimizes to the one that is committed to a particular strategy, we repeated the optimization with the network locked into each of the three observed strategies. The resulting trade-off curves are also shown fully in Figure 2-3; they reveal that each strategy can only cover a limited range of amplitude constraints. Furthermore, in the region where Strategy

One and Two overlap, we observed that the network produced a worse response-time objective using Strategy One than when it was allowed to adopt the more optimal Strategy Two, suggesting a disadvantage for a network committed to a single strategy.

The shape of the trade-off curve between f_{response} and $[\mathcal{O}]_{\text{ss}}$ illustrates further characteristics of the the search space confronting the optimization, which may have commonalities with that confronting evolution. A non-linear optimization may be classified as convex (that a local optimum is globally optimal) and non-convex (that multiple local optima exist). Convex problems obey a continuity principle: a small change in the value of a constraint parameter produces a new optimum near the previous solution. Moreover, when sequential perturbations are applied, the resulting parametric optimal solutions do not change directions (they are monotonous). Using these criteria, we examined the trade-off curve in Figure 2-3 and found a non-monotonic behavior: that f_{response} increases with $[\mathcal{O}]_{\text{ss}}$ except at $6.0 \leq [\mathcal{O}]_{\text{ss}} \leq 15.6 \mu M$, indicative of a non-convex space.

To examine the search landscape encountered by the network, we systematically varied each of the six parameter variables, μ_1 – μ_6 , starting from an optimum corresponding to Strategy One, Two, or Three and examined the effect on the response-time and output-amplitude objectives. For completeness, we also examined a base strategy with the four “forward” reactions ($\mu_1, \mu_3, \mu_4, \mu_6$) being as fast as possible and the two “opposing” reactions (μ_2, μ_5) being as slow as possible (for which $f_{\text{response}} = 0.000577 \text{ s}$ and $[\mathcal{O}]_{\text{ss}} = 15.4 \mu M$). Although this one-dimensional parametric variation covers only a small portion of the search space, the results are informative to examine closely related strategies observed above. The results, shown in Figure 2-4, portray a sample of the search landscape that contains both non-convex and insensitive regions. For example, the effect of μ_1 on the response-time reveals two local optima (one at low and and one at high μ_1 values), and insensitive region at intermediate values, near the value attained in Strategy One. As a result of this non-

convexity, we observed multiple local optima both at low (e.g., Strategies One and Two) and high (e.g., Strategy Three and the base strategy) values of μ_1 . Moreover, the insensitivity suggests we should observe some local optima with different optimal values of an insensitive parameter. Indeed, we observed multiple local optima whose optimal parameter values varied in the neighborhood of insensitive regions (data not shown).

Interestingly, Figure 2-4 also illustrates a striking range of variability among the various rate constants. The rate constants k_2 and k_5 have very little effect on f_{response} and $[\mathcal{O}]_{\text{ss}}$ for all Strategies. At the other extreme, k_3, k_4 , and k_6 strongly affect f_{response} , with smaller rates generally leading to worse (slower) response-time objective. Likewise, k_1 and k_3 have strong effects on $[\mathcal{O}]_{\text{ss}}$ for all strategies, with slower rates producing smaller $[\mathcal{O}]_{\text{ss}}$, but k_4 and k_6 show strong effects on $[\mathcal{O}]_{\text{ss}}$ only for Strategies One and Two, with slower rates generally increasing $[\mathcal{O}]_{\text{ss}}$. This Figure illustrates some of the trade-offs inherent in moving between strategies. The Strategy Two f_{response} behavior for k_4 indicates the ability to reach a shorter response-time by increasing k_4 to its maximum value ($\mu_4 = 10^3$) and thus switching to Strategy One; however, the $[\mathcal{O}]_{\text{ss}}$ behavior indicates that this change also decreases $[\mathcal{O}]_{\text{ss}}$. This parallels the observations of Figure 2-3 showing that Strategy Two is optimal for intermediate amplitude and Strategy One at low amplitude, with a smaller value of the response-time objective.

A fundamental challenge of biological network modeling is to extract conceptual understanding of networks and the strategies that they adopt. Strategies are likely to be selected because of their desirable functional properties across a wide range of conditions and time scales (including evolutionary time scales), and it is unclear to what extent snapshots of biological networks are expected to behave optimally. Nevertheless, there is ample evidence to suggest that sufficient optimization has occurred that much can be learned by studying biological systems from this perspective. With

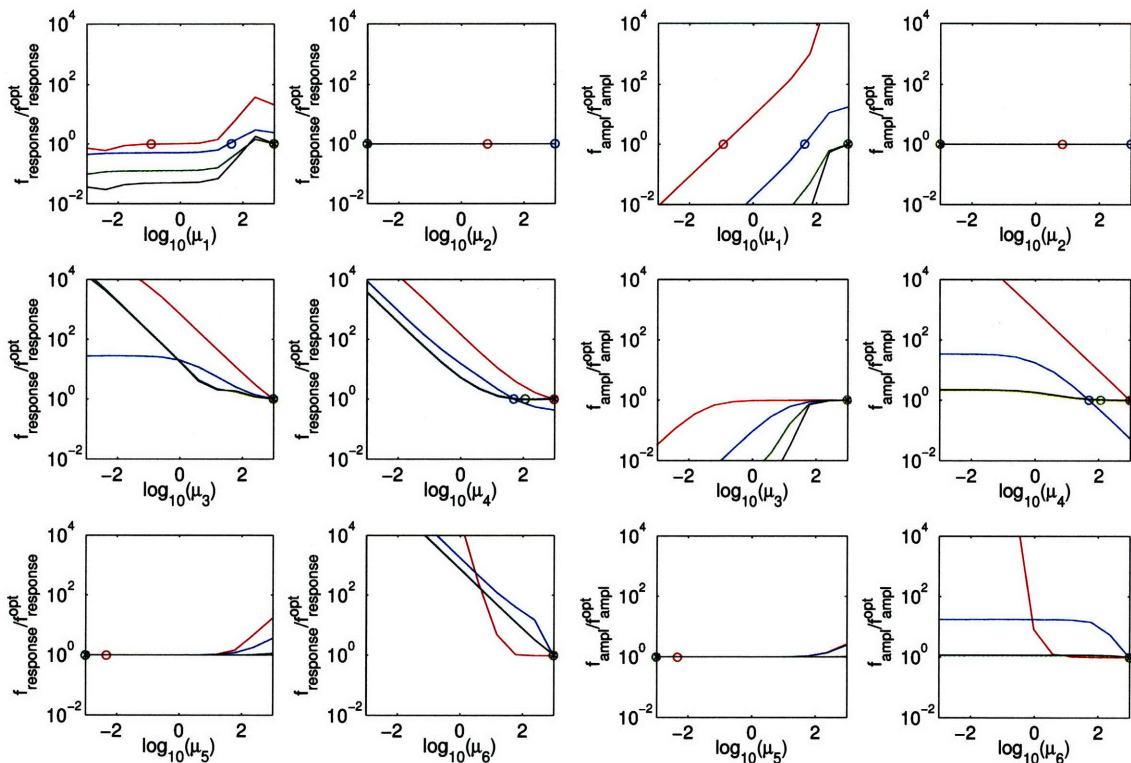


Figure 2-4: One-dimensional parametric perturbation of response-time and amplitude objectives for four strategies. Legends: red, Strategy One (with amplitude of $10^{-4} \mu M$); blue, Strategy Two (with amplitude of $10^0 \mu M$); green, Strategy Three (with amplitude of $20 \mu M$); and black, the base strategy, with all “forward” reactions being maximized and all “opposing” reactions being minimized. The objective values are normalized to the optimal values of each respective strategy. The symbols represent the unperturbed parameter values.

this goal, we have undertaken to understand in more detail the conceptual basis for the strategies adopted here.

Strategy One is particularly counterintuitive in that the first forward reaction, which is the first step in generating the output signal, is actually slowed to minimize response-time. An important research area involves dissecting network topology and parameter combinations to understand underlying operating principles. No general methodologies exist that are guaranteed to extract this type of understanding from biological network models, which are inherently non-linear, but a number of tools have been applied to the problem [27, 69, 126, 107, 62]. Here we have adopted a combination of perturbation analysis and objective decomposition to study the

mechanistic basis for Strategy One.

Starting from the Strategy One optimum with $[\mathcal{O}]_{ss} = 1.0 \times 10^{-4} \mu M$, the value of μ_1 was systematically varied across the range of $[10^{-3}, 10^3]$ with all other parameter variables fixed at their optimal values, as was done for the data in Figure 2-4. The value of the response-time objective was monitored, as were the separate rise-time and decay-time objectives that comprise the response-time. The results, shown in Figure 2-5d and plotted with a linear rather than logarithmic ordinate scale, demonstrate that the rise-time has two minima (one at low and the other at high μ_1 values), while the decay-time increases monotonically with increasing μ_1 ; the result is an overall response-time that also has two minima at low and high μ_1 , with the former being a global minimum. Thus, in the neighborhood of this Strategy One optimum, the observation is that choosing a smaller μ_1 value produces a better solution than maximizing μ_1 to its upper bound.

To understand the selection of a lower μ_1 value further, we compared the previous Strategy One to Strategy Three, which is the the best local optimum when μ_1 is constrained to its upper bound (Figure 2-5g). In this Strategy, μ_4 optimized to $10^{2.06}$, a value between its bounds. When we subjected Strategy Three to the same decomposition analysis, we discovered that rise, decay, and response-time objectives are qualitatively similar (but quantitatively of much smaller amplitude; Figure 2-5h) to Strategy One results: rise- and response-times have minima at both low and high μ_1 , while decay-time monotonically increases with increased μ_1 . For this strategy, choosing a smaller μ_1 value also produces a smaller response-time solution than maximizing μ_1 , but the difference is much smaller. Trajectories for many of the reaction species and the response-time objective, as well as their sensitivities to μ_1 variation are plotted in Figure 2-6. Each panel of the Figure is composed of two tiles, the first covering the initial 0.05 s after the input is switched on and the second cover the initial 0.05 s after 10-s input pulse is switched off. As these trajectories indicate, the

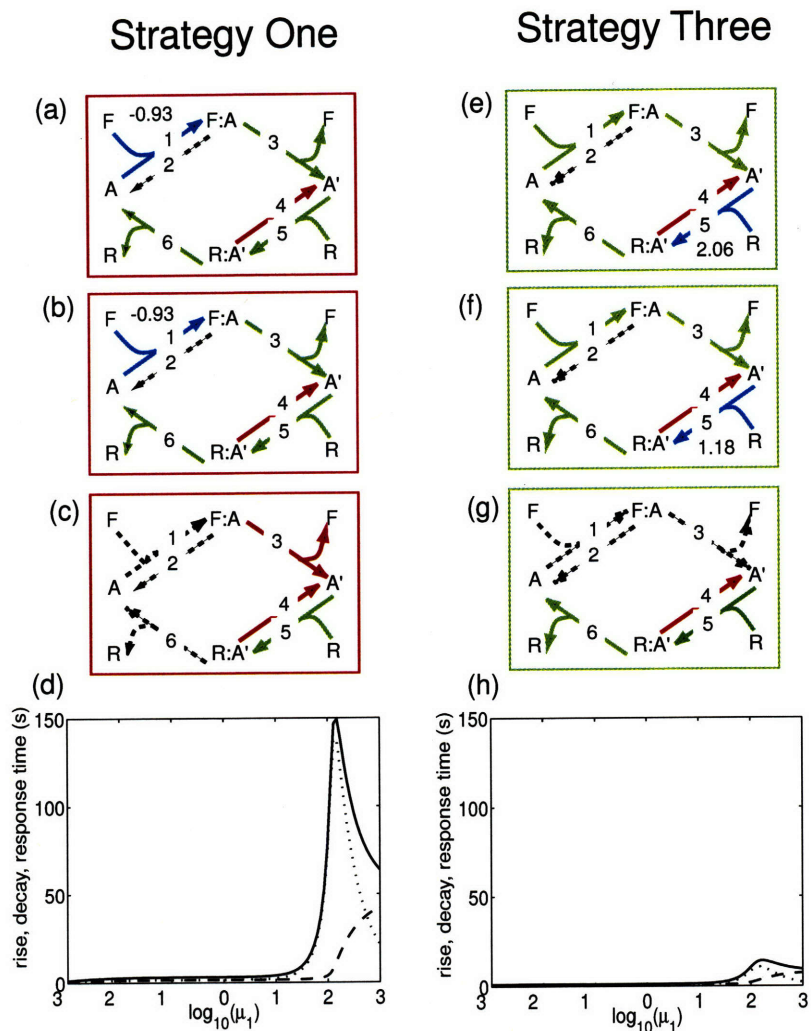


Figure 2-5: Decomposition and sensitivity analysis to uncover mechanisms of optimal strategies. (a),(e), response (rise+decay) time optimal networks; (b),(f), rise-time optimal networks; (c),(g), decay-time optimal networks; (d),(h), sensitivities of rise, decay, and response-time objectives to rate constant multiplier μ_1 . (a-d), applied to Strategy One; (e-h), applied to Strategy Three strategy; Dotted line, rise-time; dashed line, decay-time; solid line, response-time.

system dynamics is much faster than 0.05 s, and this network is capable of very fast responses. The traces for $[A']$ show an interesting difference between Strategy One and Strategy Three — namely, for Strategy One the concentration of A' rises sharply and levels off essentially instantly at the start of the initial 0.05-s window, but that for Strategy Three rises more slowly (to a higher value) and thus has a longer (worse) response-time objective. The adjacent panels in the Figure show trajectories for $\frac{\partial[A']}{\partial\mu_1}$,

which indicate the first-order changes to the $[A']$ trajectories for increasing k_1 . Both trajectories show an increase in the pulse-on steady state value of $[A']$, but the Strategy Three trajectory shows a brief, sharp overshoot of $\frac{\partial[A']}{\partial\mu_1}$. There is a trade-off inherent in increasing k_1 : on one hand the initial reaction velocity toward the pulse-on steady state is faster, which has the effect of shortening the response-time, but on the other hand pulse-on steady state is further away (occurs at higher $[A']$), which has the effect of lengthening response time. For Strategy One, this trade-off resolves itself in favor of small k_1 and low $[A']_{ss}$ because the effect of moving the intermediate steady state dominates the effect on reaction rate. For Strategy Three, the pattern of dominance switches and higher k_1 and $[A']_{ss}$ result.

Only one side of the same trade-off exists for the decay portion of the objective. Increasing k_1 increases the drop from the pulse-on $[A']_{ss}$ to the pulse-off steady state zero value of $[A']$, but k_1 does not directly affect the reaction rate for the disappearance of A' . Figure 2-5 and 2-6 confirm this behavior, in that the decay objective is worse for increasing k_1 in both Strategies One and Three with an increasing $[A']_{ss}$. Moreover, for Strategy One, at the time just before decay, a large pool of R ($[R] = 15.98 \mu M$) is available and the sensitivity trajectory of the unbound reverse enzyme is non-trivial ($\frac{\partial[R]}{\partial\mu_1} = -0.225 \mu M$), indicating that an increase in μ_1 decreases the pool of available R , and thus, increases the decay-time. In contrast, we found that Strategy Three, at the same time just before decay, both the pool of R ($[R] = 0.495 \mu M$) and its sensitivity trajectory ($\frac{\partial[R]}{\partial\mu_1} = 6.8 \times 10^{-5} \mu M$) are much smaller, indicating that a change in μ_1 barely affects the pool of R before decay. These sensitivities suggest that Strategy One selects lower μ_1 because of two reasons: first, because higher μ_1 produces a worse rise-time, mechanistically because the speed-up in kinetics of the output is exceeded by the increase in its steady state; and second, because higher μ_1 reduces the pool of available reverse enzyme that may speed up the decay-time.

To examine mechanisms employed by both strategies further, we decomposed

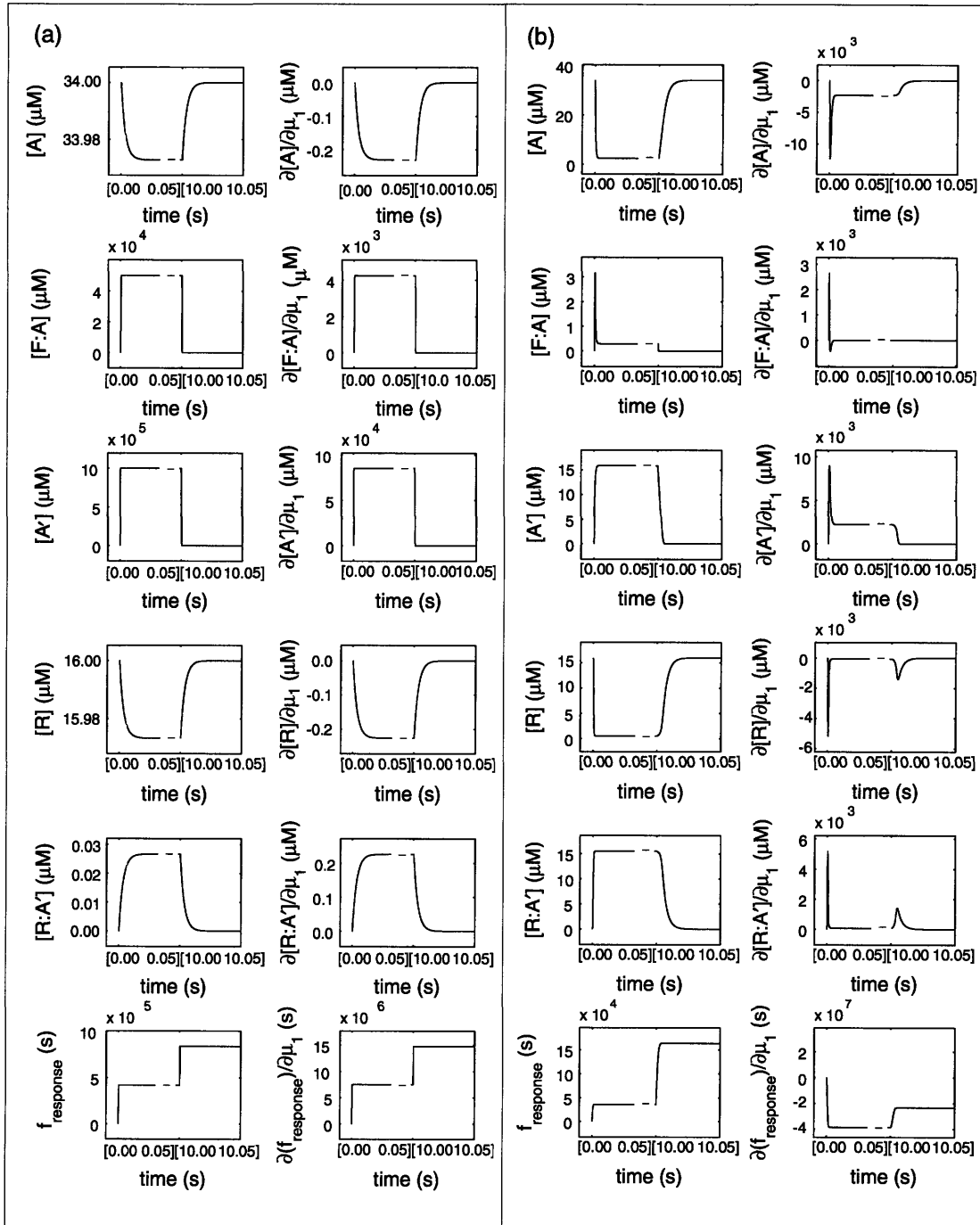


Figure 2-6: State and sensitivity trajectories of two alternative strategies of a one-step network. (a) Trajectories for Strategy One; (b) trajectories for Fast Start.

the problem by optimizing the network separately for rise- and decay-time. The resulting rise-time optimal networks, shown in Figure 2-5b and f, are very close to the two response-time optima networks in Figure 2-5a and e; The optimal decay-time networks, provided in Figure 2-5c and g, also classified to two optima: the Strategy One chooses maximal association and minimal dissociation rates of R and A' , minimal modification rate of $F : A$, and insensitive other rates; the Strategy Three chooses maximal association and minimal dissociation rates of R and A' , maximal modification rate of $R : A'$, and insensitive other rates. The existence of the two rise-time local optima agree with the two potential local optima observed in the sensitivity analysis described above. To determine the contribution of both rise- and decay-times, we merged the two rise- and decay-time optimal networks by selecting rate constants whose qualitative optimal values are each obtained by adding the objective sensitivities of each parameter in both networks. The resulting merged networks have qualitatively similar behavior as the observed response-time optimal networks. These results further support the two reasons behind the selection of lower μ_1 in the Strategy One mentioned in previous paragraph.

We extended the optimization and analysis to a two-step and a two-layer network, whose reaction schemes are given in Figure 2-1e and f. The multi-objective optimization results, provided in Figure 2-7, demonstrate that optimal strategies are clustered as well, this time with a greater number of observed strategies than the one-step network. Examination of the optimal networks showed that both of the two-step and two-layer networks also minimize an association rate to allow a sufficient pool of R to be available for decay; in these networks, two choices of association rates are possible and are selected across parameter variations (data not shown).

Utilizing the trade-off behavior as a framework, we compared the two-step and two-layer networks to the one-step network. We found that, despite having more degrees of freedom to optimize, both networks produced higher response-time objective

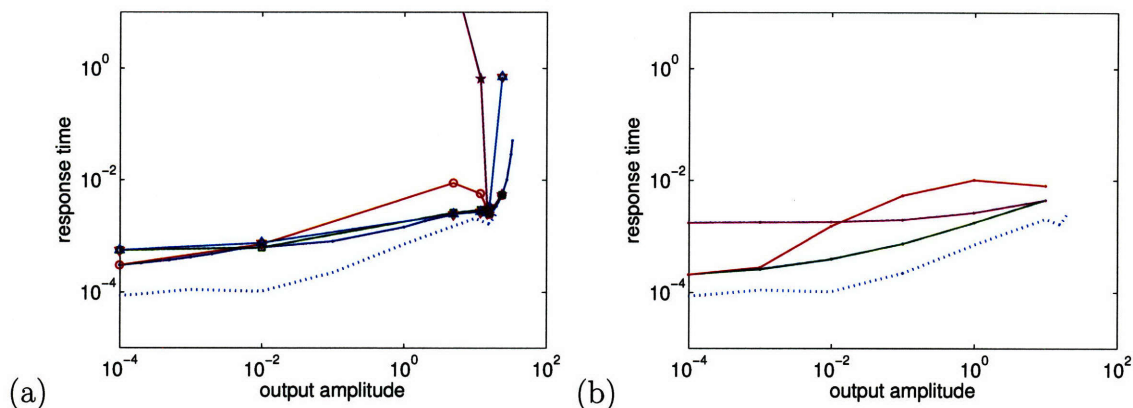


Figure 2-7: Trade-off curve for response-time and output-amplitude objectives for variant networks. (a) the two-step network given in Figure 2-1e; (b) the two-layer network given in Figure 2-1f. Legends: solid lines, multiple strategies in each network; dotted line, trade-off curve for one-step network.

values than the one-step network. This slower response-time can be rationalized because both two-step and two-layer networks contain extra intermediate entities that increase the latency of the network.

We extended the optimization framework to analyze a network consisting of two modification steps and a three-layer kinase cascade, whose scheme is shown in Figure 2-1g. We found that the trade-off has increased the latency than these networks (Figure 2-8), and that the optimal networks utilized a similar mechanism of maintaining a sufficient pool of reverse enzyme, although this time with many more alternative implementation strategies that minimize the consumption of reverse enzyme during the rise period.

We have studied optimization as a framework to discover design principles in biological networks. The characteristics of an optimization landscape and resulting optimal solutions were examined for an enzymatic network commonly found in signal transduction systems. Furthermore, we developed methods to extract knowledge out of the observed optimal solutions based on sensitivity and objective decomposition. We hope to extend the optimization framework to analyze and uncover design principles in larger-sized networks.

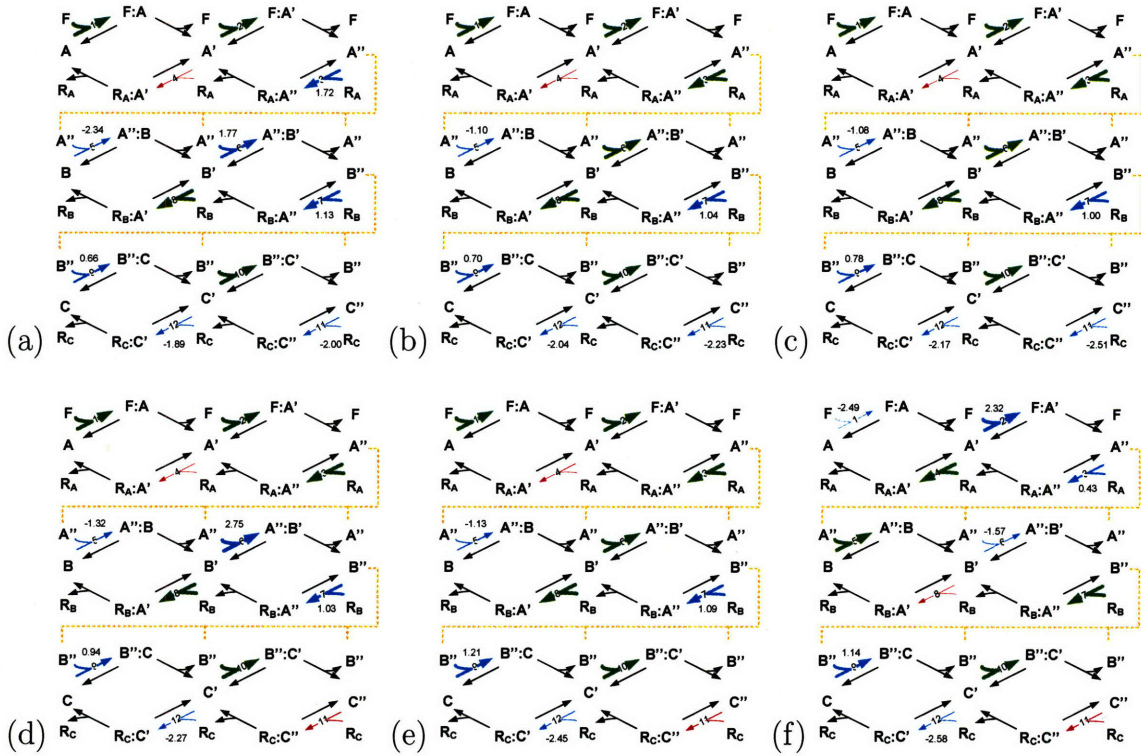


Figure 2-8: Response-time optimal kinase cascade networks at various amplitude constraints. (a) the optimal network for $[O]_{ss} = 10^{-4}$; (b) the optimal network for $[O]_{ss} = 10^{-3}$; (c) the optimal network for $[O]_{ss} = 10^{-2}$; (d) the optimal network for $[O]_{ss} = 10^{-1}$; (e) the optimal network for $[O]_{ss} = 10^0$; (f) the optimal network for $[O]_{ss} = 10^1$.

Chapter 3

Differentiating signal from transient noise in non-linear signal transduction networks

3.1 Abstract

Recent studies have developed preliminary wiring diagrams for a number of important and paradigmatic biological networks, including examples involved with key decision and control pathways in living systems. However, the design principles governing the construction and operation of biochemical networks remain mostly unknown. To address this research question, we investigate the application of optimization techniques as a set of tools for the discovery of design principles in biological networks. Optimization allows exploration of alternative network topology, parameterization, or both, and evaluation of the relative fitness of their operational strategies. One clear use of optimization is model specification and parameter estimation. Additionally, optimization may be a useful tool for understanding relationships between network structure and function and for exploring trade-offs among multiple goals. Here, we studied the ability of an enzymatic cycle to distinguish signal from noise as measured by its ability to produce a rapid but delayed response (RDR). We discovered that a two-step non-linear enzymatic cycle network can produce a response the sharpness of whose transition can only be matched by a cascade of about 100 linear reactions in series. Moreover, our studies demonstrated that this RDR allows the enzymatic network to filter transient noise effectively. Essentially, the novel mechanism employed by the optimal network hinges on the reverse enzyme acting as a reservoir that serves to delay output response; when the reverse enzyme is fully exhausted, rapid response

ensues. The operation of this mechanism in natural biological networks has some support from published literature both in bacterial chemotaxis and in cascades involving mitogen-activated protein kinase (MAPK). Finally, we examined the trade-off between the ability to produce a sharp RDR and output amplification in a family of related network structures and found that networks with multiple activation steps have advantages for producing a sharper RDR across wider ranges of output amplitude. Together these results show how complex functional behaviors of biochemical networks can be revealed through dynamic optimization.

Keywords: design principle, noise filtering, bacterial chemotaxis, mitogen-activated protein kinase, biological networks, systems biology, network motif, network optimization

3.2 Introduction

Living cells carry out the intricate job of making proper decisions in controlling simple and complex signal transduction networks that operate successfully in spite of stochastic inputs and network components. For example, patterning of the *Drosophila* embryo is controlled precisely despite stochastic fluctuations in physico-chemical systems [64, 20]. However, stochastic variations has been observed within a cell and among cells, in the form of different conditions of molecular machineries in a network [5] and stochasticity in the networks' reactive systems due to small numbers [5, 1]. Despite these variations, cell can make decisions and control their behaviors so accurately. The design principles governing the construction and operation of these well-controlled signal transduction networks are not fully understood. Knowledge of the design principles may allow us to understand the specific biological phenomena in context better, as well as the evolutionary pressures acting on the systems. Moreover, an understanding of biological design strategy may also guide the development of diagnostic and therapeutics, and could also contribute our knowledge of systems control theory. Recent studies have mapped the preliminary wiring diagrams for important transcriptional and regulatory networks and have developed mathematical models to

estimate the networks' observed dynamic behaviors in effort to better understand how molecular networks are controlled in living cells [94, 30, 1, 41, 22]. Here, we examine a substructure common to many of these network models, that of forward and reverse enzyme catalyzed modifications, to delineate optimal operational strategies and to understand relationships between network architecture and function.

To discover underlying design principles of signaling networks, we apply optimization techniques to probe mathematical models of these networks. Optimization techniques allow both exploration and evaluation of operational strategies inherent to any given network topology or across a set of related topologies. Using optimization techniques to study evolving systems could be particularly relevant in living organisms, as optimization, whether local or global, is considered to be one of the major driving forces in evolution [36, 112]. This approach may help relate network structure to function by understanding the trade-offs among multiple objectives.

We formulated the optimization using fitness functions as objectives, dynamic parameters (rate constants and initial concentrations) as optimization variables, and dynamic models with fixed topologies as constraints. In principle, both topology and parameters may be optimized simultaneously, however, in practice current numerical techniques (e.g. [18, 12]) are not efficient enough to guarantee global optimality. Thus, we have chosen to vary the topology manually and optimize for the dynamic parameters. The decomposition of the optimization into topology and dynamic parameters also is motivated by the large disjunction in the state of knowledge between topology and parameters; in most systems, dynamic parameters lack sufficient data to be regressed from while topology has been better studied. Various techniques have been used to develop network circuitry and models from experimental data [41, 22]. Here, we use our recently developed optimization techniques [24] to probe existing models and to uncover design principles.

The network studied here, called the enzymatic cycle [83, 32], comprises paired en-

zymatic activation and deactivation reactions (e.g. phosphorylation/dephosphorylation, or methylation/demethylation), and is common to many signal transduction networks such as bacterial chemotaxis [27, 65, 68] and MAP kinase cascades [47]. Enzymatic cycle networks are observed in MAP kinase cascades in multiple topologies, including multiple phosphorylation steps, layers of kinases and phosphatases [47]. We limit our scope to this enzymatic cycle and a variant topology with multiple activation steps, shown in Figure 3-1c, where a forward enzyme F catalyzes the activation of a substrate A to A' (and A'' , sequentially), and a reverse enzyme R deactivates the substrates. We wish to investigate the networks' ability to produce a rapid but delayed response and to distinguish signal from noise.

Previous studies on this network substructure demonstrated that the property of ultra-sensitivity; the steady state output exhibits a switch-like response to input amplitude— off for small inputs and on for larger inputs beyond some narrow transitional region [6]. Varying topology through additional activation/deactivation steps has been shown initially to produce a more sigmoidal response [28], and later, to produce a wider threshold region but poorer switch-like behavior [60]. These implicate that the ultra-sensitivity property serves to filter low-amplitude noise from high-amplitude signal. Here we asked whether these enzymatic cycle networks could differentiate signal from noise through their *transient* (as opposed to steady-state) behaviors by subjecting the networks to a step change in input and optimizing dynamic parameters to produce a rapid but delayed response (RDR). An RDR is the capability of a network to produce first a delay period with essentially no output followed by a rapid, switch-like response to its steady state output amplitude upon stimulation with a step increase from zero input (Figure 3-1b). While previous work focused on the steady state amplitude of a network output as a function of input concentration, here the focus is on the transient output response as a function of time. The inherent ability to generate an RDR in appropriately configured enzymatic cycle

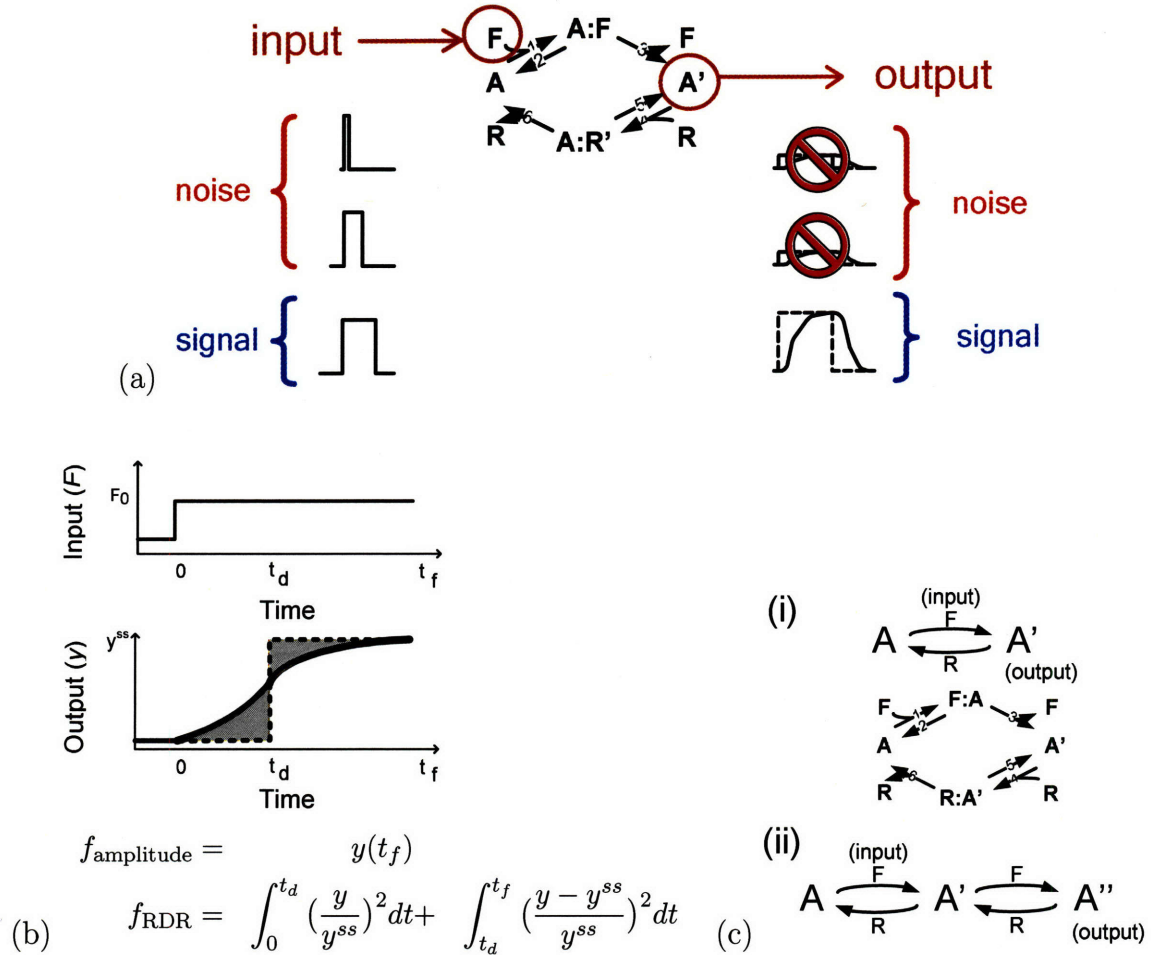


Figure 3-1: Transient noise filtering problem formulation. (a) Illustration of the filtering function. (b) Objectives, which are amplitude, and Rapid but Delayed Response (RDR). (c) The schematic of networks being studied, which are: (i) a one-step enzymatic activation and deactivation reactions, and (ii) a two-step enzymatic activation-deactivation with the output of the first step acting as a substrate for the second step enzymatic reactions. All the topologies assume elementary (uni- and bi-molecular) reactions.

topologies could endow networks with the ability to filter noises of short duration while responding to signals of longer duration.

We report here that nonlinear enzymatic cycle networks can produce an RDR using many fewer components than linear networks. Furthermore, we observed good correlation between the quality of a network's RDR and its ability to filter transient noise. Further analysis revealed that optimal enzymatic cycle networks utilize a novel strategy where the deactivation enzyme in the cycle acts as a reservoir of limited

capacity that serves to produce the delay and filter short-lived noise. Finally, we examined the trade-off between the RDR and output amplification and discovered that added modification sites, cascade steps, and number of reverse enzymes provide more handles to implement the optimal strategy to exhaust the deactivation enzyme, which allows the generation of a sharp RDR across a wide range of output amplitudes.

3.3 Methods

Biochemical network structures.

We examined two reaction schemes depicted in Fig. 3-1c: (i) a one-step enzymatic activation-deactivation, and (ii) a two-step enzymatic activation-deactivation with the output of the first step acting as an input for the second step. The enzymatic activation and deactivation reactions were modeled as elementary uni- and bi-molecular reactions with the Michaelis–Menten enzymatic reaction scheme that consists of two reversible binding fluxes and one irreversible modification flux. The two-step models assumed that the activation is sequential, modeled after the phosphorylation of Tyr185 and Thr183 in ERK1 [125]. In the perturbation study (Table 3.1), the enzymatic reaction steps (consisting of binding and modification reactions) were substituted by a linear (non-enzymatic) reaction.

Dynamic simulation setup.

The input to the biochemical network was the activation enzyme (F), whose temporal trajectory was specified externally from the dynamic model and took the form of a step change or a square pulse with varying width. The output was defined to be the final activated states of the substrate, which were the A' for the one-step and A'' for the two-step.

The dynamic simulation was formulated as deterministic differential equations

representing the mass-action kinetics that consist of uni- and bi-molecular reaction rate fluxes and state conservation equations. For example, the equation for $\frac{dA}{dt}$ in the one-step network (Fig. 3-1c(i)) is represented as:

$$\begin{aligned}
 w_1 &= \mu_1 k_{\text{on}}^{[F:A]} [F][A] \\
 w_2 &= \mu_2 k_{\text{off}}^{[F:A]} [F : A] \\
 w_6 &= \mu_6 k_{\text{cat}}^{[R:A']} [R : A'] \\
 \frac{dA}{dt} &= -w_1 + w_2 + w_6
 \end{aligned}$$

where $[A]$, $[F]$, $[F : A]$, $[R : A']$ are the state variables, w_i are the reaction fluxes, k_i are the rate constants, and μ_i are unit-less multipliers of the rate constants. The complete differential equation model is provided in Table A.2 and A.3. The canonical values of the rate constants are obtained from a combination of ERK1 and *Xenopus* oocyte MAP kinase [28, 51, 22] and are provided in Supplementary Table A.1. For the current work, the initial conditions were taken as all substrates (A,B) and enzymes (F,R) unbound and inactive, unless otherwise noted. This condition corresponds to the steady state in the absence of input. The initial condition values were obtained from the literature [22] and provided in Supplementary Table A.1.

Dynamic optimization setup.

We examined two objective functions: the output amplitude ($f_{\text{amplitude}}$), which measures the output at the final time, and the RDR objective (f_{RDR}), which measures the sharpness of the delayed response to steady state when a network is stimulated

by a step change input.

$$f_{\text{amplitude}} = y(t_f) \quad (3.1)$$

$$f_{\text{RDR}} = \int_0^{t_d} \left(\frac{y}{y^{ss}}\right)^2 dt + \int_{t_d}^{t_f} \left(\frac{y - y^{ss}}{y^{ss}}\right)^2 dt \quad (3.2)$$

Mathematically, the RDR objective measures the normalized square difference between the output and a target trajectory, which takes the form of a step change shifted by t_d . This squared formulation is conceptually similar to minimizing the red area in Fig. 3-1b; the height of the target trajectory was dynamically adjusted to the final steady state value corresponding to the parameter selected during each optimization run. We used delay times (t_d) of 1, 0.1, and 0.01 s, with the majority of results using 0.1 s. The final time (t_f) was selected to allow the signal to reach steady state, which is numerically defined here as the condition with the dynamic residual value at most 10^{-9} ; a final time of 100 s was sufficient for the work reported here.

The optimization variables were the rate constant multipliers (μ_i) of the canonical values, and in some cases, the initial concentration of the reverse enzyme (R_0). The multipliers were searched logarithmically in a space bounded by $[10^{-3}, 10^{+3}]$, and the initial concentration was searched in the space bounded by $[0.1, 100] \mu M$. For the current work we elected to search this relatively large range of parameter values to examine the quality of our numerical procedures; in other applications, limiting $\mu_i k_{\text{on}}$ to diffusion limited rates, and other physical limits, could be useful.

Numerical tools.

Simulations were performed with the DSL48S program [131] within the ABACUSS package version 2 [67]. The problem formulation for the work reported here involves differential algebraic equations (DAEs) that are both stiff and sparse. The differential equations are fundamental to the kinetic chemical model of the biochemistry; the

algebraic equations enter through a flux model formulation, which is not necessary but is more efficient because it reduces the need for repeated computations and can produce better sparsity patterns [33]. A stiff solver is necessary for efficient computation of at least some of simulations made because parameter ranges spanning 6 orders of magnitude were involved, creating the opportunity for vastly different timescales within one simulation. The system of equations is sparse because the time derivative of each activity or concentration (state variables) depends on only a small number of other state variables. The DSL48S program uses a *staggered corrector* method[131] to simultaneously solve differential algebraic equations and their sensitivities, a large-scale sparse linear algebra package MA48 [58] to make efficient use of sparsity, and a Backward Difference Formula (BDF)[75] to deal with stiffness.

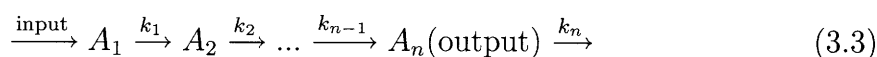
The solution for the set of rate constant multipliers (μ_i variables) that minimize the objective function for a given differential equation model, set of initial conditions, and constraints was solved with the *control parameterization* algorithm [74], where the system dynamics (including objective function and constraints as well as their derivatives with respect to the optimization variables) were solved for given sets of optimization variables, and this information was used iteratively to carry out nonlinear optimization in the space of the optimization variables to minimize the objective function subject to the constraints. The *control parameterization* algorithm may be more efficient for a stiff system than other alternative methods [66]. Ideally the optimization problem with embedded dynamics should be solved by a deterministic global optimization. Currently we are working on methodology to achieve that goal [12]. For the current work, local optimization was carried out from a large number of random starting points. The local optimization was carried out with the nonlinear programming (NLP) solver NPSOL, which implements the Successive Quadratic Programming (SQP) algorithm [99]. NPSOL terminates each local optimization when a Karush–Kuhn–Tucker (KKT)[57, 90] point is reached, to within a convergence tolerance—the

tolerance to reach zero with limited machine precision— of 10^{-6} (The complete definition of the convergence tolerance is given in reference [99]). The best optimum from the set of multi-start local optima was taken as the global optimum. To check the convergence to the global optimum for the one-step network, ten times as many runs were performed and no new lowest local optima were observed. Problems with six (twelve) parameter variables were repeated from 1,000 (20,000) sets of starting points using random initial points, and the lowest local optima were observed ≈ 10 (3–5) times. Typically each optimization run in a multi-start set required 3 (75) s for a six- (twelve-) variable problem using a single 2.8-GHz Intel Pentium III Xeon processor.

Because of the presence of insensitive regions near the local optima, we categorized the optima by weighting to their sensitivities. Here, we computed one-dimensional sensitivity on each optimization variable in some lowest local optima and categorized the variable based on its location on the sensitivity curves. We are working on more rigorous analysis based on multi-dimensional sensitivities. With these sensitivities, we categorized optima with slightly different parameter values to the same group if the different values correspond to the flatness of the search space.

Comparison to linear networks.

Rapid but delayed response (RDR) has been observed in a cascade of Continuously Stirred Tanks (CST) in chemical engineering, or equivalently, in a cascade of linear (non-enzymatic) reactions in series. The cascade of linear reactions followed this scheme:



$$\begin{aligned} \frac{d[A_1]}{dt} &= \text{input} - k_1[A_1] & A_j(t_0) &= 0 \quad \forall j \in \{1 \leq j \leq n\} \\ \frac{d[A_j]}{dt} &= k_{j-1}[A_{j-1}] - k_j[A_j] & & \forall j \in \{2 \leq j \leq n\} \end{aligned} \quad (3.4)$$

where k_i are uni-molecular reaction rate constants. To examine rapid delayed response as a function of cascade length, we ran two studies:

1. We optimized for the RDR objective (f_{RDR}) for a cascade with fixed length n by varying the rate constants k_i , and we repeated for different length n .
2. We simulated the dynamics of these linear networks by varying the cascade length n , constraining the rate constants to be the same across all steps for a given cascade length n , adjusting the rate constant values using the equation $k_n = \frac{n}{t_d}$ and measured the values of the RDR objective (f_{RDR}) for varying n .

The simulation study corresponds to the cascade of CSTs problem in Chemical Engineering, and we used the study to verify the capability of the RDR objective formulation to measure sharpness of the ensuing RDR. Ideally, we would like to optimize for the RDR objective for any cascade length, however, the optimization expense grows combinatorially with cascade length; fortunately our results for short cascades show that the simulation heuristic above was essentially equivalent to the optimization. The correctness of the heuristic in the simulation was verified by comparing the RDR objective values and rate constants to the corresponding values obtained from optimization for cascade length up to ten; we found that for a cascade of the same length, simulation and optimization produced objective values within 0.24% and rate constants within 1% of each other. Furthermore, with increasing cascade length and fixed t_d , both simulation and optimization produced monotonically decreasing values of the RDR objective, suggesting that the RDR objective may be used to measure the sharpness of RDR for a fixed t_d . To generalize for other t_d values, we optimized with a different t_d value ($t_d = 0.1$, previously $t_d = 1$) for cascade lengths up to ten, and ob-

tained a t_d -normalized RDR objective ($\frac{1}{t_d}f_{\text{RDR}}$) within the optimization convergence tolerance and t_d -normalized rate constants ($t_d k_i$) within 0.01%. This is consistent with the fact that normalization converts the RDR objective into a unit-less measure, and led us to use the normalized objective to compare results at different t_d values.

Relating the rapid but delayed response and transient noise filtering.

To establish relationships between the rapid but delayed response (RDR) and transient noise filtering, we subjected a network optimized for the RDR objective to a series of pulse inputs with varying width. Furthermore, we constructed the power spectrum of the two-step enzymatic cycle network and the 100-step linear networks by subjecting the network to pulses of inputs with width varying logarithmically in $[10^{-4}, 10^1]$ s range, and measured the maximum output amplitude for time less than 100 s. A more rigorous method to construct power spectrum is to take the Discrete Fourier Transform (DFT) of the output response when the network is subjected to a random input covering the whole spectrum, however, we did not use this method because the DFT requires even time-spacing of the output spectrum, which is computationally expensive for the five orders of magnitude range of frequency. The two-step enzymatic cycle network is parameterized for three cases: RDR objective optimized ($\log_{10}(\mu_i) = [2.46, 0.68, 1.04, 2.54, 0.80, 2.60, 3, -2.26, 2.97, 2.35, 3, 3]$), response-time optimized [24] ($\log_{10}(\mu_i) = [3, -1.29, 3, 3, -2.22, 3, 1.31, -2.57, 3, -3, -1.52, 3]$), and *Xenopus* oocyte MAPK parameterized with the same activation and deactivation rates ($\log_{10}(\mu_i) = [1.70, 1.18, -0.73, 0, 0, 0]$).

3.4 Results

Nonlinear enzymatic cycle networks can produce a rapid but delayed response (RDR) more effectively than linear networks.

We first optimized linear and non-linear enzymatic cycle networks of varying length for the rapid but delayed response (RDR)(Figure 3-1a,b).The linear networks are shown in Equation 3.3 (Methods) and the non-linear enzymatic cycle networks in Figure 3-1c. In the linear network, the product of each reaction was the reactant for the next, and all reactions were first order and irreversible. The non-linear enzymatic cycle networks also had the product of one reaction being the reactant of the next, but each reaction was non-linear in that it was catalyzed in the forward direction by one enzyme and in the reverse by another. The rate constants for each network were optimized to an RDR objective, a measure of delayed response sharpness (see Methods), with time delay of 0.1 s.

The results, shown in Figure 3-2, demonstrate that linear networks of increasing length produce sharper output responses. Output trajectories for linear networks of length 4, 10, and 100 are shown with increased RDR sharpness. Non-linear enzymatic cycle networks of one and two steps are also shown, with the longer network also demonstrating a steeper response. Interestingly, the relatively short non-linear enzymatic cycle networks were more capable of producing steeper RDR responses than much longer linear networks. For example, the one-step enzymatic network was as effective as a 10-step linear network; likewise, the two-step non-linear matched the performance of a 100-step linear network. The non-linear enzymatic cycle network contains far fewer components (six or nine), than the equivalently performing linear networks (10 or 100). We next tried to establish a relationship between sharp RDRs and transient noise filtering, the ultimate property of our interest.

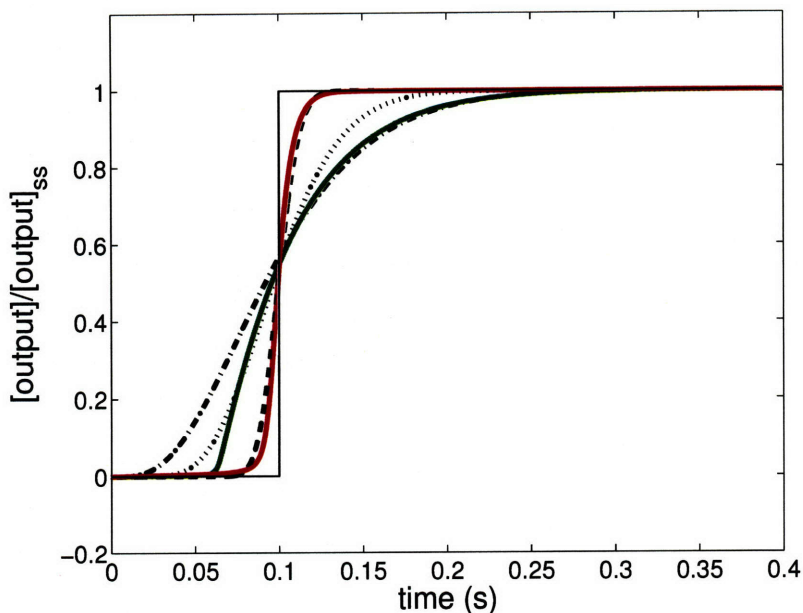


Figure 3-2: Response trajectories of RDR optimal linear and non-linear enzymatic cycle networks with various lengths. Black, the output trajectory of linear networks with length of 4 (dot-dashed), 10 (short-dotted), and 100 (long-dotted) steps; green, the one of a one-step enzymatic network; red, the one of a two-step enzymatic network. The optimal one-step and two-step enzymatic cycle networks produced a rapid but delayed response (RDR) the sharpness of whose transition can only be matched by a cascade of 10 and 100-step linear reactions in series.

Optimal RDR enzymatic cycle networks filter out transient noise effectively.

To analyze the relationship between RDR and the ability to filter transient noise, we stimulated the RDR-optimal two-step enzymatic network and the 100-step linear network with a series of square-wave pulse inputs of varying duration and examined the corresponding responses. Our expectation was that input pulses with duration shorter than the delay time t_d of the RDR would be filtered out as noise but that longer pulses would be delayed but transmitted intact as signal. The data are shown in Figure 3-3a. The top panel gives the train of input pulses used in this study, and the next two panels give the resulting trajectories for the two-step enzymatic and the 100-step linear network, respectively. The results show filtering of short inputs

and transmission of long inputs relative to the delay time, but the two-step enzymatic network acted as a more discriminating filter. For example, the longest duration noise pulse of 0.05 s was filtered out almost completely (to 0.5% of maximal response) in the two-step enzymatic network, but was barely filtered (to 99.99% of maximal response) in the 100-step linear network.

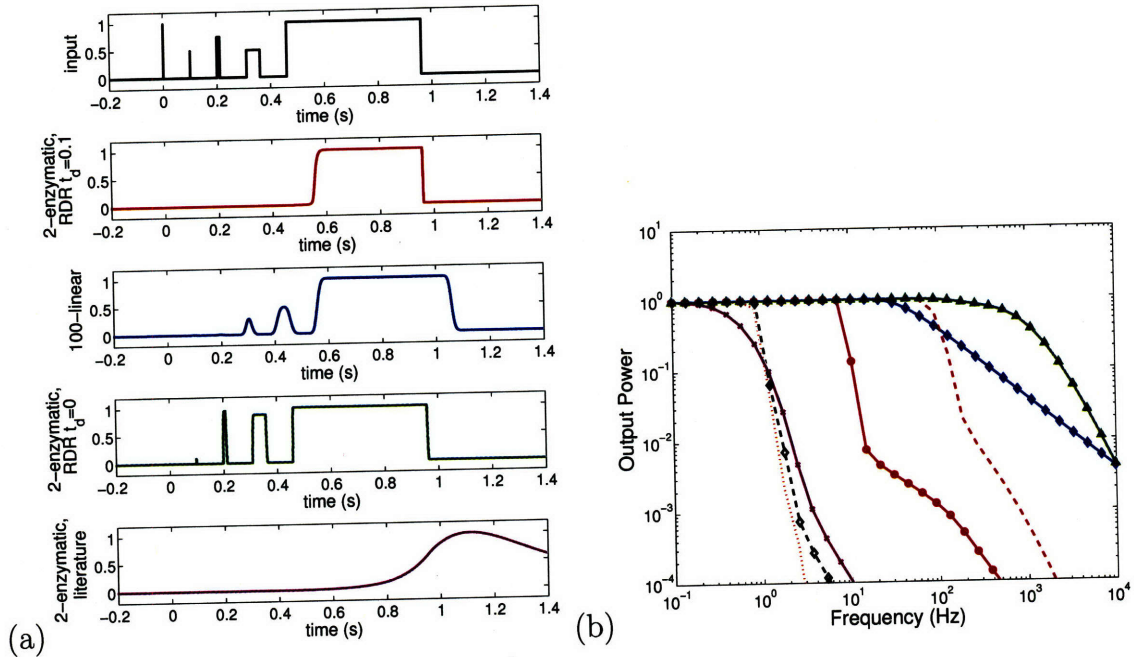


Figure 3-3: The relationship between RDR-optimality and transient noise filtering in a two-step enzymatic cycle and a 100-step linear network. (a) The output trajectories when the networks are subjected to inputs of pulse trains with varying duration. Top panel, the input trajectory of pulse trains with varying duration; lower panels, the output response for different network topology and parameterization. Black, input; red, RDR-optimized two-step enzymatic cycle with $t_d = 0.1$; blue, a cascade of 100-step linear reactions; green, RDR-optimized two-step enzymatic cycle with $t_d = 0$; magenta, literature-parameterized two-step enzymatic cycle. (b) The power spectra of a two-step enzymatic cycle and a 100-step linear networks with various parameterizations. Red, two-step RDR optimized enzymatic cycle with $t_d = 0.1$ (solid), $t_d = 1$ (short-dotted), and $t_d = 0.01$ (long-dotted); green, RDR-optimized two-step enzymatic cycle with $t_d = 0$; magenta, literature parameterized two-step enzymatic cycle; black dotted, two-step enzymatic cycle with literature-parameterized activation rates and RDR-optimized deactivation rates. blue, a cascade of 100-step linear reactions;

We then constructed the power spectrum for both networks to more rigorously analyze transient noise filtering. A power spectrum curve plots the amplitude of

transduced output as a function of its timescales (frequencies). The power spectrum of a transient process typically consists of two portions: a flat region when the stimulation timescale is longer than the process timescale such that the process is in steady-state, and a decreasing region when the stimulation timescale is shorter such that the process is in transient. The transitional frequency between the two portions relates to the division between timescales treated as signal (steady portion) from that treated as noise (transient portion). For our case, we expect this to be affected by t_d value. The slope in the unsteady portion measures the relative effectiveness of a network at rejecting transient noise; a steeper slope translates to a more effective noise filter. The power spectrum curves, shown in Figure 3-3b, possess a log-log scale slope of -6.7986 for the two-step enzymatic network, and one of -1.0000 for the 100-step linear network. This quantifies the superior noise filtering capability of the RDR-optimal enzymatic network compared to the 100-step linear network.

To differentiate the contribution of topology from parameterization in our study of noise filtering effectiveness, we examined power spectra of other parameterizations in both enzymatic and linear networks. To determine the contribution of topology, we parameterized the 100-step linear networks randomly and obtained power spectra with slope of -1.0000 despite having different parameters (data not shown). This slope similarity suggests that the non-linear enzymatic network topology is necessary to observe filtering effectiveness. To determine the contribution of RDR-optimized parameterization, we examined three sets of other parameters in the enzymatic network: first, literature-obtained parameters of the *Xenopus* oocyte MAPK [28] where the deactivation-step kinetics were assumed to be similar to the activation-step kinetics; second, a hybrid of the literature-obtained activation-step from *Xenopus* oocyte MAPK and RDR-optimized deactivation-step parameters; and third, parameters when the network was optimized for rapid *undelayed* response (RDR objective with $t_d = 0$) [24]. From these parameterizations, we obtained power spectrum curves with

slopes of -4.2414 , -6.5604 and -2.3045 , respectively. These data suggest a large contribution of RDR-optimized parameterization, in particular, optimized deactivation rate constants, to filtering effectiveness. Finally, we also examined the sensitivity to delay time (t_d) by re-optimizing for different t_d values (from 0.1 to 1 and 0.01 s) and discovered that the power spectra slopes decrease with t_d ($t_d = 1, 0.01$ produced a slope of $-7.3419, -4.8188$, respectively). Taken together, these results suggest that the non-linear topology of the enzymatic cycle coupled with RDR-optimal parameterization can lead to efficient filtering of transient noise.

RDR response is promoted by the timed depletion of reverse enzyme in optimal enzymatic cycles.

To understand the operational strategy that RDR-optimal two-step enzymatic networks use to produce a sharp delay response, we examined state trajectories and observed a central role for unbound reverse enzyme (R). The trajectories, shown in Figure 3-5c, reveal that when the output started to rise rapidly, the concentration of unbound reverse enzyme (R) dropped to essentially zero. This observation suggests that in an RDR-optimal network, one critical role of R is to delay response; exhausting the reservoir of R at an appropriate time (t_d) allows networks to produce a sharp RDR.

To further characterize the role of the reverse enzyme in generating the RDR, we perturbed topology and initial conditions for the two-step enzymatic networks to identify the contributions of various components in producing the RDR. The network topology is perturbed by collapsing each of the four enzymatic reaction steps (two for activation and two for deactivation) into a linear reaction and re-optimizing for the RDR objective. The results, summarized in Table 3.1, show that perturbation of either or both deactivation steps, but not the activation steps, caused a failure to produce a sharp RDR. Alternatively, we also characterized the role of reverse enzyme

by perturbing its initial condition (R_0). Here, R_0 was perturbed to two cases that qualitatively prevent the network from implementing the optimal RDR strategy, first, when R_0 is in excess, and second, when R_0 is scarce. In the case of excess, depletion is prevented by stoichiometry; in the case of scarcity, there is insufficient material to last through the delay given the available range of rate constants. The results, the last two rows of Table 3.1, show that networks with excess or scarce R_0 could not produce sharp RDR. These results further demonstrate the importance of maintaining and depleting a reservoir of reverse enzyme in the optimal strategy.

Table 3.1: Perturbation of network structure and initial conditions to examine the mechanism of optimal Rapid but Delayed Response (RDR). LLN: the Length of a Linear reaction Network connected in series producing an equivalent f_{RDR}/t_d value.

Network structure	best f_{RDR}/t_d	LLN
one-step enzymatic activation-deactivation	0.05	20
two-step enzymatic activation-deactivation	0.017	190
two-step enzymatic with zero deactivation	0.117	5
two-step enzymatic with first deactivation substituted by a one-step linear reaction	0.089	7
two-step enzymatic with second deactivation substituted by a one-step linear reaction	0.056	17
two-step enzymatic with first and second deactivations substituted by a two-step linear reactions	0.110	5
two-step enzymatic with first and second deactivations substituted and merged to a one-step linear reaction	0.094	6
two-step enzymatic with second step activation reaction substituted by a one-step linear reaction	0.019	150
one-step linear activation-deactivation reactions	0.162	3
two-step enzymatic with reverse enzyme in excess ($\frac{R_0}{A_0} = 1.2$)	0.080	8
two-step enzymatic with reverse enzyme scarce ($\frac{R_0}{A_0} = 0.0003$)	0.087	8

To understand the consequences of this strategy, we simulated the behavior of the RDR-optimal two-step enzymatic network, where the initial concentrations of the forward enzyme F_0 or reverse enzyme R_0 were varied and the output monitored. Because the network exhausts R at a specific t_d , we expected that stimulating the network with different input amplitudes (F_0) or reverse enzyme initial conditions (R_0) would produce responses at different t_d . To test this, we took the parameters

optimized for $t_d = 0.1$ s, $F_0 = 20 \mu M$, $f_{\text{amplitude}} = 10$, and perturbed the model with first, input amplitude $F \in [18, 22.5]$ in 0.5 intervals, and second, initial conditions of the reverse enzyme $R_0 \in [15.6, 16.5]$ in 0.1 intervals. The results, shown in Figure 3-4a and b, produced sharp RDRs at earlier t_d with increased F_0 or reduced R_0 . We observed that the RDRs were still as sharp despite being simulated at conditions different from the ones used in the optimization, suggesting some robustness to the strategy for producing sharp RDRs. Moreover, for low values of F_0 the reservoir of R could not be depleted (by the relative balance of kinetic of activation and deactivation steps), and so a very small steady-state output amplitude was produced. To further investigate the sensitivity of the steady-state output amplitude to input amplitude F_0 , we varied $F_0 \in [10, 100] \mu M$ in $10^{-5} \mu M$ intervals. We found that the output amplitude increased rapidly once F_0 exceeded a threshold, with a maximum slope of at least $1.9 \cdot 10^{+08} \frac{\mu M \text{ output}}{\mu M \text{ input}}$, and the curve shown in Figure 3-4c.

Structure and function relationships of enzymatic cycles measured by their ability to resolve conflicting objectives.

To analyze the trade-off behavior of the one- and two-step enzymatic cycle networks, we looked at the manifestation of the two principles to produce sharp RDR as described above. The two network topologies differ in that the two-step network has an extra enzymatic cycle, and thus, an extra modified state through which the reverse enzyme may interact with substrate to delay response. From these topological differences, we found the two-step network to produce sharper RDR at a wider range of amplitudes than the one-step network presumably because of longer reaction steps in cascade and more freedom for the reverse enzyme to delay response.

To examine the behavior of the networks when faced with possibly competing objectives, we optimized both the one- and two-step enzymatic networks for the RDR objective while constraining the output amplitude to a specified value and varied this

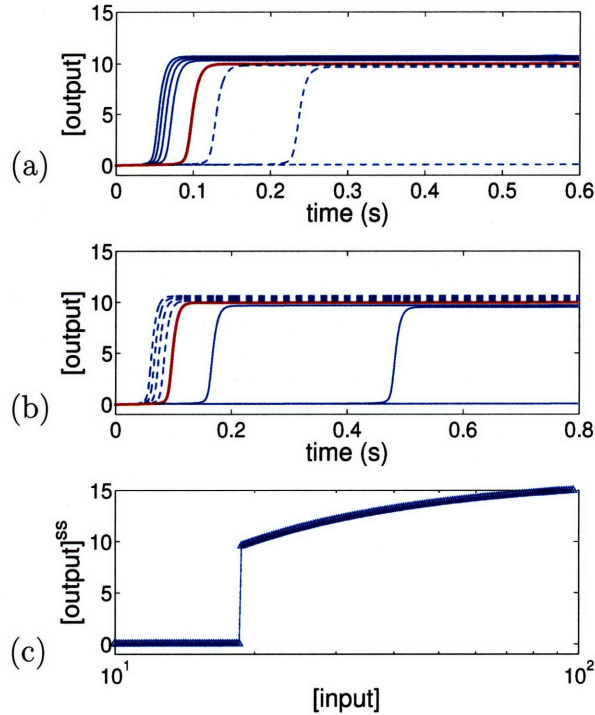


Figure 3-4: The dynamic behavior of an RDR optimal two-step cycle network. (a) Sensitivity of response trajectories to input amplitude (F_0). Red, baseline ($F_0 = 20 \mu M$); blue, increasing (solid) or decreasing (dotted) F_0 in $0.5 \mu M$ interval. (b) Sensitivity of response trajectories to reverse enzyme initial condition (R_0). Red, baseline ($R_0 = 16 \mu M$); blue, increasing (solid) or decreasing (dotted) R_0 in $0.1 \mu M$ interval. (c) The steady-state curve of the RDR-optimal two-step enzymatic network shows an all-or-none behavior.

value in a parametric optimization formulation. We chose this method because other multi-objective methods [129, 98] may be more efficient computationally, but are not well developed for a non-linear dynamically embedded problem as formulated here. The results of this constrained optimization, Figure 3-5a, showed the existence of a trade-off between amplitude and RDR quality. In the absence of the constraint, each network obtained its best RDR optimum (f_{RDR} of 0.0488 for the one-step and 0.00148 for the two-step network); in the presence of an amplitude constraint, each network achieved the same optimum when the appropriate constraint was applied ($f_{\text{amplitude}}$ of 0.502 for the one-step and 18.000 for the two-step network). Interestingly, application of higher or lower amplitude constraints led to optima that scored worse by the RDR objective, which illustrates the presence of a trade-off. The two-step network achieves

a better RDR objective for all values of the amplitude constraint studied, particularly for smaller amplitudes, and has a narrower range of sensitivity. Both networks appear relatively insensitive to the amplitude constraint at low amplitudes.

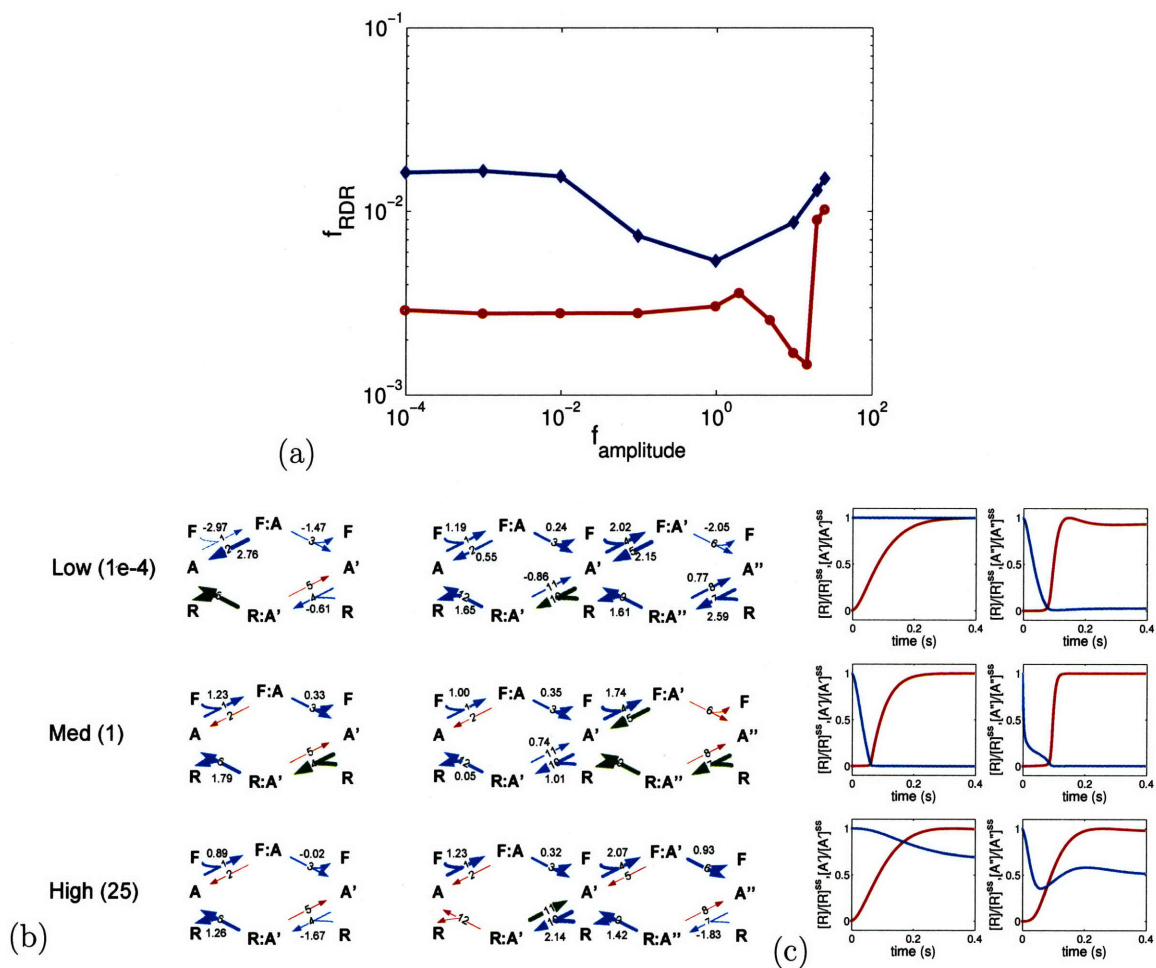


Figure 3-5: Trade-offs in one- and two-step enzymatic cycle networks. (a) Trade-off curves between RDR and amplitude objectives. Blue, one-step enzymatic cycle network; red, two-step enzymatic cycle network. (b) Optimal networks at various amplitudes. Top, low amplitude ($f_{\text{amplitude}} = 10^{-4}$); middle, medium amplitude ($f_{\text{amplitude}} = 10^0$); bottom, high amplitude ($f_{\text{amplitude}} = 25$). Arrow colors indicate the location of optimal rate constant multipliers relative to bounds: red, at lower bound; green, at upper bound; blue, in-between bounds. The number next to blue arrows indicates the optimal value (in log-scale). Arrow thickness is proportional to the log scale of optimal value, normalized by respective lower and upper bounds. (c) Corresponding trajectories of output (red) and unbound reverse enzyme (blue) at low, medium, and high amplitudes.

To delineate the operational strategies that produce these trade-off behaviors, we

examined the optimal one- and two-step enzymatic networks with various amplitude constraints. The optimal networks are shown in Figure 3-5b, and the corresponding trajectories of the output response A' or A'' and the unbound reverse enzyme R are shown in Figure 3-5c. At a low amplitude value, the one-step network selected low activation rate constants; the corresponding trajectories show that the network failed to produce a sharp RDR, and R was barely consumed. These observations are consistent with an interpretation that the low amplitude constraint caused the one-step enzymatic network to fail to produce enough A' to exhaust R . In comparison, the two-step enzymatic network chose high activation and deactivation rate constants for the first enzyme cycle and a low modification rate constant for the second cycle; the corresponding trajectories show that the network produced a sharp RDR, and the R was exhausted. These results suggest that the extra cycle in the two-step enzymatic network allows it to satisfy a low amplitude constraint and implement the RDR optimal strategy simultaneously. At an intermediate amplitude value, both the one- and two-step enzymatic networks produced their lowest RDR objective values, and corresponding trajectories showed that these networks produced sharp RDRs and exhausted R . Furthermore examination showed that both optimal networks chose the maximal association and the minimal dissociation rate constants between R and either A' (for one-step) or A'' (for two-step), suggesting that, when at a non-extreme amplitude value, the ideal network construction is to maximize the binding of the activated substrate to the reverse enzyme. At a high amplitude value, both of these networks chose optimal parameters with a low association rate constant between A' or A'' and R ; the corresponding trajectories show that both networks failed to produce sharp RDRs and to exhaust R . These results are consistent with an interpretation that at high amplitude, substrate conservation forces insufficient amounts of available substrate to exhaust R . These trade-off analyses provided a new perspective on the relationship between observed network topologies and their

ability to satisfy conflicting objectives.

3.5 Discussion

We studied the transient noise filtering capability of enzymatic cycle networks by examining the rapid delayed response (RDR) property. We discovered that the RDR property and transient noise filtering can be enhanced through rate constant optimization of enzymatic cycle networks, and that very short enzymatic cycle networks, which are non-linear, can outperform much longer cascades of purely linear reactions.

The RDR optimal strategy depends on the parameterization of deactivation kinetics but not activation kinetics. We obtained a similar quality of RDR responses when the activation kinetics are either fixed in the optimization or collapsed into a linear reaction, but much worse RDR responses upon perturbation of the deactivation kinetics. Furthermore, the best RDR optimal networks selected maximal association and minimal dissociation rate constants between the activated kinase and the reverse enzyme; interestingly, these optimal rate constants are consistent with the observed tight binding between ERK2 and Pyst1 [77, 115], Hog1 and Ptp2, Ptp3 [122, 26, 117], and Fus3 and Msg5 [59], and a class of phosphatases called STYX [87].

The enzymatic cycle network implements optimal RDR mechanistically by the timed depletion of reverse enzyme. We observed a central role of the unbound reverse enzyme as a reservoir that filters out short duration input noise, and subsequent depletion of unbound reverse enzyme allows rapid response. Moreover, this RDR-parameterization is sensitive to the initial condition of the reverse enzyme; interestingly, tight regulation of the abundance of reverse enzyme has been observed in the *puckered (puc)* gene, a dual-specificity phosphatase [39], and elsewhere [44, 50, 21, 82].

The RDR-optimal strategy described above required some expression of phosphatase in unstimulated cells. Some phosphatases, such as MKP1 and MKP2, have

been shown to be transcriptionally induced [29, 70, 114] and their degradation reduced [71] by activated kinases. These mechanisms of negative feedback do not necessarily conflict with the filtering requirement for the existence of phosphatase in unstimulated cells, as feedback may occur over a longer time scale while filtering may function over a shorter time scale. A study by Brondello, et al. [70] indicated the absence of MKP1 expression in unstimulated CCL39 cells, however, whether the absence of phosphatase expression is applicable to other cell lines and phosphatases remains unclear.

The RDR is complementary to, but different from, ultra-sensitivity [6, 28], which is a steady-state property that measures the sigmoidalness of output as a function of constant input. The RDR is a dynamic property that represents a step-like response to the time duration of an input stimulus. Thus, an RDR network filters short transient noise, and an ultra-sensitive network filters small amplitude steady-state noise. In this study, we observed both properties simultaneously; stimulating an optimal network with step-change input of increasing amplitude produced an RDR with varying delay time and a strong ultra-sensitivity (all-or-none) steady-state behavior. Furthermore, we discovered that by RDR-optimizing the deactivation step rate constants of a two-step enzymatic cycle network originally studied by Ferrell, et al. [28, 73], we obtained a network with an enhanced ultra-sensitive behavior, which may explain the wild type behavior observed experimentally [73].

Additional modification sites resolve conflicting objectives better. When a one-step and a two-step enzymatic cycle networks were compared by their ability to resolve objectives, we found that the two-step cycle produces a good RDR objective in a wide range of amplitudes, while a one-step cycle requires that the amplitudes be at a narrower range. This wider range suggests that multi-site cycle networks possess enhanced flexibility to select amplitude while maintaining good RDR.

The RDR objective itself does not correspond exactly to the problem of filter-

ing transient noise, but it served as a useful surrogate for design purposes. The RDR-optimized enzymatic cycle networks produced good transient noise filtering; interestingly, linear networks with similar RDR objective values did not. We selected the RDR objective formulation because it is computationally easier than alternative formulations, such as measuring the slope in the power spectrum directly. The RDR-objective formulation produced a smooth search space and thus standard non-linear optimization tools (conjugate gradient, SQP) may be used, while optimizing the slope of the power spectrum directly produced a numerical problem with theoretically unknown shape of search landscape and standard tools failed to optimize to reasonable solutions (data not shown).

Transient behavior of signaling networks across multiple timescales has been examined experimentally in the canonical signal transduction system of bacterial chemotaxis [38], however, transient noise filtering properties have not been explored in detail. Here, we looked at how transient filtering functionality affects the construction and operation of enzymatic cycles on short timescales. We looked at principles and mechanisms affecting the observed transient filtering in enzymatic cycle networks and how variant topologies filter out transient noise. Filtering transient noise may be an integral function of signal transduction networks, and, as reflected by its name, the network should transmit information with fidelity to avoid making erroneous decisions, some of which have life-or-death consequences.

The non-linearity of enzymatic cycles allows production of RDR more efficiently than a cascade of linear reactions. The cascade of linear reactions is closely related to the problem of a series of Continuously-Stirred Tanks (CSTs) in chemical engineering, where adding tanks in series produces a steeper RDR with the limit of infinite numbers converging to a Plug-Flow Reactor (PFR). Here, simple non-linearity in the reaction scheme, in the form of bilinear reactions, allows development of an optimal operational strategy with much enhanced performance over the linear case.

We demonstrated increased utilization of existing models using optimization techniques. The optimization framework allows us to decouple the reaction mechanisms, where typically are supported by many publications, from the parameterization of a fixed network topology, where less data are available. We hope expanded utilization of optimization techniques will uncover further design principles in biological applications.

Chapter 4

Compact Modeling through

Protein Domain Network

Representation Applied to Study

Signaling Specificity

Abstract

Modeling is an important tool to study complex signal transduction networks, however, the current standard approach, the Mass-Action (MA) kinetics, generates models of which number of species scales combinatorially to the number of protein domain sites and modification states, and are difficult to build and compute when applied to large networks. Methods to reduce model sizes have been proposed, yet, they require generation of MA models at least initially and the resulting reduced models may be inaccurate at different parameterizations. Here, we propose a method to generate compact models by representing the problem as a protein domain network (DN), with nodes defined both as unbound and bound states of each protein domain in a modification state, and arcs defined as binding and modification reactions. This method generates models whose number of species scales quadratically in the worst case to the number of protein domain sites and modification states, a tremendous saving over the combinatorial scaling in the number of species of the MA models. We evaluated the accuracy of the DN method and its alternative implementation approaches by applying to a suite of test systems, and discovered that: (1) in eight orders of magnitude parameter variations, a particular approach produces trajectories

that are within 2.4% of the ones of an MA model; (2) both MA and DN models possess qualitatively similar parametric behavior; and (3) when compared to reduced MA models obtained from species elimination, DN models are more accurate in six orders of magnitude variation in parameters. We subsequently applied this DN method to uncover determinants of signaling specificity in yeast Mating (M) and Filamentous Growth (FG) pathways, where an MA model is estimated to consist of more than 10^7 species and is too large to simulate, and where a DN model consists of only 132 state variables. From the constructed DN model, we discovered two findings. First, Ste5, a scaffold protein thought of as a determinant for M specificity, is a poor specificity factor when acting as a passive tethering protein. However, if Ste5 promotes cooperative binding to its partners, it would enhance M specificity. Second, the sharing of a binding domain, as observed in the common domain of Fus3 MAP kinase that is shared between its activators, deactivators, and targets, promotes unexpected emergent behavior. Together, these results demonstrate the utility of the proposed DN method to construct models for the study of signal transduction networks which otherwise require too many species to be developed with the MA approach.

Keywords: large scale systems biology, model reduction, computational biology, signal transduction, design principles, signaling specificity

4.1 Introduction

Quantitative modeling is an important tool to study complex biological systems. Modeling complements quantitative and systematic experimentations, captures a subset of current knowledge in the field, and indicates some of the gaps in the systematic picture of our knowledge space. Here, we focus on modeling of signal transduction networks that typically consist of interacting proteins, some of which consists of multiple domain and modification sites. A standard method to model these networks is the Mass-Action (MA) kinetics, where a signal transduction system is represented as a network with nodes defined as protein complexes in each binding and modification states, and arcs defined as binding and modification reactions. A key facet of MA models of signal transduction networks is the large number of protein complex species, which include multiple sites of binding domains and modification sites and corresponding activity, as well as heterogeneous complexes composed of various and

variant binding members. This number of species grows combinatorially with the number of proteins, binding domains, and modification sites, making modeling exercise difficult to build and compute. Furthermore, such a large number of species and reactions makes it a challenge to determine parameters experimentally. While this significant challenge must be fully addressed to understand the biological basis for observed biological complexity, important questions has been probed with relatively minimalist parameter sets for the models, where the binding parameters are limited to paired interactions or cooperativity among few paired interactions [19, 100].

Several approaches have been proposed to manage this complexity, which include automatic generation of MA models with a user interface based on interactions of molecular domains (BIONETGEN [88]) and generation species on-the-fly during a stochastic dynamic simulation (MOLECULIZER, [76]). These approaches produce models with a number of species that scales combinatorially, and thus, may not be practical for a large-sized system. To reduce the size of MA models, several model reduction methods [46, 124, 42, 91, 89] have been developed, which include lumping, sensitivity analysis to eliminate insignificant species and reactions, and time-scale analysis. These approaches are not ideal because they require the generation of an MA model that may be intractable in large networks (at least initially), and furthermore, the methods evaluated the most accurate reduced model at a specific parameterization and thus, may potentially produce inaccuracy when evaluated at different parameterizations. In fact, a recent study have suggested that one of these approaches, the species elimination, may produce grossly inaccurate reduced models when evaluated at different parameterizations [72]. This potential inaccuracy is problematic because, in a typical model development process, we do not have *a priori* information about model parameterization. Thus, we need a reduced sized model that is accurate across different parameterizations.

With an eye toward using abstraction for selective model simplification, which may

lead more directly to intuitive understanding, we have developed and implemented a modeling method based on a reaction network of protein domains that lead to compact models. Our goals were to develop a method that, first, is easily extended to any larger sized networks while avoiding the generation of an MA model, second, is accurate even when parameters are varied, and third, can represent observed interactions in natural systems such as cooperative binding. The applicability of alternative approaches to implement the method are examined and, where appropriate, applied to understand signaling specificity in yeast Mating and Filamentous Growth pathways.

The method is applied to study the signaling specificity in yeast Mating (M) and Filamentous Growth (FG) pathways [52, 59, 109]. In particular, we would like to understand the roles of Ste5, a scaffold protein, and the MAPK common domain [123] in relation to signaling specificity. The Ste5 is accepted to be one of the determinants of M response specificity [85], yet, both pathways are known to be stimulated by Pheromone [109] and both to produce MAP kinase activation via Ste5 [52, 59], suggesting that Ste5 may promote both M and FG responses and thus, may be a poor specificity determinant. These apparent conflicting observations encourage us to study further the role of Ste5 to the specificity of M over FG signals. Moreover, previous studies have uncovered a common domain in MAPK that is shared between MAPK activator, phosphatase, and downstream kinase target(s) [123]. This unusual architecture of the network encourages us to study how the sharing of this common domain may affect network functions.

4.2 Methods

Overview of the protein domain network (DN) representation

We proposed a model reduction method to simulate dynamics of protein complexes based on protein binding domain network (DN) representation. The method can be

expressed mathematically as:

$$\frac{dT_{pk}}{dt} = + \sum_{s \in \sum_j S_{[i,i']}} v_s w_s^{\text{mod}} \quad \forall p \in P, k \in K_p, i \in I_{pk}, i' \in P_i \quad (4.1)$$

$$\frac{dD_i^{\text{unb}}}{dt} = \sum_{i' \in P_i} (-w_{[i,i']}^{\text{on}} + w_{[i,i']}^{\text{off}}) \quad \forall i \in I, i' \in P_i \quad (4.2)$$

$$\frac{dD_{[i,i']}^{\text{bnd}}}{dt} = \sum_{i' \in P_i} (+w_{[i,i']}^{\text{on}} - w_{[i,i']}^{\text{off}}) + \sum_{s \in S_{[i,i']}} v_s w_s^{\text{mod}} \quad \forall i \in I, i' \in P_i \quad (4.3)$$

$$C_s = g_m^s(D_{[i,i']}^{\text{bnd}}, T_{pk}) \quad \forall s \in S, [i, i'] \in S_s^{-1} \quad (4.4)$$

$$w_{[i,i']}^{\text{on}} = f^{\text{on}}(k_{[i,i']}^{\text{on}}, D_i^{\text{unb}}, D_{i'}^{\text{unb}}) \quad \forall i \in I, i' \in P_i \quad (4.5)$$

$$w_{[i,i']}^{\text{off}} = f^{\text{off}}(k_{[i,i']}^{\text{off}}, D_{[i,i']}^{\text{bnd}}) \quad \forall i \in I, i' \in P_i \quad (4.6)$$

$$w_s^{\text{mod}} = f^{\text{mod}}(k_s^{\text{mod}}, C_s) \quad \forall s \in S \quad (4.7)$$

where p indexes proteins; j indexes domain sites; k indexes modification states; s indexes modification reactions; i indexes a domain variable for a particular protein p domain j and modification state k ; T_{pk} is the the total protein p concentration in a modification k state; D_i^{unb} is an unbound form domain variable; $D_{[i,i']}^{\text{bnd}}$ is a bound form domain variable; C_s is a modification specie; $w_{[i,i']}^{\text{on}}$, $w_{[i,i']}^{\text{off}}$ are binding fluxes; w_s^{mod} is a modification flux; P is the set of all proteins; K_p is the set of all modification states of protein p ; I is the set of all protein p domain j in modification state k ; I_{pk} is the set of all domain j for a protein p in modification state k ; S is the set of all modification reactions; P_i is the set of binding partners of domain variable i ; $S_{[i,i']}$ is the set of modification reactions whose modification specie C_s depends on the binding of domain variable i to its partner i' ; S_s^{-1} is the set of $[i, i']$ interactions influencing the modification reaction s .

Counting the numbers of species and reactions and their scaling

To count the numbers of species and reactions for MA and DN models listed in Table 4.1, we examined the implemented models when models are available and estimated the numbers otherwise. To estimate the number of species and reactions, in particular for MA method applied to yeast mating and filamentous growth pathways, we used the following formula: for an MA model of a network containing N^P proteins with each protein p possessing N_p^D number of binding domains, and N_p^M number of modification states, the number of species of an MA model is $O\left(\sum_{p \in N^P} \sum_j^{N_p^D} \binom{N_p^D}{j} \prod_{k \in K_{pj}} N_k^M\right)$, where K_{pj} is the set of proteins in a complex to protein p and occupy j out of possible N_p^B binding sites; the number of reactions of a network with N_{species} number of species are bounded by $o(N_{\text{species}})$ for a sparsely connected network, and by $O(N_{\text{species}}^2)$ for a fully-connected network.

Relationship between Domain Network (DN) and Mass-Action (MA) kinetic models

We started from an MA model given by:

$$\begin{aligned}
 \dot{x}_x &= \sum_i v_{i,x} w_i^{\text{on}} + \sum_i v_{i,x} w_i^{\text{off}} + \sum_i v_{i,x} w_i^{\text{mod}} & (4.8) \\
 w_i^{\text{on}} &= k_i^{\text{on}} x_{i1} x_{i2} \\
 w_i^{\text{off}} &= k_i^{\text{off}} x_i \\
 w_i^{\text{mod}} &= k_i^{\text{mod}} x_i
 \end{aligned}$$

where x_x are MA species; $w_i^{\text{on}}, w_i^{\text{off}}, w_i^{\text{mod}}$ are bimolecular association, uni-molecular dissociation, and uni-molecular modification fluxes, respectively; k are rate constants; and $v_{i,x}$ are reaction coefficients. We lumped MA species (x_x) possessing the same domain state to DN species \hat{x}_d , which are given by $\hat{x}_d = \sum_d M_{dx} x_x$, to Equation 4.8,

to obtain:

$$\sum_p M_{dx} \dot{x}_x = \sum_d M_{dx} \sum_i v_{i,x} w_i^{\text{on}} + \sum_d M_{dx} \sum_i v_{i,x} w_i^{\text{off}} + \sum_d M_{dx} \sum_i v_{i,x} w_i^{\text{mod}} \quad (4.9)$$

which simplifies to:

$$\begin{aligned} \dot{\hat{x}}_d &= \sum_{i'} v_{i',d} w_{i'}^{\text{on}} + \sum_{i'} v_{i',d} w_{i'}^{\text{off}} + \sum_d \sum_i M_{dx} v_{i,x} w_i^{\text{mod}} \\ w_{i'}^{\text{on}} &= k_{i'}^{\text{on}} \hat{x}_{i_1'} \hat{x}_{i_2'} \\ w_{i'}^{\text{off}} &= k_{i'}^{\text{off}} \hat{x}_{i'} \\ w_i^{\text{mod}} &= k_i^{\text{mod}} x_i \end{aligned} \quad (4.10)$$

In Equation 4.10, both MA association and dissociation fluxes ($w_i^{\text{on}}, w_i^{\text{off}}$) reduced to DN binding fluxes ($w_{i'}^{\text{on}}, w_{i'}^{\text{off}}$) that are functions of DN variables only. No simplification is observed for the modification fluxes (w_i^{mod}). Thus, to obtain an exact lumped DN model, we also mapped, one-to-one, the modification species from MA to DN.

Alternative approaches to compute modification species

To compute modification species as a function of domain variables, we examined four alternative approaches that differ on the level of assumptions. Each approach is a special case of Equations 4.1-4.7, and summarized as follow:

1. Approach 1: Differential relationship of modification species and domain variables. The modification species are computed using differential equations with association fluxes that are derived from the differential form of $g_{m,\text{eqm}}^s$ developed in Approach 3. No domain variables are assumed at steady-state. The approach can be expressed mathematically with this equation being a realization of Equation 4.4:

$$\frac{dC_s}{dt} = \sum_{[i,i'] \in S'_s} \left(w_{[i,i']}^{\text{on}} \prod_{[i_2,i'_2] \in S'_s, [i_2,i'_2] \neq [i,i']} \frac{D_{[i_2,i'_2]}^{\text{bnd}}}{T_{p_2}} \right) - \left(\sum_{[i,i'] \in S_s^{-1}} k_{[i,i']}^{\text{off}} \right) C_s - w_s^{\text{mod}}$$

Applied to System I, the expression is:

$$\frac{d[S(1,1)]}{dt} = w_{[S21,K11]}^{\text{on}} \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} + w_{[S11,E11]}^{\text{on}} \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} - (k_{[S11,E11]}^{\text{off}} + k_{[S21,K11]}^{\text{off}})[S(1,1)] - w_{S(1,1)}^{\text{mod}}$$

2. Approach 2: Algebraic relationship of modification species and domain variables. The modification species are computed from algebraic functions of domain variables using $g_{\hat{m},\text{eqm}}^s$ developed in Approach 3. No domain variables are assumed at steady-state. Applied to System I, the expression of $g_{\hat{m},\text{eqm}}^s([S(1,1)])$ below is a realization of Equation 4.4:

$$\frac{[S(1,1)]}{T_{S1}^{\text{tot}}} = \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}}$$

3. Approach 3: Equilibrium binding. Both the unbound and bound domain variables are assumed to be at equilibrium with negligible modification fluxes. The resulting algebraic equations are used to compute the modification species. The approach can be expressed mathematically with these equations replacing Equations 4.2, 4.3, 4.4:

$$0 = \sum_{i' \in P_i} (-w_{[i,i']}^{\text{on}} + w_{[i,i']}^{\text{off}})$$

$$C_s = g_{\hat{m},\text{eqm}}^s(D_{[i,i'] \in S'_i}^{\text{bnd}}, T_{jk})$$

The modification species function $g_{\hat{m},\text{eqm}}^s$ is derived by constructing an MA model, assuming equilibrium binding, and analytically solving a modification specie as a function of the domain variables in MATHEMATICA(R) (Wolfram Research, Inc.). Applied to System I, the function $g_{\hat{m},\text{eqm}}^s([S(1,1)])$ takes a simple form of products of molar fraction of bound domains to a scaffold protein, given by:

$$\frac{[S(1,1)]}{T_{S1}^{\text{tot}}} = \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}}$$

and solved together with these domain variable equations after equilibrium as-

sumption:

$$\begin{aligned}
\frac{dD_{[S11]}^{\text{unb}}}{dt} &= 0 & \frac{dD_{[S11,E11]}^{\text{bnd}}}{dt} &= w_{[S11,E11]}^{\text{on}} - w_{[S11,E11]}^{\text{off}} = 0 \\
\frac{dD_{[S21]}^{\text{unb}}}{dt} &= 0 & \frac{dD_{[S21,K11]}^{\text{bnd}}}{dt} &= w_{[S21,K11]}^{\text{on}} - w_{[S21,K11]}^{\text{off}} = 0 \\
&& \frac{dD_{[S21,K21]}^{\text{bnd}}}{dt} &= w_{[S21,K12]}^{\text{on}} - w_{[S21,K12]}^{\text{off}} = 0
\end{aligned}$$

4. Approach 4: Pseudo steady-state on the bound domains. The bound domain variables are assumed at pseudo steady-state, and the resulting algebraic equations are used to compute modification species. The approach is expressed mathematically with these equations replacing Equations 4.2, 4.3, 4.4:

$$\begin{aligned}
\frac{dD_{[i,i']}^{\text{bnd}}}{dt} = 0 &= \sum_{i' \in P_i} (+w_{[i,i']}^{\text{on}} - w_{(i,i')}^{\text{off}}) + \sum_{s \in S_{[i,i']}} w_s^{\text{mod}} \\
C_s &= g_{m,\text{pssa}}^s(D_{[i,i'] \in S'_s}^{\text{bnd}}, T_{jk})
\end{aligned}$$

Applied to System I, the approach is given by:

$$\begin{aligned}
\frac{dD_{[S11,E11]}^{\text{bnd}}}{dt} &= w_{[S11,E11]}^{\text{on}} - w_{[S11,E11]}^{\text{off}} = 0 \\
\frac{dD_{[S21,K11]}^{\text{bnd}}}{dt} &= w_{[S21,K11]}^{\text{on}} - w_{[S21,K11]}^{\text{off}} - w_{[S(1,1)]}^{\text{mod}} = 0 \\
\frac{dD_{[S21,K21]}^{\text{bnd}}}{dt} &= w_{[S21,K12]}^{\text{on}} - w_{[S21,K12]}^{\text{off}} + w_{[S(1,1)]}^{\text{mod}} = 0 \\
[S(1,1)] &= f_{\text{pssa}}^{S(1,1)}(D_{[S11,E11]}^{\text{bnd}}, D_{[S11,K11]}^{\text{bnd}}, T_{S1}^{\text{tot}})
\end{aligned}$$

The $f_{\text{pssa}}^{S(1,1)}$ is evaluated numerically from the MA model of System I because we could not find any simple analytical algebraic expression.

The detail dynamic models of each approach applied to System I is given in Supplementary Materials.

Accuracy measurements of alternative approaches to implement DN representation

The accuracy of four alternative approaches are examined by comparing the trajectories of state variables and their sensitivities generated by Approaches 1-4 to Approach 0 in System I. We utilized the same underlying parameters to simulate all approaches. We measured accuracy of approaches by comparing trajectories of a downstream state K_p and its sensitivities to rate constants and initial conditions. We compared trajectories generated by Approaches 1-4 to the ones by Approach 0 using the following accuracy measure: $f_{ij}^{\text{accuracy}} = \frac{1}{t_f} \frac{\int_0^{t_f} (x_{ij}(t) - x_{i0}(t))^2 dt}{\max_{t=0}^{t_f} x_{i0}(t)^2}$, where x_{ij} are the i -th state trajectory of Approach j . The integral function is chosen to compare across the whole trajectories instead of at a specific time point. We examined f^{accuracy} in both the canonical parameter and one-dimensional parametric perturbation to all of the rate constants and initial conditions bounded by $[10^{-3}, 10^5]$ fold of the canonical values.

The parametric behavior of Approach 1 is compared to the one of Approach 0 by comparing the shape of a function in parametric space in System II and III. We chose to measure a responsiveness function, given by $f_{ij}^{\text{response}} = \int_0^{t_f} \left(\frac{x_{ij} - x_{ij}^{ss}}{x_{ij}^{ss}} \right)^2 dt$, with i being a state variable, chosen to be K_p^{tot} and j being the approach number.

Comparing the accuracy of DN and species elimination reduced models at parameter variations

To compare the accuracy of DN method to species elimination model reduction methods, we applied both methods to System III. The accuracy is measured using f_{ij}^{accuracy} defined above, with the state variable being compared being the total amount of activated form of C . The fully enumerate MA model is constructed with BIONET-GEN [88]. The reduced MA models are constructed by eliminating species whose trajectories have the smallest norm when the MA model is evaluated at a canonical

parameter value, and the reactions into and out of eliminated species being set to zero. We varied the number of eliminated species to be 25, 50, 100, and 164 species eliminated, the last number is chosen to construct a reduced model containing the same number of species as the DN model.

Model systems

The following four systems are utilized to develop and apply the method and are described as follows:

1. System I consists of three proteins, S, E, K , with S containing two binding sites, each binds E and K . Protein complexes are represented by $S(i, j)$, $i \in \{0, 1\}, j \in \{0, 1, 2\}$, where 0 represents unbound state of a domain, and $\{i, j\} > 0$ represents a bound state of a domain. Protein K may be modified to K_p via the complex $S(1, 1)$. We assumed independent binding in the systems by assigning the same rate constants between paired domain interactions.
2. System II consists of the same components as System I, however, we introduced cooperative binding by assigning different values of dissociation rate constants when both E and K binds simultaneously to S . The values are assigned as $k_{[S11,E11]_{\text{for } [S(1,1)]}}^{\text{off}} = \beta k_{[S11,E11]_{\text{for } [S(1,0)]}}^{\text{off}}$ and $k_{[S21,K11]_{\text{for } [S(1,1)]}}^{\text{off}} = \beta k_{[S21,K11]_{\text{for } [S(0,1)]}}^{\text{off}}$, with β being a cooperativity factor. Applied to this system, Approach 1 is given by the specific realization of Equation 4.4, 4.6 to be:

$$\begin{aligned} \frac{d[S(1, 1)]}{dt} &= w_{[S21,K11]}^{\text{on}} \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} + w_{[S11,E11]}^{\text{on}} \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \\ &\quad - \beta (k_{[S11,E11]}^{\text{off}} + k_{[S21,K11]}^{\text{off}}) [S(1, 1)] - w_{S(1,1)}^{\text{mod}} \\ w_{[S11,E11]}^{\text{off}} &= k_{[S11,E11]}^{\text{off}} (D_{[S11,E11]}^{\text{bnd}} + (\beta - 1)[S(1, 1)]) \\ w_{[S21,K11]}^{\text{off}} &= k_{[S21,K11]}^{\text{off}} (D_{[S21,K11]}^{\text{bnd}} + (\beta - 1)[S(1, 1)]) \end{aligned}$$

Approach 2 and 3 are given by the following specific realization of Equation 4.4:

$$(D_{[S11,E11]}^{\text{bnd}} - [S(1, 1)])(D_{[S21,K11]}^{\text{bnd}} - [S(1, 1)]) = \beta [S(1, 1)](T_{S1}^{\text{tot}} - D_{[S11,E11]}^{\text{bnd}} - D_{[S21,K11]}^{\text{bnd}} + [S(1, 1)])$$

Approach 4 is not applied to this system.

3. System III consists of a dimerizing scaffold protein S with three binding partners A, B, C , with A being defined as an input with one modification state and each of B, C possessing two modification states (unphosphorylated, phosphorylated). Each of B, C can be modified only when it binds to a dimerized scaffolds together with at least one other phosphorylated binding partner of S serving as a kinase cascade.
4. System IV consists of multiple proteins of the yeast Mating (M) and Filamentous Growth (FG) pathways. The scope is defined as the signal transduction unit, from Ste4, a receptor, to Ste12 and Tec1, transcription factors of M and FG. The input is defined as the Ste4 concentration, and outputs are defined as time integral of active Ste12 homo-dimers binding to Mating genes for M, and as time integral of active Ste12 and Tec1 hetero-dimers binding to FG genes for FG.

The DN method applied to yeast Mating and Filamentous Growth pathways

Model scope. The scope of the network is from Ste4 to Ste12 and Tec1 binding to mating and filamentous growth genes. Each of the kinases in the cascade (Ste11, Ste7, Fus3, Kss1) are modeled with three modification states, unmodified, singly phosphorylated, and doubly phosphorylated forms. The rest of the kinases are modeled with two modification states. The input is Ste4, the outputs are time integral of activated Ste12 homo-dimer bound to mating gene, and time integral of activated Ste12 and Tec1 bound to filamentous growth gene. The integral function is selected to evaluate over the whole trajectory instead of just at a specific time point. Each protein interaction is represented by a separate domain in each protein, with the exception

of Fus3 and Kss1 MAPKs that are known to share domain between their activator, phosphatase, and downstream kinases [123]. In the dissection part, we also created a version of the model with Fus 3 binding separate domains to Ste7 and to Ste12, Tec1, and Msg5. Only Fus3 has a known phosphatase (Msg5), for the other kinases, we assumed uni-molecular dephosphorylation reactions. We included the Fus3 promoted degradation of Tec1 that were found [93, 108] to be a major determinant of specificity in these pathways. To balance the degradation, we added a small constant synthesis term for Tec1. For other proteins, we assumed no synthesis and degradation terms. The binding cooperativity of Ste5 to activated Ste7 and Fus3 is implemented by assigning a multiplier β in front of the dissociation rate constants of Ste5 to activated Ste7, and Ste5 to Fus3.

Parameterization. The model parameters consist of rate constants and initial conditions. The initial conditions are obtained from a global yeast expression experiment [110], and being converted from abundance to concentration using the typical yeast cell volume [97] of 37.5 *fL*. The rate constants are manually adjusted to produce model satisfying the experimentally observed phenotypic behavior, summarized in Table 4.3.

To evaluate quality of a parameterization, we developed both unperturbed and perturbed cases of the model with the same underlying parameterization (except for the specific perturbation in a case) and compared the observed states using formulation developed in Table 4.4. The quality of a parameterization is evaluated by simultaneously simulating the whole set of model cases. All the set of case models consist of 18278 state variables. The simulation is implemented in ABACUSSII [67] and each run takes 800 s in a 2.8-GHz Intel Pentium III Xeon processor.

We chose the following five general rules to select parameter values. First, association rate constants are selected at a typical value for two proteins. Second,

dissociation, phosphorylation, and dephosphorylation rate constants of Ste4, Ste5, Ste20, Ste11, Ste7, Ste12, Tec1 are selected to produce sufficient activation for a given input range. Third, dissociation rate constants of Ste5 to Fus3 are smaller than the one of Ste5 to Kss1. Forth, interactions of Fus3 and Msg5 are parameterized with an all-or-none dephosphorylation rate constants developed recently [25]. This parameterization is chosen to allow added differentiability of low and high input responses that are lost with added cascade length. Msg5 initial condition is adjusted to provide an appropriate threshold value. Fifth, interactions of Fus3 to Ste12 and Fus3 to Tec1 are chosen as a fraction of the ones of Fus3 to Msg5 to reduce interference of the all-or-none behavior. We adjusted a total of 63 rate constants.

To obtain a set of parameterization satisfying the phenotypic constraints formulated above, we adjusted some rate constants by 10 or 100 fold and re-evaluate the constraints. We chose the rate constants being adjusted based on our understanding of the network operation. A more rigorous method to parameterize using constrained optimization may be possible but its development is beyond this paper.

4.3 Results

The protein Domain Network (DN) representation is a method to produce compact and accurate dynamic models of signal transduction networks.

To model signal transduction networks while avoiding the combinatorial explosion in the number of species of the Mass-Action (MA) kinetics method, we studied an alternative problem representation to construct dynamic models, called the protein Domain Network (DN). In this method, signal transduction is represented as a network, with nodes defined as both an unbound form of each protein domain in each

Table 4.1: Test systems being used to validate and apply protein-domain network representation of dynamic models. Numbers in parentheses indicate the number of fluxes and additional variables implemented for numerical stability. An asterisk in the number indicates an estimate.

Sys.	Descriptions	Number of state variables (fluxes)	
		MA models	DN models
I	A scaffold protein with two partners	9 (15)	10 (7)
II	A scaffold protein with two partners and binding cooperativity	9 (15)	10 (7)
III	A dimerizing scaffold protein with three binding partners, each with two modification states	194 (1686)	30 (14)
IV	Mating and Filamentous Growth pathway, with non-dimerizing Ste5	$3 \cdot 10^3(10^3 - 10^7)^*$	131 (764)
V	Mating and Filamentous Growth pathway, with dimerizing Ste5	$2 \cdot 10^7(10^7 - 10^{14})^*$	132 (770)

modification state and multiple bound forms to a partner, and arcs defined as binding and modification fluxes. With this definition, we produced DN models whose number of state variables scales quadratically to the number of protein domains and modification states in the worst case, a tremendous saving over the combinatorial scaling of the MA method. Applied to test systems summarized in Table 4.1, the DN method produces much less species than the MA method; for example, when applied to yeast Mating and Filamentous Growth pathways, a DN model consists of 132 species while an MA model requires more than 10^7 species and is too large to develop.

To develop the DN method, we examined the relationship between MA and DN models by applying both methods to a test system (System I), with reaction schemes illustrated in Figures 4-1a and b, and models provided in Figure 4.2. Our goal is to develop a DN method that produces models that most accurately representing corresponding MA models while avoiding any construction of MA models. In the DN method, the above node and arc definitions qualify the method as a lumping approach, where MA model species are lumped into DN model species, with a lumping procedure being guided by biological insights. This linear lumping is given by $\hat{x} = Mx$, where x

are MA species, \hat{x} are DN variables, and M is a linear lumping matrix. In a general lumping approach [63, 46], an application of M may produce reduced models that are either exact or approximate. When the DN's M was applied to the System I MA model, we discovered that the resulting DN model may be exact: the DN model consists of binding fluxes which reduced to a function of only DN variables, and a modification flux that remains a function of an MA specie (x_s); and this latter flux may be computed exactly when x_s is also mapped one-to-one from MA to DN (Methods, Supplementary Materials). Because we required that any DN models to be functions of DN variables only to avoid any construction of MA models that may be infeasible, we must compute x_s from DN variables.

To compute x_s , we considered a reverse lumping function that computes any x from \hat{x} , given by $x = g_{\hat{m}}(\hat{x})$. If the reverse lumping function $g_{\hat{m}}$ is exact, we may compute x_s exactly, and we would obtain an exact DN model. Thus, we may rank alternative $g_{\hat{m}}$ functions based on how close they are to exact functions and how extend-able they are when applied to large networks. The previous lumping method [46] has developed a linear function $g_{\hat{m}}(\hat{x}) = \hat{M}\hat{x}$ based on a generalized inverse of M ($M\hat{M} = I$); when applied here, we could not obtain a rule from the resulting \hat{M} that is extend-able to large networks (Supplementary Materials). To obtain $g_{\hat{m}}$ functions extend-able to large networks, we used some intuition about the systems which produced non-linear approximate functions, described below; however, first, we must extend the necessary and sufficient condition for the existence of an exact lumping dynamics when $g_{\hat{m}}$ is non-linear. We were able to extend the proof of the exact lumping method for any non-linear $g_{\hat{m}}(\hat{x})$ satisfying $Mg_{\hat{m}}(\hat{x})M = M$, with details given in the Supplementary Materials.

Within this DN method, we examined four alternative reverse lumping functions $g_{\hat{m}}$ by evaluating their accuracy and extendability as applied to test systems. The best of these functions was used to model signaling specificity of the yeast Mating

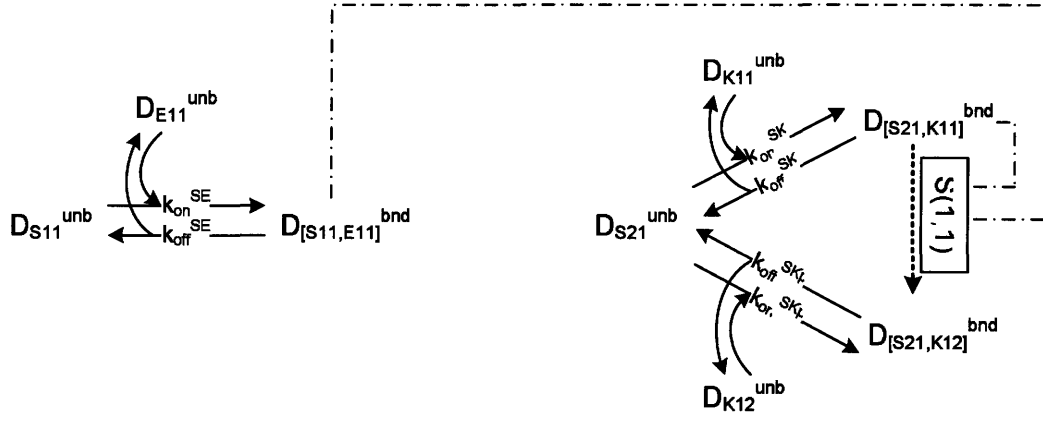
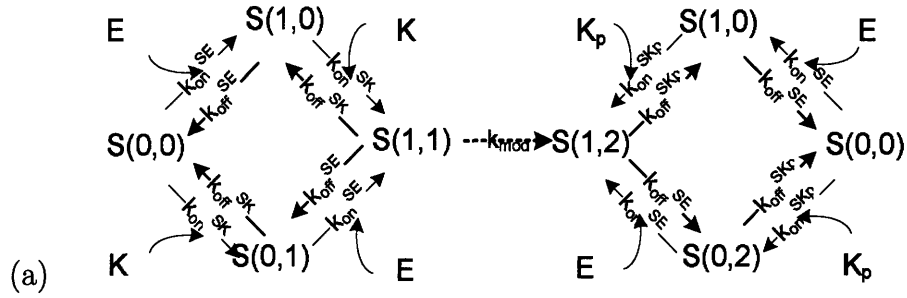


Figure 4-1: Reaction schemes for System I. (a) The mass-action kinetics representation. The network consists of a scaffold protein S with two binding partners E and K ; Protein K may be modified to K_p when both E and K binds to S . Protein complexes are represented as $S(i, j)$, where i indicates states of the first domain of S (0, unbound; 1, bound to E) and j indicates states of the second domain of S (0, unbound; 1, bound to K ; 2, bound to K_p). Solid lines, binding fluxes; short dotted lines, modification fluxes. (b) The protein-domain network representation. D_i^{unb} , an unbound form domain variable i ; $D_{[i,i'] }^{\text{bnd}}$, a bound form domain variable of i to its partner i' ; Solid lines, binding fluxes; short dotted lines, modification fluxes; dash-dotted lines, dependencies of a modification specie to bound-state protein domains.

and Filamentous Growth pathways. The MA (Approach 0) includes the most detailed treatment used throughout this study. Approaches 1-4 simplify Approach 0 in ways that are described in depth in the Methods section and Supplementary Materials, briefly indicated here:

1. Approach 1: The x_s are computed *differentially* from bound domains with expressions of association fluxes derived from Approach 3; however, domains are not in steady-state.

2. Approach 2: The x_s are computed *algebraically* from bound domains with expressions of x_s obtained from Approach 3; however, domains are not in steady-state.
3. Approach 3: The x_s are computed *algebraically* from bound domains, with expressions obtained by assuming that domains are in binding equilibrium (steady-state) with negligible modification fluxes.
4. Approach 4: The x_s are numerically computed from bound domains, and the bound domains are in steady-state without neglecting modification fluxes.

Numerical studies uncover the accuracy and limitations of the DN method.

To evaluate the accuracy of Approaches 1-4, we applied them to System I, comparing trajectories of a state variable and its parametric sensitivities to the ones of Approach 0 under equal underlying parameters. The accuracy was measured by a normalized integrated squared difference of the trajectories (f^{accuracy} , Methods). We examined the accuracy for canonical parameter values derived from the yeast Mating pathways (Supplementary Table 4.5) as well as one-dimensional parametric perturbations in the range of $[10^{-3}, 10^5]$ fold of their canonical values. The results, shown in Figures 4-2a-c for a state trajectory and Figures 4-2f-h for a parametric sensitivity trajectory, demonstrate that Approach 1 is the most accurate (lower f^{accuracy}), followed by Approaches 2, 4, and 3. We observed consistent ranking in other parameter perturbations being studied (data not shown). This ranking is also consistent with the ranking of weaker approximations being applied; for example, Approach 2 should be more accurate than 3 because the earlier does not neglect modification fluxes while the latter does, and Approach 1 should be more accurate than 2 because the earlier assumes molar fractional expression in the association fluxes only while

the latter assumes molar fractional expression throughout. At the canonical parameterization, Approaches 1-4 produced f^{accuracy} of $10^{-3.24}$, $10^{-2.85}$, $10^{-1.35}$, and $10^{-2.53}$, respectively, which translates to integrated differences of 2.40%, 3.76%, 21.1%, and 5.43%. Furthermore, Approach 1 produced a worst case accuracy across examined parameterizations ($f^{\text{accuracy}} = 10^{-2.94}$, when $k_{\text{off}}^{\text{SK}}$ is 10^{-3} fold) that is better than that of other approaches ($f^{\text{accuracy}} = 10^{2.3}$ when E_0 is 10^{-3} fold for Approaches 2-4). To provide further intuition to the values of f^{accuracy} , we examined trajectories for two parameterizations: first, at the canonical values, and second, at the values where the accuracy of Approach 1 is the worst. The results, shown in Figures 4-2d,e,i,j, indicate that the trajectories of Approach 1 reproduce the trajectories of Approach 0 with a maximal trajectory differences at all times of 6.82% (for canonical value parameterization) and 20.0% (for worst case parameterization). We concluded that Approach 1 is superior to the other approaches, and we are interested to characterize more about parametric dependence of its accuracy.

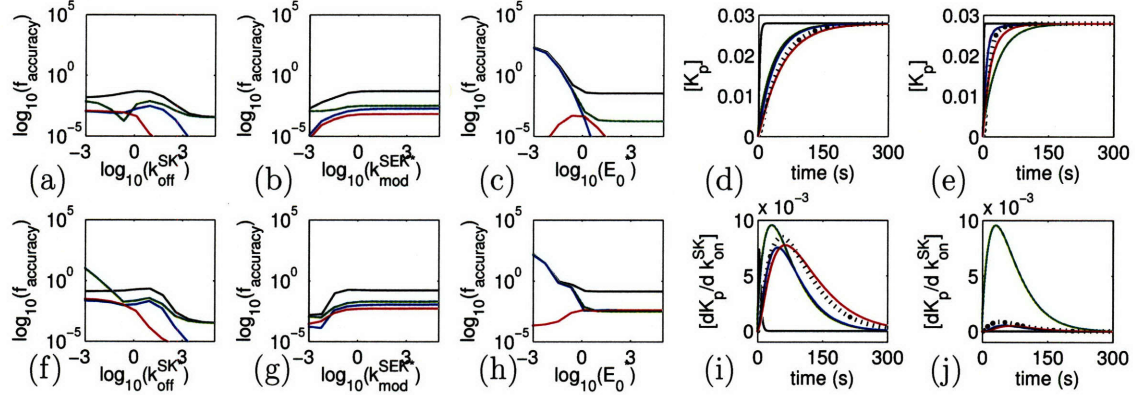


Figure 4-2: Accuracy quantification of four approaches (Approaches 1-4) to implement DN representation applied to System I. Accuracy is measured by integrated squared difference of a trajectory of an approach to the one of Approach 0 (f^{accuracy}); (a-c), accuracy of a state (K_p) as a function of a network parameter; (f-h), accuracy of a parametric sensitivity ($\frac{\partial K_p}{\partial k_{\text{on}}^{\text{SK}}}$) as a function of a network parameter; (d), trajectories of state K_p at the canonical parameter values; (e), trajectories of state K_p when $k_{\text{off}}^{\text{SK}}$ is 10^{-3} times its canonical value; (i,j), trajectories of parametric sensitivity $\frac{\partial K_p}{\partial k_{\text{on}}^{\text{SK}}}$ at parameters defined in (d) and (e). Legends: dotted black, Approach 0; red, Approach 1; blue, Approach 2; black, Approach 3; green, Approach 4.

To explore the limitations of Approaches 1-4, we examined the trends indicating how accuracy changes with parameters. Theoretically, because Approach 1 is derived from Approach 3, which assumes negligible modification fluxes, we expect less accuracy as the ratio of modification fluxes to binding fluxes becomes more significant. The resulting accuracy trends, shown in Figures 4-2a-h, agree with the expected behavior; less accuracy is observed at slower dissociation (Figures 4-2a,f) and faster modification rate constants (Figures 4-2b,g).

To examine how Approaches 1-4 may be extended to larger networks, we looked into the resulting equations when applied to System I (Methods and Supplementary Materials). When Approach 3 is applied to System I, the resulting algebraic equation shows that one may compute any combinatorial binding complex from a multiplication of molar fractional domain variables to the scaffold protein (Methods, Supplementary Materials). To examine whether this molar fractional expression can be extended to larger networks with independent bindings, we analyzed systems with more domains of a scaffold protein, more modification states of a partner, and a case when scaffold protein is also modified. We proved that the molar fractional expression can be extended to these larger systems (Supplementary Materials) and thus Approach 1-3 may be extended to any larger networks with independent bindings. In contrast, Approach 4 computed x_s numerically, so we were unable to extend it to larger networks.

The extension of Approaches 1-4 to networks with cooperative binding was similarly analyzed. For Approach 1, a cooperative binding may be expressed by modifying the dissociation fluxes of x_s differential equation and the binding fluxes of domains that bind cooperatively (equations are in Methods and Supplementary Materials). For Approaches 2 and 3, the presence of cooperative binding requires re-derivation of the algebraic expression to compute x_s from domain variables (equations are in Methods and Supplementary Materials). This re-derivation makes extension of Ap-

proaches 2 and 3 to large networks with cooperative binding difficult. We were unable to extend Approach 4 to networks with cooperative bindings.

To explore the accuracy of Approach 1 to represent networks with cooperative binding, we applied Approach 1 to model a system with a similar reaction scheme to System I and cooperative binding when both E and K bind to S (System II). The cooperative binding was implemented by reducing dissociation rate constants of $[S(1, 1)]$ by a factor β when compared to the constants of $[S(1, 0)]$ and $[S(0, 1)]$. This cooperative binding is represented in Approach 1 by multiplying β to the dissociation fluxes in the differential equation of $S(1, 1)$ and modifying the binding fluxes of domains that bind cooperatively (the DN model is provided in the Supplementary Materials). We examined two properties of the Approach 1 model, its accuracy, and parametric dependence of its dynamic behavior compared to Approach 0 model. The accuracy is examined for Approach 1 as well as Approach 2 by randomly varying all parameters simultaneously and recording the resulting trajectory when compared to Approach 0 using f^{response} measure. The results, whose histograms are shown in Figure 4-3, demonstrate that both Approach 1 and 2 produced an average f^{accuracy} of $10^{-16.6}$, $10^{-16.1}$, and worst case runs ($f^{\text{accuracy}} \geq 10^{-5}$, the cut-off value is chosen arbitrarily) were observed 225/5000 and 548/5000 times, respectively, suggesting that Approach 1 is more accurate than Approach 2. The dynamic behavior is represented as an arbitrarily chosen responsiveness function (f^{response} , Methods), whose behavior is examined in two-dimensional parametric variations bounded by $[10^{-3}, 10^5]$ fold of the canonical values. The results, given in Figures 4-4a and b, show that the two models produce similarly shaped contour plots. When both models are compared, the resulting f^{response} values are within 28.12% on average, with a maximal difference of 127.19%.

To examine the utilization of Approach 1, we compared the parametric behavior of dynamic models of Approaches 0 and 1. We applied both approaches to a larger

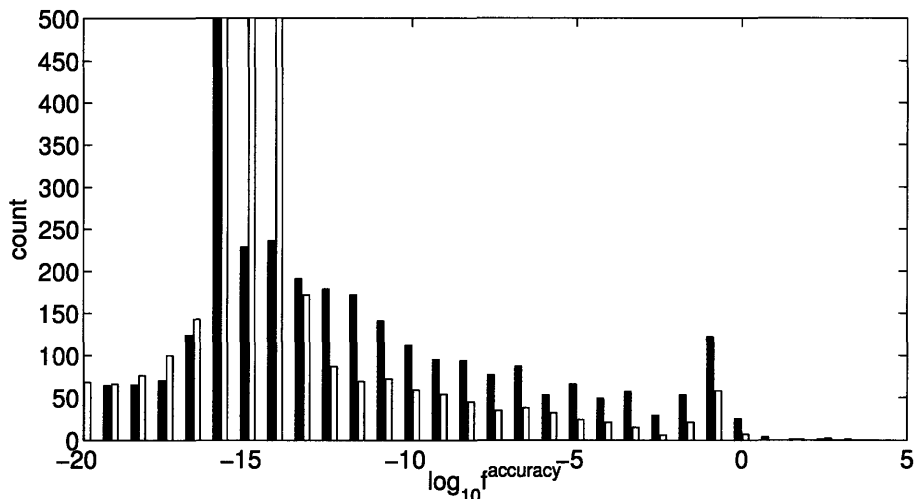


Figure 4-3: The histogram of f^{accuracy} of Approach 1 and 2 applied to System II, for randomized parameter variations in six-orders of magnitude ranges. Legends: white, Approach 1; black, approach 2.

network consisting of a dimerizing scaffold protein Ste5 with three binding partners, the first partner as an input with one modification state, and the next two partners as a kinase cascade with two modification states in each partner (System III). In this system, the Approach 0 model contains 194 species and 1686 fluxes, while the Approach 1 model contains 30 species and 14 fluxes (models are provided in Supplementary Materials). To examine the parametric behavior of both models, we varied two arbitrarily chosen parameters and measured an arbitrarily chosen responsiveness function (f^{response} , Methods). The resulting parametric behavior, two of which are given in Figures 4-4c and d, show that both models also generate similar shaped contour plots. Furthermore, the resulting f^{response} values differ by a mean of 8.29% and a maximum of 34.3%.

To compare the accuracy of Approach 1 to the species elimination method when parameters are varied, we simulated, with parameterizations randomly and simultaneously varied in six orders of magnitude, both System III's Approach 1 model and reduced MA models with varying numbers of eliminated species. We examined the reduced MA models with 25, 100, and 164 eliminated species, with the eliminated

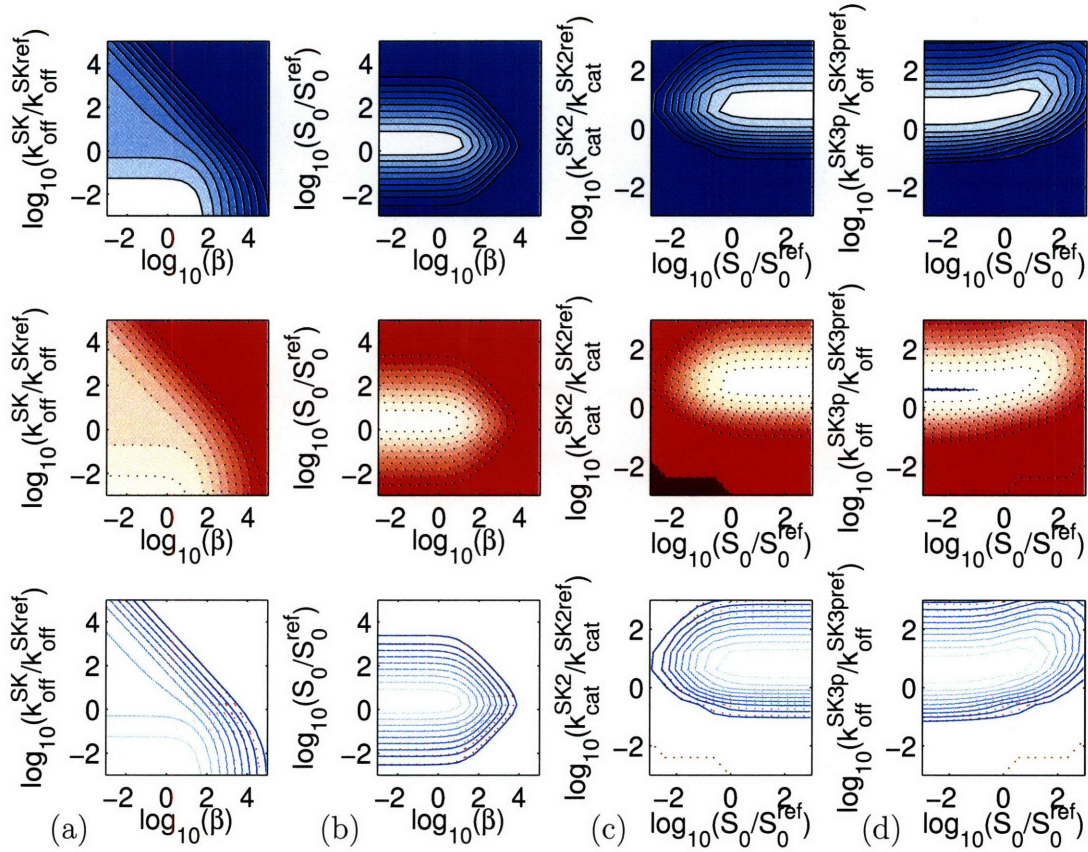


Figure 4-4: Comparing parametric behavior of Approach 1 to the ones of Approach 0. Both approaches are applied to a system similar to System I with added binding cooperativity (System II) and to a system of dimerizing scaffold proteins with four kinase partners (System III). (a) Contour plot of a responsiveness function (f^{response}) as a function of a cooperative factor (β) and a dissociation rate constant ($k_{\text{off}}^{\text{SK}}$) applied to System II. Top, Approach 1; middle, Approach 0; bottom, both; Legends: blue, Approach 1; red, Approach 0; (b) Contour plot similar to (a) as a function of cooperative factor (β) and an initial condition (S_0). (c) Contour plot for System III as a function of an initial condition (S_0) and a modification rate constant ($k_{\text{cat}}^{\text{SK}2}$). (d) Contour plot similar to (d) as a function of an initial condition (S_0) and a dissociation rate constant ($k_{\text{off}}^{\text{SK}3p}$).

species chosen to be MA species with the smallest trajectory norm at the canonical parameterization. The results, given in Figure 4-5, show that the reduced MA models with an increased number of eliminated species produce higher frequency of inaccurate simulations; and the peak frequencies are observed at the worst case f^{accuracy} possible ($f^{\text{accuracy}} = 10^0$) for 100 and 164 eliminated species models. In comparison, the Approach 1 model simulations are most frequently observed at $f^{\text{accuracy}} = 10^{-4}$, a

much lower value than the the one of the reduced models. When both approaches are compared near the worst case accuracy ($10^0 \geq f^{\text{accuracy}} \geq 10^{-0.25}$), the Approach 1 model is observed 120 times out of $5 \cdot 10^4$ runs, while the reduced MA models with 25, 100, and 164 eliminated species are observed 149, 806, and 1553 times, respectively. This suggests that the Approach 1 model is less likely to produce grossly inaccuracy solutions than the species eliminated reduced MA models.

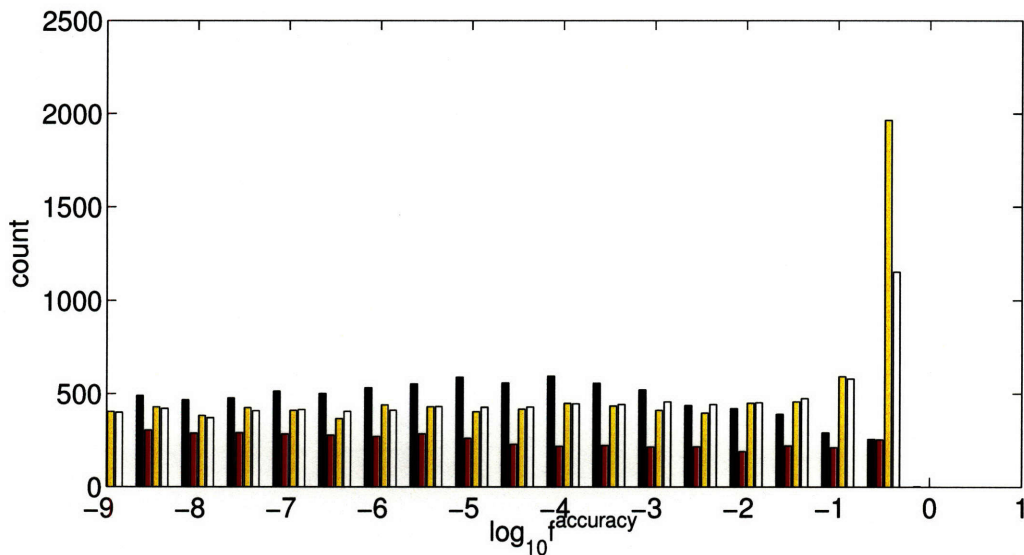


Figure 4-5: Accuracy of Approach 1 compared to species elimination method applied to System III. Black, Approach 1 (with 31 state variables); red, reduced MA model with 25/194 species eliminated; yellow, reduced MA model with 100/194 species eliminated; white, reduced MA model with 163/194 species eliminated.

Application of domain-network representation to study signaling specificity in yeast Mating and Filamentous Growth pathways.

We subsequently applied Approach 1 to study signaling specificity in the yeast Mating (M) and Filamentous Growth (FG) network. The network scope, whose overview is given in Figure 4-6 and its DN method scheme is given in Figure 4-7, is limited to be the signal transduction unit from Ste4, a membrane receptor acting as an input,

kinase cascades that may be activated with and without Ste5 scaffold protein, and Ste12 and Tec1 transcription factors acting as outputs for M and FG responses. Each interaction in this network is assigned a separate domain in the participating proteins, except Fus3 and Kss1 that each shares a common domain for its interactions with activators, phosphatases, and target kinases [123]. We initially assumed independent binding in the model. To examine the significance of Ste5 dimerization, we developed two model versions, with (System IV) and without Ste5 scaffold protein dimerizing (System V). The resulting Approach 1 model consists of up to 132 state variables and 902 fluxes, while an Approach 0 model may consist of more than $2 \cdot 10^7$ state variables and $2 \cdot 10^7 - 7 \cdot 10^{14}$ fluxes.

The Approach 1 model is parameterized to reproduce experimentally observed phenotypic behavior [52, 59, 93, 108] summarized in Table 4.3. The parameters consist of initial conditions and rate constants, with the conditions obtained from global expression measurement [110] and the rate constants manually adjusted guided by literature [54] and given in the Supplementary Material. To evaluate the quality of the parameterizations, we developed perturbation cases from the model that correspond to mutational and over-expression studies performed experimentally. We evaluated the behavior of these perturbed models to determine if they can reproduce experimentally observed behavior; these behaviors are expressed as constraints summarized in Table 4.4. We found a set of 11 parameterizations, with some rate constants differ by at least 10 fold, that satisfy all the constraints.

To determine the role of Ste5 scaffold protein, we varied its expression simultaneously with the network input. The results, Figure 4-8a, showed that the M response forms a single peak at high input, intermediate Ste5 expression, while the FG response forms a ring to the lower input, both at lower and higher Ste5 expression to M response. When we examined both responses at the canonical value of Ste5 expression, the network produced FG response at low input, and M response at high input.

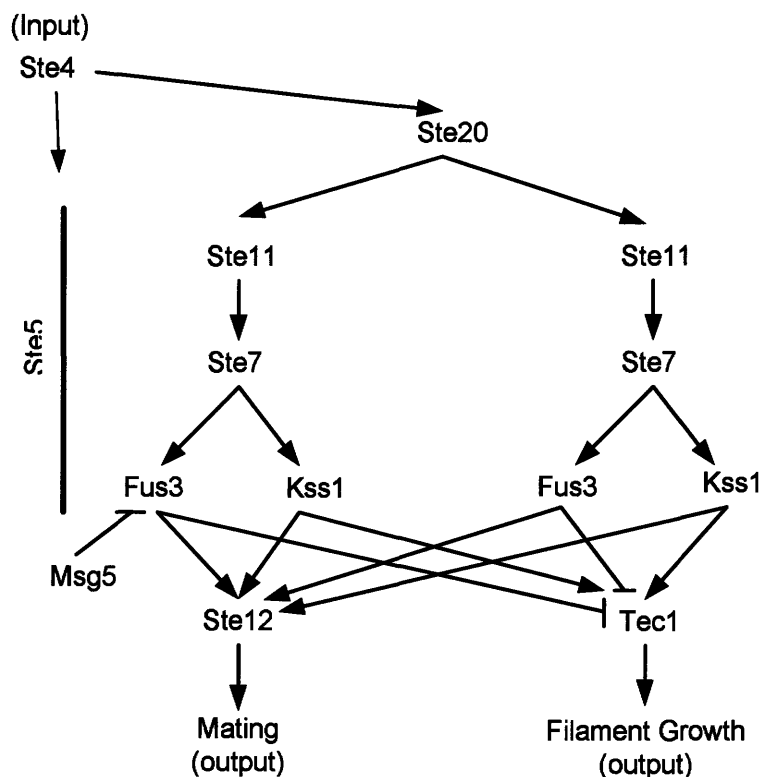


Figure 4-6: The scope of yeast Mating (M) and Filamentous Growth (FG) pathways studied in this paper. The network consists of a G-protein receptor (Ste4), a cascade of kinases that can be activated in the presence and absence of a scaffold protein (Ste5), two MAP kinases (Fus3 and Kss1) for each pathway, and two transcription factors for each (Ste12 and Tec1). The input to the network is the amount of activated Ste4 and it is shared between both pathways, and the outputs are defined as time-integrated of activated Ste12 homo-dimers binding to Mating gene for M pathway, and hetero-dimer Ste12-Tec1 binding to Filamentous Growth gene for FG pathway.

The FG response contour plot suggests that FG are activated at some over-expressed Ste5, contrary to the assigned function of Ste5 as the determinant for M response specificity. We repeated the analysis to a model with dimerizing Ste5 (System V, Figure 4-8b), and to 10 other feasible parameterization of the network (Supplementary Materials), and we consistently observed the high FG response at high Ste5 expression. We hypothesized that the trend is a consequence of non-cooperativity in the model, however, we do not have a rigorous computational tool to rule out other parameterizations in this model that may produce similar FG response behavior.

To determine if Approach 1 may represent cooperative binding in this system, we

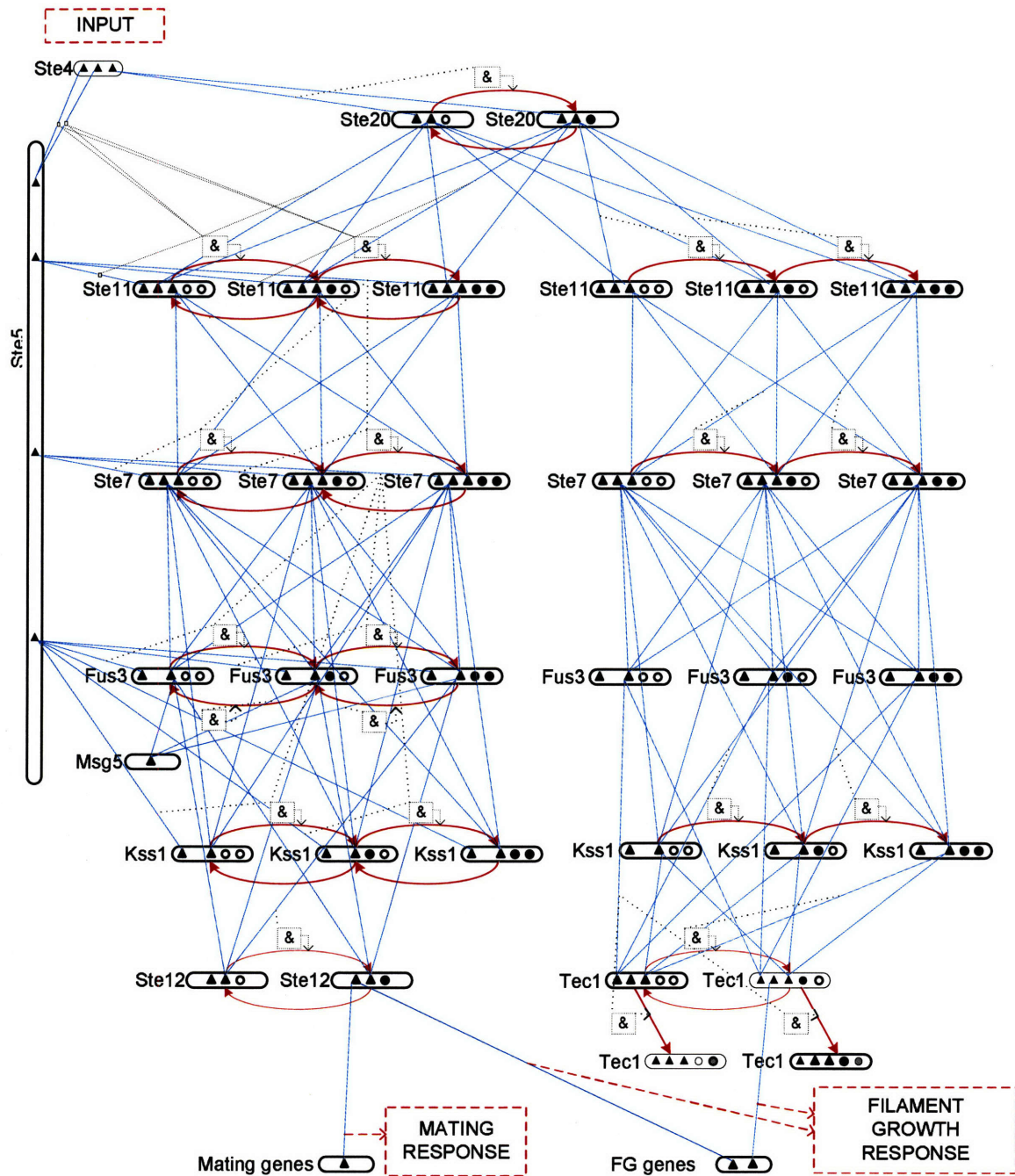


Figure 4-7: The reaction scheme in domain network representation of the yeast Mating (M) and Filamentous Growth (FG) pathways. Legends: triangle, a protein domain; open circle, an inactive modification state; solid circle, an active modification state; blue line, a binding interaction; red line, a modification reaction; black dotted line, a dependency interaction between bound domains and modification species; &, combination of multiple dependency interactions to compute a modification specie.

modified the network to include cooperative binding between Ste5, activated Ste7, and Fus3. We expected that increased cooperativity would reduce FG response because

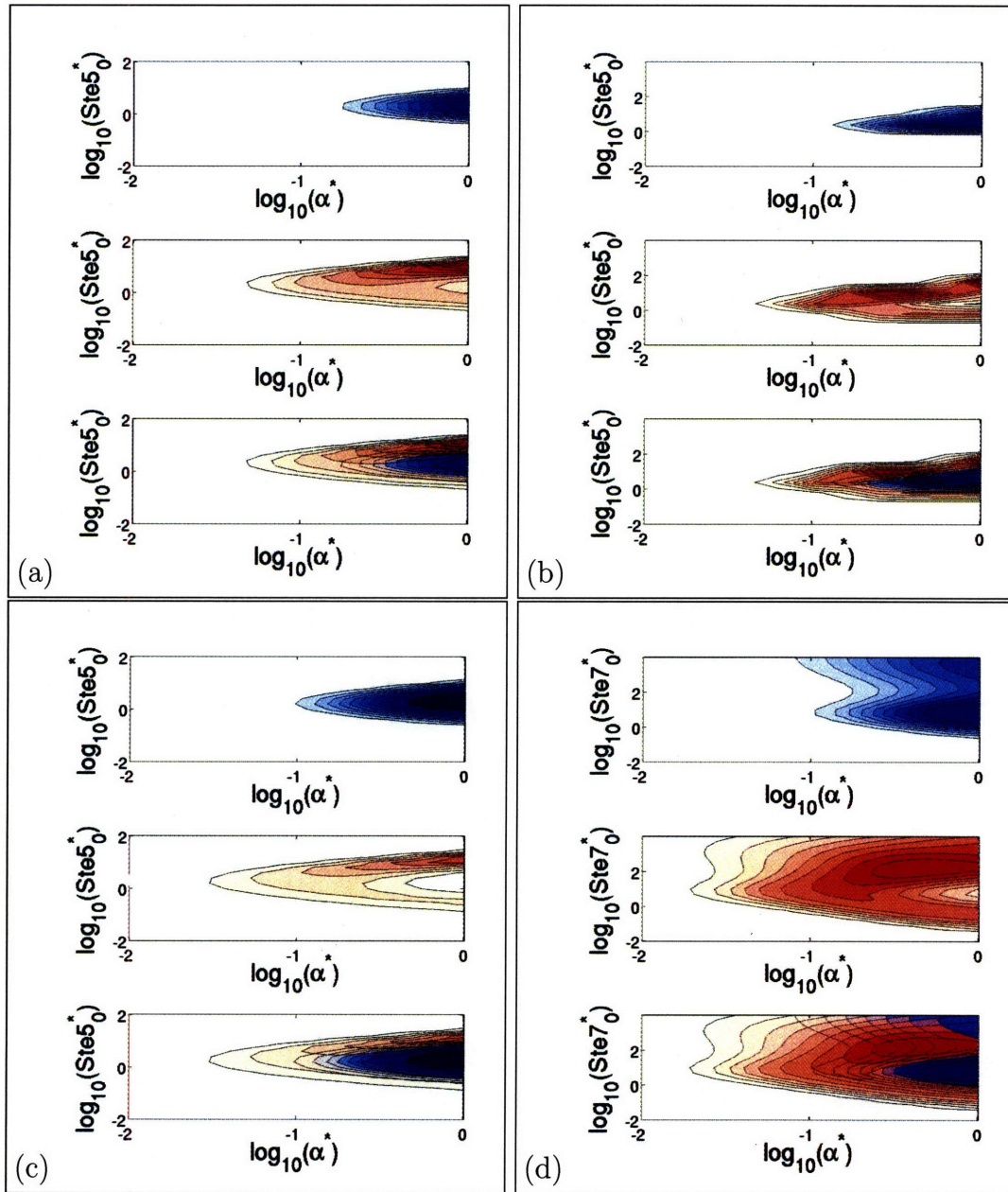


Figure 4-8: Yeast Mating (M) and Filamentous Growth (FG) signal specificity as a function of selected protein expressions and pathways input. (a) M and FG responses as a function of Ste5 for a parameterization on a non-dimerizing Ste5 model (System IV); (b) similar as (a) on a dimerizing Ste5 model (System V); (c) similar as (a) but with added cooperative binding between activated Ste7, Ste5, and Fus3, implemented by multiplying relevant dissociation constants by 0.3 fold of values in (a); (d) M and FG responses as a function of Ste7 in System IV. Legends: blue, M response; red, FG response. α^* , normalized input; $Ste5_0$, normalized Ste5 initial condition. Top, M response only; middle, FG response only; bottom, an overlay of both M and FG responses.

increased cooperativity would increase Fus3 activation, which consequently increase Fus3-promoted Tec1 degradation and decrease FG response. The results, Figures 4-8c, showed reduced FG response at high Ste5 with increased cooperativity, consistent with our expected behavior.

To determine the role of MAPK common domain, we varied the expression of the domain's activator, Ste7 MEK, and the network input in the Approach 1 model and examined the responses. The results, Figure 4-8d for System IV and a similar figure for System V (data not shown), showed that M response produced two activation peaks at high input, both low and high Ste7 expression, and FG response forming activation peaks at the lower input surrounding the lower peak of M response. Our hypothesis is that the two M response peaks are a consequence of the Fus3 common domain; increased Ste7 will occupy Fus3 domain more and out-compete both Msg5, a phosphatase and negative regulator of M, and Ste12, an MAPK-APK and positive regulator of M, with an overall results producing emergent behavior.

To examine this hypothesis, we perturbed the Approach 1 model with two cases, first, vary the affinity between Ste7 and the Fus3 common domain, and second, examine an alternative network where Ste7 binds to a separate domain in Fus3. The results, Figure 4-9, showed that increased Ste7-Fus3 affinity produced more activated Fus3, less unbound Fus3 common domain, and reduced Msg5-Fus3 complexes, all of which are consistent with Ste7 out-competes Msg5 to bind to Fus3 common domain, resulting to reduced Fus3 dephosphorylation. We also observed that activated Ste12 and M response have bifurcating behavior at the lower Ste7-Fus3 affinity, consistent with the interpretation that the effect of Ste7-Fus3 affinity to Ste12 activation is both positive that more Fus3 is activated, and negative that Ste12 is also out-competed by Ste7 to bind to Fus3. These emergent behaviors disappeared in a network with separate domain in Fus3 to bind Ste7 (Figures 4-9f-j).

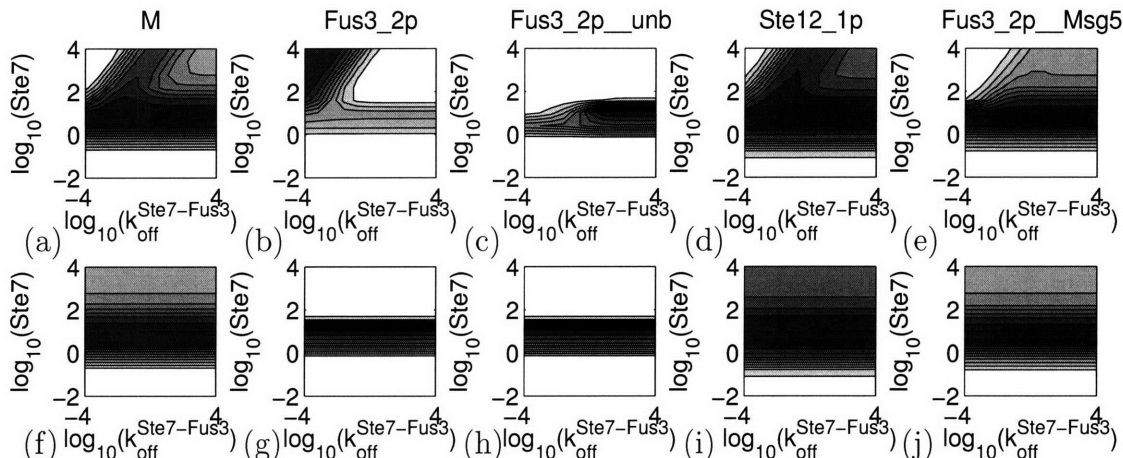


Figure 4-9: Perturbation studies to examine the emergent behavior of the Fus3 common domain. (a) Mating response for System IV; (b) activated Fus3; (c) unbound Fus3 common domain; (d) activated Ste12; (e) a complex of Msg5 and Fus3; (f-j) similar to (a-e) for a network with separate binding domain between Ste7 and Fus3.

4.4 Discussion

We proposed protein Domain Network (DN) representation to model signal transduction networks compactly. When compared to the standard Mass-Action (MA) kinetics, the DN representation utilized an alternative network definition, with protein domains in each modification state as nodes, and binding and modification fluxes as arcs. This definition produced compact dynamic models that may be applied to large signal transduction networks where MA approach may be infeasible because it requires too large number of species. During the preparation of this manuscript, several groups have also proposed a protein domain representation to reduce complexity [96, 49]. Compared to [96], we formulated a generalized representation of the domain network and its extension to networks with cooperative binding while Borisov and groups [96] only considered independent bindings and their representation requires classification of proteins. Compared to [49], our approach considered the presence of modification fluxes in these networks and alternative approaches to compute them, while Conzelmann and groups only considered binding fluxes in the networks. In essence, their results confirmed that the binding fluxes in our DN models

reduced to functions of DN variables only, and they provided an alternative systematic method to compute these fluxes.

To develop DN approach, we looked into its relationship to MA models in test systems. In the framework of model reduction methods (reviewed in [91]), the DN approach can be categorized into lumping approach, where MA species are lumped into DN species with a linear mapping guided by biological insights. The lumping of species may produce exact reduced model when, in addition to the domain lumping, we mapped the modification species one-to-one from MA to DN models. Because we require that DN method to avoid construction of any MA models, we explored alternative approaches to compute the modification species only as a function of DN variables and discovered an approach that produce accurate and compact models.

When compare to other model reduction methods, the DN method produced superior performance. Three alternative methods are available to reduce a reaction system (reviewed in [91]): lumping, sensitivity analysis, and time-scale analysis. First, automatic methods to lump a non-linear ODE system [45, 46] search for the mapping function M from an MA to a reduced model by maximizing accuracy. However, these approaches require construction of MA models, and is currently limited for a linear inverse mapping from reduced to MA models. Furthermore, the resulting M mapping does not possess a structure that can be extended to larger networks. Second, sensitivity analysis [91, 124, 111, 72] chooses a subset of MA species and MA reaction fluxes that may be eliminated based on their lack of contribution to measured outputs. Here, we found that reduced models with less than 144 (out of the original 194) species produce worst case accuracy much more frequently than the proposed DN model containing only 30 species. Third, a Krylov-based model reduction method [42] may compute the mapping M systematically to minimize deviation to data, however, this approach is currently limited to linear systems only, and when applied to the bilinear MA models, it produced a reduced model that lose its sparsity structure, and

did not obtain overall reduction in the computational time (data not shown). Finally, all the fore-mentioned model reduction methods require a construction of MA models at least initially, which may be intractable for a large network.

The DN method has several key limitations. First, the method's number of nodes scales quadratically to the number of modification states, which may scale combinatorially to the number of modification sites. Recent experiments studied that the number of modification sites in EGFRs to be around 20 [133], which translates to 2^{20} (10^6) states, however, it is not clear whether all the combination are being utilized biologically. The current knowledge in ERK system suggests that the phosphorylation are sequential [125] If all combinations of modification sites are found to be important, a similar change of representation may be utilized to produce compact models. Second, the DN method is not appropriate to represent stochastic dynamic system. The DN reverse mapping functions rely on average concentration and do not contain distribution information. The importance of stochastic dynamic systems may depend on specific systems, for example, the stochasticity of reactions are found to be unimportant in yeast mating system [5], but are crucial to model bacteriophage lambda behavior [1].

The DN method is applied to study signaling specificity in yeast Mating (M) and Filamentous Growth (FG) pathways. We discovered two interesting behaviors of the DN models: first, that the Ste5 scaffold protein that is thought to be determinant of Mating response also contribute to significant amount of FG response, and thus, Ste5 does not promote specificity when Ste5 is a passive tethering protein with independent binding, however, when Ste5 also promote cooperative binding to its kinase partners, then the signaling specificity is enhanced; second, that the Fus3 common domain that is shared between Ste7 (an MEK), Msg5 (a phosphatase) and Ste12 (an MAPK-APK) contributes to an emergent behavior of high M response at both canonical and high Ste7 initial conditions. Interestingly, a recent study on Ste5 [106] suggested

that allosteric binding is observed in Ste5, and that the actual picture of the network architecture is more complex than what we considered here. This result demonstrated how the DN model produced behaviors consistent to our expected understanding of the system, and may be applied to study other large signal transduction networks.

$$\begin{aligned}
[\dot{E}] &= -w_1 + w_2 - w_3 + w_4 - w_5 + w_6 & (4.11) \\
[\dot{K}] &= -w_7 + w_8 - w_9 + w_{10} \\
[\dot{K}_p] &= -w_{11} + w_{12} - w_{13} + w_{14} \\
[S(\dot{0}, 0)] &= -w_1 + w_2 - w_7 + w_8 - w_{11} + w_{12} \\
[S(\dot{1}, 0)] &= +w_1 - w_2 - w_9 + w_{10} - w_{13} + w_{14} \\
[S(\dot{0}, 1)] &= -w_3 + w_4 + w_7 - w_8 \\
[S(\dot{0}, 2)] &= -w_5 + w_6 + w_{11} - w_{12} \\
[S(\dot{1}, 1)] &= +w_3 - w_4 + w_9 - w_{10} - w_{15} \\
[S(\dot{1}, 2)] &= +w_5 - w_6 + w_{13} - w_{14} + w_{15} \\
\\
w_1 &= k_{\text{on}}^{SE} [S(0,0)][E] \\
w_2 &= k_{\text{off}}^{SE} [S(1,0)] \\
w_3 &= k_{\text{on}}^{SE} [S(0,1)][E] \\
w_4 &= k_{\text{off}}^{SE} [S(1,1)] \\
w_5 &= k_{\text{on}}^{SE} [S(0,2)][E] \\
w_6 &= k_{\text{off}}^{SE} [S(1,2)] \\
w_7 &= k_{\text{on}}^{SK} [S(0,0)][K] \\
w_8 &= k_{\text{off}}^{SK} [S(0,1)] \\
w_9 &= k_{\text{on}}^{SK} [S(1,0)][K] \\
w_{10} &= k_{\text{off}}^{SK} [S(1,1)] \\
w_{11} &= k_{SK_p}^{\text{on}} [S(0,0)][K_p] \\
w_{12} &= k_{SK_p}^{\text{off}} [S(0,2)] \\
w_{13} &= k_{SK_p}^{\text{on}} [S(1,0)][K_p] \\
w_{14} &= k_{SK_p}^{\text{off}} [S(1,2)] \\
w_{15} &= k_{\text{mod}}^{SEK} [S(1,1)] \\
\\
T_{S1}^{\text{tot}} &= 0 & (4.12) \\
D_{[E11]}^{\text{unb}} = D_{[S11]}^{\text{unb}} &= -v_1 + v_2 \\
D_{[K11]}^{\text{unb}} &= -v_3 + v_4 \\
D_{[K21]}^{\text{unb}} &= -v_5 + v_6 \\
D_{[S21]}^{\text{unb}} &= -v_3 + v_4 - v_5 + v_6 \\
D_{[S11,E11]}^{\text{bnd}} &= +v_1 - v_2 \\
D_{[S21,K11]}^{\text{bnd}} &= +v_3 - v_4 - v_7 \\
D_{[S21,K21]}^{\text{bnd}} &= +v_5 - v_6 + v_7 \\
[S(1,1)] &= g_{\tilde{m}}(D, T) \\
\\
v_1 &= k_{\text{on}}^{SE} D_{[E11]}^{\text{unb}} D_{[S11]}^{\text{unb}} \\
v_2 &= k_{\text{off}}^{SE} D_{[S11,E11]}^{\text{bnd}} \\
v_3 &= k_{\text{on}}^{SK} D_{[K11]}^{\text{unb}} D_{[S21]}^{\text{unb}} \\
v_4 &= k_{\text{off}}^{SK} D_{[S21,K11]}^{\text{bnd}} \\
v_5 &= k_{SK_p}^{\text{on}} D_{[K21]}^{\text{unb}} D_{[S21]}^{\text{unb}} \\
v_6 &= k_{SK_p}^{\text{off}} D_{[S21,K21]}^{\text{bnd}} \\
v_7 &= k_{\text{mod}}^{SEK} [S(1,1)]
\end{aligned}$$

$$\begin{aligned}
T_{S1}^{\text{tot}} &= [S(0,0)] + [S(0,1)] + [S(0,2)] \\
&\quad + [S(1,0)] + [S(1,1)] + [S(1,2)] & (4.13) \\
D_{[E11]}^{\text{unb}} &= [E] \\
D_{[K11]}^{\text{unb}} &= [K] \\
D_{[K21]}^{\text{unb}} &= [K_p] \\
D_{[S11]}^{\text{unb}} &= [S(0,0)] + [S(0,1)] + [S(0,2)] \\
D_{[S11,E11]}^{\text{bnd}} &= [S(1,0)] + [S(1,1)] + [S(1,2)] \\
D_{[S21]}^{\text{unb}} &= [S(0,0)] + [S(1,0)] \\
D_{[S21,K11]}^{\text{bnd}} &= [S(0,1)] + [S(1,1)] \\
D_{[S21,K21]}^{\text{bnd}} &= [S(0,2)] + [S(1,2)] \\
\\
\text{Approach 1:} \\
[S(\dot{1}, 1)] &= +v_1 \frac{D_{[S21]}^{\text{unb}}}{T_{S1}^{\text{tot}}} + v_3 \frac{D_{[S11]}^{\text{unb}}}{T_{S1}^{\text{tot}}} \\
&\quad - (k_{\text{off}}^{SE} + k_{\text{off}}^{SK} + k_{\text{mod}}^{SEK}) [S(1,1)] & (4.14) \\
\\
\text{Approach 2:} \\
\frac{[S(1,1)]}{T_{S1}^{\text{tot}}} &= \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} & (4.15) \\
\\
\text{Approach 3:} \\
D_{[S11,E11]}^{\text{bnd}} = 0 &= v_1 - v_2 & (4.16) \\
D_{[S21,K11]}^{\text{bnd}} = 0 &= v_3 - v_4 \\
D_{[S21,K21]}^{\text{bnd}} = 0 &= v_5 - v_6 \\
D_{[E11]}^{\text{unb}} = D_{[S11]}^{\text{unb}} = 0 &= -v_1 + v_2 \\
D_{[K11]}^{\text{unb}} = 0 &= -v_3 + v_4 \\
D_{[K21]}^{\text{unb}} = 0 &= -v_5 + v_6 \\
D_{[S21]}^{\text{unb}} = 0 &= -v_3 + v_4 - v_5 + v_6 \\
\frac{[S(1,1)]}{T_{S1}^{\text{tot}}} &= \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \\
\\
\text{Approach 4:} \\
[S(\dot{1}, 0)] &= 0 = +w_1 - w_2 - w_9 + w_{10} - w_{13} + w_{14} & (4.17) \\
[S(\dot{0}, 1)] &= 0 = -w_3 + w_4 + w_7 - w_8 \\
[S(\dot{0}, 2)] &= 0 = -w_5 + w_6 + w_{11} - w_{12} \\
[S(\dot{1}, 1)] &= 0 = +w_3 - w_4 + w_9 - w_{10} - w_{15} \\
[S(\dot{1}, 2)] &= 0 = +w_5 - w_6 + w_{13} - w_{14} + w_{15}
\end{aligned}$$

Table 4.2: Equations applied to System I. Equation 4.11, Mass Action (MA) kinetics models; Equation 4.12, Domain Network (DN) models; Equation 4.13, definition of the DN lumping; Equation 4.14–4.17, four approaches to implement $g_{\tilde{m}}$ within the DN approach.

Table 4.3: Experimental phenotypic behaviors being used to parameterize models of yeast mating and filamentous growth pathways.

No	Experimental results	Ref.
0	Wt produced FG response at low pheromone and M response at high pheromone	[109]
1	Ste4 and Ste5 has major contribution to the activation of Kss1 and FG.	[59]
2	Δ Fus3 greatly elevated levels of active Kss1.	[59],[52]
3	Kss1 dependence on Ste5 is bypassed by increasing levels of Ste11 and Ste7.	[59]
4	To be activated, Kss1 does not need to bind to Ste5, but Fus3 must bind to Ste5.	[59]
5	Over-expression of Ste11 inhibits Fus3 unless Ste5 is over-expressed.	[59]
6	Over-expression of Ste11 and Ste7 in Δ Ste5 increases Kss1 activation (to the base case).	[59]
7	Msg5 phosphatase inhibits constitutive activation of Fus3. Δ Msg5 constitutively activated Mating based on increased Fus3pp. FG was inhibited.	[59]
8	Kss1 inhibits Mating by interacting with Ste12.	[52]
9	Kinase defective Fus3-K42R suppress FG response of Δ Fus3.	[52]
10	Kss1 also has inhibitory function to FG that does not require Ste7, and arises from binding of inactive Kss1 to Tec1.	[52]
11	Δ Fus3 can still activate mating via Kss1 that is 0.1 of wt. Δ Fus3,Fus3-K42R reduces down the mating frequency to 0.0001 from wt.	[52]

Table 4.4: Nonlinear constraint formulation to represent the phenotypic behaviors listed in Table 4.3. M, mating response; FG, filamentous growth response; over, overexpression; Δ , deletion mutation; Fus3-K42R, mutational to Fus3 with reduced activity; Kss1^{nobind}, mutation to non-binding Kss1.

Cases	Facts	Variables	Constraint expressions
Low/high alpha factor	0	M, FG	high.M \geq low.M, high.FG \leq low.FG
Ste5 Δ	1	Kss1 _{pp}	Ste5 Δ .Kss1 _{pp} \leq 0.5*baseline.Kss1 _{pp}
	4	Fus3 _{pp}	Ste5 Δ .Fus3 _{pp} \leq 10 ⁻⁶
Fus3 Δ	2	Kss1 _{pp}	Fus3 Δ .Kss1 _{pp} \geq baseline.Kss1 _{pp}
	9	FG	Fus3-K42R.FG \leq Fus3 Δ .FG
	11	M	Fus3 Δ .M \leq 0.1 * baseline.M
Fus3-K42R	9	FG	Fus3-K42R.FG \leq Fus3 Δ .FG
	11	M	Fus3-K42R.M \leq 10 ⁻⁴ * baseline.M
Ste11 ^{over} , Ste7 ^{over} , Ste5 Δ	3	Kss1 _{pp}	Ste11Ste7 ^{over} Ste5 Δ .Kss1 _{pp} \geq Ste5 Δ .Kss1 _{pp}
Ste11 ^{over}	5	Fus3 _{pp}	Ste11 ^{over} .Fus3 _{pp} \leq baseline.Fus3 _{pp}
Ste11 ^{over} , Ste5 ^{over}	5	Fus3 _{pp}	Ste11Ste5 ^{over} .Fus3 _{pp} \geq Ste11 ^{over} .Fus3 _{pp}
Msg5 Δ	7	Fus3 _{pp} , M	Msg5 Δ .(Fus3 _{pp} ,M) \geq baseline.(Fus3 _{pp} ,M)
Kss1 ^{nobind}	8	M	Kss1 ^{nobind} .M \geq baseline.M
Kss1 ^{over}	10	FG	Kss1 ^{over} .FG \leq baseline.FG

4.5 Supplementary Materials

4.5.1 Models for System I

Approach 0: A Mass-Action (MA) kinetic model

The Mass-Action (MA) kinetic of System I (Figure 4-1a) is given as follows:

$$\begin{aligned}
 [\dot{E}] &= -w_1 + w_2 - w_3 + w_4 - w_5 + w_6 & (4.18) \\
 [\dot{K}] &= -w_7 + w_8 - w_9 + w_{10} \\
 [\dot{K}_p] &= -w_{11} + w_{12} - w_{13} + w_{14} \\
 [S(\dot{0}, 0)] &= -w_1 + w_2 - w_7 + w_8 - w_{11} + w_{12} \\
 [S(\dot{1}, 0)] &= +w_1 - w_2 - w_9 + w_{10} - w_{13} + w_{14} \\
 [S(\dot{0}, 1)] &= -w_3 + w_4 + w_7 - w_8 \\
 [S(\dot{0}, 2)] &= -w_5 + w_6 + w_{11} - w_{12} \\
 [S(\dot{1}, 1)] &= +w_3 - w_4 + w_9 - w_{10} - w_{15} & (4.19) \\
 [S(\dot{1}, 2)] &= +w_5 - w_6 + w_{13} - w_{14} + w_{15}
 \end{aligned}$$

$$\begin{aligned}
 w_1 &= k_{\text{on}}^{SE} [S(0, 0)][E] \\
 w_2 &= k_{\text{off}}^{SE} [S(1, 0)] \\
 w_3 &= k_{\text{on}}^{SE} [S(0, 1)][E] \\
 w_4 &= k_{\text{off}}^{SE} [S(1, 1)] \\
 w_5 &= k_{\text{on}}^{SE} [S(0, 2)][E] \\
 w_6 &= k_{\text{off}}^{SE} [S(1, 2)] \\
 w_7 &= k_{\text{on}}^{SK} [S(0, 0)][K] \\
 w_8 &= k_{\text{off}}^{SK} [S(0, 1)] \\
 w_9 &= k_{\text{on}}^{SK} [S(1, 0)][K] \\
 w_{10} &= k_{\text{off}}^{SK} [S(1, 1)] \\
 w_{11} &= k_{SK_p}^{\text{on}} [S(0, 0)][K_p] \\
 w_{12} &= k_{SK_p}^{\text{off}} [S(0, 2)] \\
 w_{13} &= k_{SK_p}^{\text{on}} [S(1, 0)][K_p] \\
 w_{14} &= k_{SK_p}^{\text{off}} [S(1, 2)] \\
 w_{15} &= k_{\text{mod}}^{SEK} [S(1, 1)]
 \end{aligned}$$

Where the w represents binding and modification fluxes, $S(i, j)$ represents various protein complexes with E, K, K_p as partners. The indices in S are defined such that $i \in \{0, 1\}$ represents unbound and bound state to E , and $j \in \{0, 1, 2\}$ represents unbound and bound states to K, K_p , respectively

Lumping to reduce the number of states

To reduce the number of states, we represent the network as a protein-domain network (DN). The DN models consist of domain variables D , and total variables T , both are constructed by lumping species in MA models using the following linear

relationship:

$$\begin{aligned}
T_{S1}^{\text{tot}} &= [S(0, 0)] + [S(0, 1)] + [S(0, 2)] + [S(1, 0)] + [S(1, 1)] + [S(1, 2)] & (4.20) \\
D_{[E11]}^{\text{unb}} &= [E] \\
D_{[K11]}^{\text{unb}} &= [K] \\
D_{[K21]}^{\text{unb}} &= [K_p] \\
D_{[S11]}^{\text{unb}} &= [S(0, 0)] + [S(0, 1)] + [S(0, 2)] \\
D_{[S11,E11]}^{\text{bnd}} &= [S(1, 0)] + [S(1, 1)] + [S(1, 2)] \\
D_{[S21]}^{\text{unb}} &= [S(0, 0)] + [S(1, 0)] \\
D_{[S21,K11]}^{\text{bnd}} &= [S(0, 1)] + [S(1, 1)] \\
D_{[S21,K21]}^{\text{bnd}} &= [S(0, 2)] + [S(1, 2)]
\end{aligned}$$

where $D_{[pij]}^{\text{unb}}$ represents the unbound state of protein p domain i modification j , $D_{[pij,d']}^{\text{bnd}}$ represents the bound state of protein p domain i modification j with protein domain modification d' , and T_{pj}^{tot} represents the total amount of protein p in modification j .

Applied to the System I MA model (Equation 4.18), this mapping produced the following DN dynamic equations:

$$\begin{aligned}
T_{S1}^{\text{tot}} &= 0 & (4.21) \\
D_{[E11]}^{\text{unb}} &= -w_1 + w_2 - w_3 + w_4 - w_5 + w_6 \\
D_{[K11]}^{\text{unb}} &= -w_7 + w_8 - w_9 + w_{10} \\
D_{[K21]}^{\text{unb}} &= -w_{11} + w_{12} - w_{13} + w_{14} \\
D_{[S11]}^{\text{unb}} &= -w_1 + w_2 - w_3 + w_4 - w_5 + w_6 \\
D_{[S11,E11]}^{\text{bnd}} &= +w_1 - w_2 + w_3 - w_4 + w_5 - w_6 \\
D_{[S21]}^{\text{unb}} &= -w_7 + w_8 - w_9 + w_{10} & -w_{11} + w_{12} - w_{13} + w_{14} \\
D_{[S21,K11]}^{\text{bnd}} &= +w_7 - w_8 + w_9 - w_{10} & -w_{15} \\
D_{[S21,K21]}^{\text{bnd}} &= +w_{11} - w_{12} + w_{13} - w_{14} + w_{15}
\end{aligned}$$

We noticed that the binding fluxes $w_1 - w_{14}$ can be lumped to be a function of D only, for example:

$$\begin{aligned}
D_{[S11,E11]}^{\text{bnd}} &= w_1 - w_2 + w_3 - w_4 + w_5 - w_6 & (4.22) \\
&= k_{\text{on}}^{SE} [S(0, 0)] [E] - k_{\text{off}}^{SE} [S(1, 0)] + k_{\text{on}}^{SE} [S(0, 1)] [E] - k_{\text{off}}^{SE} [S(1, 1)] + k_{\text{on}}^{SE} [S(0, 2)] [E] - k_{\text{off}}^{SE} [S(1, 2)] \\
&= k_{\text{on}}^{SE} [E] ([S(0, 0)] + [S(0, 1)] + [S(0, 2)]) - k_{\text{off}}^{SE} ([S(1, 0)] + [S(1, 1)] + [S(1, 2)]) \\
&= k_{\text{on}}^{SE} D_{[E11]}^{\text{unb}} D_{[S11]}^{\text{unb}} - k_{\text{off}}^{SE} (D_{[S11,E11]}^{\text{bnd}}) \\
&= v_1 - v_2 \\
D_{[S21,K11]}^{\text{bnd}} &= w_7 - w_8 + w_9 - w_{10} - w_{15} \\
&= k_{\text{on}}^{SK} [S(1, 0)] [K] - k_{\text{off}}^{SK} [S(1, 1)] + k_{\text{on}}^{SK} [S(0, 0)] [K] - k_{\text{off}}^{SK} [S(0, 1)] - k_{\text{mod}}^{SEK} [S(1, 1)] \\
&= k_{\text{on}}^{SK} [K] ([S(1, 0)] + S_-) - k_{\text{off}}^{SK} ([S(1, 1)] + [S(0, 1)]) - k_{\text{mod}}^{SEK} [S(1, 1)] \\
&= k_{\text{on}}^{SK} D_{[K11]}^{\text{unb}} D_{[S21]}^{\text{unb}} - k_{\text{off}}^{SK} D_{[S21,K11]}^{\text{bnd}} - k_{\text{mod}}^{SEK} [S(1, 1)] \\
&= v_3 - v_4 - v_7 \\
D_{[S11]}^{\text{unb}} = D_{[E11]}^{\text{unb}} &= -v_1 + v_2 \\
D_{[S21]}^{\text{unb}} &= -v_3 + v_4 - v_5 + v_6 \\
D_{[K11]}^{\text{unb}} &= -v_3 + v_4 \\
D_{[K21]}^{\text{unb}} &= -v_5 + v_6 \\
D_{[S21,K21]}^{\text{bnd}} &= v_5 - v_6 + v_7
\end{aligned}$$

where:

$$\begin{aligned}
v_1 &= k_{\text{on}}^{SE} D_{[E11]}^{\text{unb}} D_{[S11]}^{\text{unb}} \\
v_2 &= k_{\text{off}}^{SE} D_{[S11,E11]}^{\text{bnd}} \\
v_3 &= k_{\text{on}}^{SK} D_{[K11]}^{\text{unb}} D_{[S21]}^{\text{unb}} \\
v_4 &= k_{\text{off}}^{SK} D_{[S21,K11]}^{\text{bnd}} \\
v_5 &= k_{SK_p}^{\text{on}} D_{[K21]}^{\text{unb}} D_{[S21]}^{\text{unb}} \\
v_6 &= k_{SK_p}^{\text{off}} D_{[S21,K21]}^{\text{bnd}} \\
v_7 &= k_{\text{mod}}^{SEK} [S(1,1)]
\end{aligned} \tag{4.23}$$

The structure of the MA model allows the binding fluxes ($w_1 - w_{14}$) to be computed as a function of the DN variables, however, the modification flux w_{15} cannot be reduced to be a function of DN variables, and left as a function of the MA species $[S(1,1)]$. We will estimate $[S(1,1)]$ from DN variables using alternative functions described later.

Similarly, the dynamic of the total modification states are obtained by lumping, to be a function of modification fluxes only:

$$\begin{aligned}
T_{S1}^{\text{tot}} &= 0 \\
T_{E1}^{\text{tot}} &= 0 \\
T_{K1}^{\text{tot}} &= -v_7 \\
T_{K2}^{\text{tot}} &= +v_7
\end{aligned} \tag{4.24}$$

Thus, the overall DN model representation of System I is given by:

$$\begin{aligned}
D_{[E11]}^{\text{unb}} = D_{[S11]}^{\text{unb}} &= -v_1 + v_2 \\
D_{[K11]}^{\text{unb}} &= -v_3 + v_4 \\
D_{[K21]}^{\text{unb}} &= -v_5 + v_6 \\
D_{[S21]}^{\text{unb}} &= -v_3 + v_4 \quad -v_5 + v_6 \\
D_{[S11,E11]}^{\text{bnd}} &= +v_1 - v_2 \\
D_{[S21,K11]}^{\text{bnd}} &= +v_3 - v_4 \quad -v_7 \\
D_{[S21,K21]}^{\text{bnd}} &= +v_5 - v_6 + v_7 \\
T_{S1}^{\text{tot}} &= 0 \\
[S(1,1)] &= g_{\hat{m}}(D, T)
\end{aligned} \tag{4.25}$$

where v are defined in Equation 4.23, and alternative $g_{\hat{m}}$ functions to be explored next.

Alternative functions to compute modification species

Approach 3: Binding equilibrium approximation. One way to compute modification species is to assume binding equilibrium, and neglect modification fluxes. In System I, this approach can be expressed as:

$$\begin{aligned}
[S(i, 0)] &= 0 = +w_1 - w_2 && -w_9 + w_{10} && -w_{13} + w_{14} \\
[S(o, 1)] &= 0 = && -w_3 + w_4 && +w_7 - w_8 \\
[S(o, 2)] &= 0 = && -w_5 + w_6 && +w_{11} - w_{12} \\
[S(i, 1)] &= 0 = +w_3 - w_4 && +w_9 - w_{10} \\
[S(i, 2)] &= 0 = && +w_5 - w_6 && +w_{13} - w_{14}
\end{aligned} \tag{4.27}$$

which simplifies to:

$$\begin{aligned}
w_1 &= w_2 \\
w_3 &= w_4 \\
w_5 &= w_6 \\
w_7 &= w_8 \\
w_9 &= w_{10} \\
w_{11} &= w_{12} \\
w_{13} &= w_{14}
\end{aligned} \tag{4.28}$$

These equations can be manipulated to compute the unbound and bound domains:

$$\begin{aligned}
D_{[S11,E11]}^{\text{bnd}} &= [S(1,0)] + [S(1,1)] + [S(1,2)] = \frac{k_{\text{on}}^{SE}}{k_{\text{off}}^{SE}} D_{[E11]}^{\text{unb}} ([S(0,0)] + [S(0,1)] + [S(0,2)]) = \frac{k_{\text{on}}^{SE}}{k_{\text{off}}^{SE}} D_{[E11]}^{\text{unb}} D_{[S11]}^{\text{unb}} \tag{4.29} \\
D_{[S21,K11]}^{\text{bnd}} &= [S(0,1)] + [S(1,1)] = \frac{k_{\text{on}}^{SK}}{k_{\text{off}}^{SK}} D_{[K11]}^{\text{unb}} ([S(0,0)] + [S(1,0)]) = \frac{k_{\text{on}}^{SK}}{k_{\text{off}}^{SK}} D_{[K11]}^{\text{unb}} D_{[S21]}^{\text{unb}} \\
D_{[S21,K21]}^{\text{bnd}} &= [S(0,2)] + [S(1,2)] = \frac{k_{\text{on}}^{SK}}{k_{\text{off}}^{SK}} D_{[K21]}^{\text{unb}} ([S(0,0)] + [S(1,0)]) = \frac{k_{\text{on}}^{SK}}{k_{\text{off}}^{SK}} D_{[K21]}^{\text{unb}} D_{[S21]}^{\text{unb}} \\
T_{S1}^{\text{tot}} &= D_{[S11]}^{\text{unb}} + D_{[S11,E11]}^{\text{bnd}} \\
T_{S1}^{\text{tot}} &= D_{[S21]}^{\text{unb}} + D_{[S21,K11]}^{\text{bnd}} + D_{[S21,K21]}^{\text{bnd}}
\end{aligned}$$

Using the approximation given in Eqn 4.27, we can express the modification specie $[S(1,1)]$ from $D_{[S11,E11]}^{\text{bnd}}$ and $D_{[S21,K11]}^{\text{bnd}}$ by solving the following equations:

$$\begin{aligned}
\frac{w_1}{w_2} = \frac{w_3}{w_4} \text{ or } \frac{w_7}{w_8} = \frac{w_9}{w_{10}} & \text{ or } [S(0,0)][S(1,1)] = [S(1,0)][S(0,1)] \tag{4.30} \\
\frac{w_1}{w_2} = \frac{w_5}{w_6} \text{ or } \frac{w_9}{w_{10}} = \frac{w_{13}}{w_{14}} & \text{ or } [S(0,0)][S(1,2)] = [S(1,0)][S(0,2)] \\
D_{[S11,E11]}^{\text{bnd}} &= [S(1,0)] + [S(1,1)] + [S(1,2)] \\
D_{[S21,K11]}^{\text{bnd}} &= [S(0,1)] + [S(1,1)] \\
D_{[S21,K21]}^{\text{bnd}} &= [S(0,2)] + [S(1,2)] \\
T_{S1}^{\text{tot}} &= [S(0,0)] + [S(1,0)] + [S(0,1)] + [S(0,2)] + [S(1,1)] + [S(1,2)]
\end{aligned}$$

These six equations above contain 10 variables, and if we specify the $D_{[S11,E11]}^{\text{bnd}}$, $D_{[S21,K11]}^{\text{bnd}}$, $D_{[S21,K21]}^{\text{bnd}}$, and T_{S1}^{tot} , we have a fully determined set of equations which can be solved for $[S(1,1)]$ to produce a molar fractional products of bound domains to the scaffold protein concentration:

$$\frac{[S(1,1)]}{T_{S1}^{\text{tot}}} = \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \tag{4.31}$$

We notice that this equation is derived from a weaker approximation than made by the binding equilibrium, that it only needs the ratio of the fluxes to be equal, and that only one of the two ratios of the fluxes being needed.

In summary, this approach solve the following system of equations:

$$\begin{aligned}
T_{S1}^{\text{tot}} &= & 0 & \tag{4.32} \\
D_{[S11,E11]}^{\text{bnd}} &= 0 = & v_1 - v_2 & \\
D_{[S21,K11]}^{\text{bnd}} &= 0 = & v_3 - v_4 & \\
D_{[S21,K21]}^{\text{bnd}} &= 0 = & v_5 - v_6 & \\
D_{[E11]}^{\text{unb}} = D_{[S11]}^{\text{unb}} &= 0 = & -v_1 + v_2 & \\
D_{[K11]}^{\text{unb}} &= 0 = & -v_3 + v_4 & \\
D_{[K21]}^{\text{unb}} &= 0 = & -v_5 + v_6 & \\
D_{[S21]}^{\text{unb}} &= 0 = & -v_3 + v_4 - v_5 + v_6 & \\
\frac{[S(1,1)]}{T_{S1}^{\text{tot}}} &= & \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} & \tag{4.33}
\end{aligned}$$

where the definition of fluxes v are given in Equation 4.23. Compared to Equation 4.25, this approach assumes steady-state and neglects modification fluxes to compute D .

Approach 1: Differential modification function based on molar fractional association fluxes.

The modification specie $[S(1,1)]$ is computed by assuming molar fractional association rate. This approach solves Equation 4.25, with a specific realization of $g_{\hat{m}}$ in Equation 4.26 to be a differential expression given by:

$$\begin{aligned}
[S(\dot{1},1)] &= & +w_3 + w_9 - w_4 - w_{10} - w_{15} & \tag{4.34} \\
[S(\dot{1},1)] &= & +k_{\text{on}}^{SE} [S(0,1)][E] + k_{\text{on}}^{SK} [S(1,0)][K] - (k_{\text{off}}^{SE} + k_{\text{off}}^{SK} + k_{\text{mod}}^{SEK})[S(1,1)] & \\
[S(\dot{1},1)] &= & +v_1 \frac{[S(0,1)]}{D_{S11}^{\text{unb}}} + v_3 \frac{[S(1,0)]}{D_{S21}^{\text{unb}}} - (k_{\text{off}}^{SE} + k_{\text{off}}^{SK} + k_{\text{mod}}^{SEK})[S(1,1)] & \\
[S(\dot{1},1)] &= & +v_1 \frac{\left(\frac{D_{S11}^{\text{unb}} D_{S21}^{\text{bnd}}}{T_{S1}^{\text{tot}}}\right)}{D_{S11}^{\text{unb}}} + v_3 \frac{\left(\frac{D_{S21}^{\text{unb}} D_{S11,E1}^{\text{bnd}}}{T_{S1}^{\text{tot}}}\right)}{D_{S21}^{\text{unb}}} - (k_{\text{off}}^{SE} + k_{\text{off}}^{SK} + k_{\text{mod}}^{SEK})[S(1,1)] & \\
[S(\dot{1},1)] &= & +v_1 \frac{D_{S21}^{\text{unb}}}{T_{S1}^{\text{tot}}} + v_3 \frac{D_{S11}^{\text{unb}}}{T_{S1}^{\text{tot}}} - (k_{\text{off}}^{SE} + k_{\text{off}}^{SK} + k_{\text{mod}}^{SEK})[S(1,1)] &
\end{aligned}$$

Approach 2: Algebraic modification function based on molar fractional expression The modification specie $[S(1,1)]$ is computed by assuming molar fractional algebraic expression derived in Equation 4.31. This approach solves Equation 4.25, with a specific realization of $g_{\hat{m}}$ in Equation 4.26 to be an algebraic expression given by:

$$\frac{[S(1,1)]}{T_{S1}^{\text{tot}}} = \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \tag{4.35}$$

Approach 4: Pseudo steady-state approximation. The pseudo steady-state approximation(PSSA) is derived by taking a steady-state approximation of these intermediate states:

$$\begin{aligned}
[S(\dot{1},0)] = 0 &= & +w_1 - w_2 & & -w_9 + w_{10} & & -w_{13} + w_{14} & \tag{4.36} \\
[S(\dot{0},1)] = 0 &= & -w_3 + w_4 & & +w_7 - w_8 & & & \\
[S(\dot{0},2)] = 0 &= & & & -w_5 + w_6 & & +w_{11} - w_{12} & \\
[S(\dot{1},1)] = 0 &= & +w_3 - w_4 & & & & +w_9 - w_{10} & & -w_{15} \\
[S(\dot{1},2)] = 0 &= & & & +w_5 & & -w_6 & & +w_{13} & & -w_{14} + w_{15}
\end{aligned}$$

These equations may be lumped into the DN variable D with a similar pseudo steady-

state approximation:

$$\begin{aligned}
D_{[S11,E11]}^{\text{bnd}} = 0 &= & v_1 - v_2 & (4.37) \\
D_{[S21,K11]}^{\text{bnd}} = 0 &= & v_3 - v_4 - v_7 \\
D_{[S21,K21]}^{\text{bnd}} = 0 &= & v_5 - v_6 + v_7
\end{aligned}$$

The difference between these equations and the exact dynamics of the combined system (Eqn 4.25) is that we assume that the domain variables D are time invariant.

In summary, this approach solves the following system of equations:

$$\begin{aligned}
T_{S1}^{\text{tot}} &= & 0 & (4.38) \\
T_{E1}^{\text{tot}} &= & 0 \\
T_{K1}^{\text{tot}} &= & -v_7 \\
T_{K2}^{\text{tot}} &= & +v_7 \\
D_{[S11,E11]}^{\text{bnd}} = 0 &= & v_1 - v_2 \\
D_{[S21,K11]}^{\text{bnd}} = 0 &= & v_3 - v_4 - v_7 \\
D_{[S21,K21]}^{\text{bnd}} = 0 &= & v_5 - v_6 + v_7 \\
D_{[S11]}^{\text{unb}} + D_{[S11,E11]}^{\text{bnd}} &= & T_{S1}^{\text{tot}} \\
D_{[S21]}^{\text{unb}} + D_{[S21,K11]}^{\text{bnd}} + D_{[S21,K21]}^{\text{bnd}} &= & T_{S1}^{\text{tot}} \\
D_{[E11]}^{\text{unb}} + D_{[S11,E11]}^{\text{bnd}} &= & T_{E1}^{\text{tot}} \\
D_{[K11]}^{\text{unb}} + D_{[S21,K11]}^{\text{bnd}} &= & T_{K1}^{\text{tot}} \\
D_{[K21]}^{\text{unb}} + D_{[S21,K21]}^{\text{bnd}} &= & T_{K2}^{\text{tot}} \\
\frac{[S(1,1)]}{T_{S1}^{\text{tot}}} &= & \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}}
\end{aligned}$$

where v are defined in Eqn 4.23.

4.5.2 Models for System II

To extend the approaches to model networks with cooperative binding, we examined the following modification to the four approaches applied to System I. Applied to System II, Approach 1 generates the following system of equations:

$$\begin{aligned}
D_{[E11]}^{\text{unb}} = D_{[S11]}^{\text{unb}} &= & -v_1 + v_2 & (4.39) \\
D_{[S11,E11]}^{\text{bnd}} &= & +v_1 - v_2 & -(\beta - 1)k_{\text{off}}^{SE}[S(1,1)] \\
D_{[K11]}^{\text{unb}} &= & -v_3 + v_4 \\
D_{[K21]}^{\text{unb}} &= & & -v_5 + v_6 \\
D_{[S21]}^{\text{unb}} &= & -v_3 + v_4 & -v_5 + v_6 \\
D_{[S21,K11]}^{\text{bnd}} &= & +v_3 - v_4 & -v_7 \quad -(\beta - 1)k_{\text{off}}^{SK}[S(1,1)] \\
D_{[S21,K21]}^{\text{bnd}} &= & & +v_5 - v_6 + v_7 \\
T_{S1}^{\text{tot}} &= & 0 \\
T_{E1}^{\text{tot}} &= & 0 \\
T_{K1}^{\text{tot}} &= & -v_7 \\
T_{K2}^{\text{tot}} &= & +v_7 \\
[S(1,1)] &= & +v_1 \frac{D_{[S21,K11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} + v_3 \frac{D_{[S11,E11]}^{\text{bnd}}}{T_{S1}^{\text{tot}}} & -\beta(k_{\text{off}}^{SE} + k_{\text{off}}^{SK})[S(1,1)] - v_7
\end{aligned}$$

Approach 2 and 3 requires a new derivation of the algebraic expression to represent cooperative binding. This can be from the following system of equations:

$$\begin{aligned}
\frac{w_1}{w_2} &= \frac{w_3}{\beta w_4} \text{ or } \frac{w_7}{w_8} = \frac{w_9}{\beta w_{10}} \text{ or} & \beta[S(0,0)][S(1,1)] &= [S(1,0)][S(0,1)] & (4.40) \\
\frac{w_1}{w_2} &= \frac{w_5}{w_6} \text{ or } \frac{w_9}{w_{10}} = \frac{w_{13}}{w_{14}} \text{ or} & [S(0,0)][S(1,2)] &= [S(1,0)][S(0,2)] \\
D_{[S_{11},E_{11}]}^{\text{bnd}} &= & [S(1,0)] &+ [S(1,1)] &+ [S(1,2)] \\
D_{[S_{21},K_{11}]}^{\text{bnd}} &= & [S(0,1)] &+ [S(1,1)] \\
D_{[S_{21},K_{21}]}^{\text{bnd}} &= & [S(0,2)] &+ [S(1,2)] \\
T_{S_1}^{\text{tot}} &= & [S(0,0)] &+ [S(1,0)] &+ [S(0,1)] &+ [S(0,2)] &+ [S(1,1)] &+ [S(1,2)]
\end{aligned}$$

which simplified to

$$(D_{[S_{11},E_{11}]}^{\text{bnd}} - [S(1,1)])(D_{[S_{21},K_{11}]}^{\text{bnd}} - [S(1,1)]) = \beta[S(1,1)](T_{S_1}^{\text{tot}} - D_{[S_{11},E_{11}]}^{\text{bnd}} - D_{[S_{21},K_{11}]}^{\text{bnd}} + [S(1,1)]) \quad (4.41)$$

Thus, the system of equations for Approach 2 applied to System II is similar to Equation 4.39, with the $g_{\hat{m}}$ equation modified to

$$(D_{[S_{11},E_{11}]}^{\text{bnd}} - [S(1,1)])(D_{[S_{21},K_{11}]}^{\text{bnd}} - [S(1,1)]) = \beta[S(1,1)](S^{\text{tot}} - D_{[S_{11},E_{11}]}^{\text{bnd}} - D_{[S_{21},K_{11}]}^{\text{bnd}} + [S(1,1)]) \quad (4.42)$$

Approach 4 is not applied to System II because of difficulty to generalize the approach to larger networks.

4.5.3 Lumping approach with linear inverse lumping applied to System I

The DN method is a lumping approach given by $\hat{x} = Mx$, where x are MA species, \hat{x} are DN species, and M is the lumping matrix.

Applied to System I, we defined:

$$x = \begin{matrix} & & & [E] & [K] & [K_p] & [S(0,0)] & [S(1,0)] & [S(0,1)] & [S(0,2)] & [S(1,1)] & [S(1,2)] \end{matrix} \quad (4.43)$$

$$\hat{x} = \begin{pmatrix} D_{[E_{11}]}^{\text{unb}} & D_{[K_{11}]}^{\text{unb}} & D_{[K_{21}]}^{\text{unb}} & D_{[S_{11}]}^{\text{unb}} & D_{[S_{11},E_{11}]}^{\text{bnd}} & D_{[S_{21}]}^{\text{unb}} & D_{[S_{21},K_{11}]}^{\text{bnd}} & D_{[S_{21},K_{21}]}^{\text{bnd}} & T_{S_1}^{\text{tot}} & S(1,1) \end{pmatrix} \quad (4.44)$$

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad (4.45)$$

The generalized inverse can be found by definition to be $M\hat{M}M = M$, which produced:

$$\hat{M} = \begin{pmatrix} 1.00 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.100 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.27 & -0.23 & 0.43 & -0.32 & -0.07 & 0.05 & 0.50 & 0 & 0 \\ 0 & 0 & 0 & -0.18 & 0.32 & 0.30 & 0.05 & -0.20 & 0.14 & -0.50 & 0 & 0 \\ 0 & 0 & 0 & 0.09 & 0.09 & -0.27 & 0.73 & -0.27 & 0.18 & -1.00 & 0 & 0 \\ 0 & 0 & 0 & 0.27 & -0.23 & -0.07 & -0.32 & 0.43 & 0.05 & 0.50 & 0 & 0 \\ 0 & 0 & 0 & 0.00 & -0.00 & -0.00 & 0.00 & 0.00 & -0.00 & 1.00 & 0 & 0 \\ 0 & 0 & 0 & -0.18 & 0.32 & -0.20 & 0.04 & 0.30 & 0.14 & -0.5000 & 0 & 0 \end{pmatrix} \quad (4.46)$$

We were unable to extract a rule from these generalized inverse strategy that may be extended to larger networks.

4.5.4 Proof for molar fractional expression to compute protein complexes in a network of larger sizes with independent binding

We will use induction to prove that a protein complex may be computed from its participating bound domains using a molar fractional expression. The proof outlines as follows:

1. Prove for a network with a scaffold protein and two partners, each with one modification state.
2. To extend 1 to more modification states of partners ,prove for a network with a scaffold protein and two partners, one partner with an additional modification state, and utilize the proof of 1.
3. To extend 1 to more partners of a scaffold protein, prove for a network with a scaffold protein binding three partners.
4. To extend 1 to a network with scaffold proteins in more modification states, prove for a network with an added modification state to the scaffold protein.
5. To extend 1 to a network with more scaffold proteins, prove for a network with an added scaffold protein.

Problem 1. For a network consisting of a scaffold protein S with two partners, A and B, each with one modification state, the binding equilibrium produced the following equations:

$$\begin{aligned}
 K_a^{SA} &= \frac{[S(1,0)]}{[S(0,0)][A]} = \frac{[S(1,1)]}{[S(0,1)][A]} & (4.47) \\
 K_a^{SB} &= \frac{[S(0,1)]}{[S(0,0)][B]} = \frac{[S(1,1)]}{[S(1,0)][B]}
 \end{aligned}$$

Using the following lumped definition,

$$\begin{aligned}
 S(0,*) &= S(0,0) + S(0,1) & (4.48) \\
 S(1,*) &= S(1,0) + S(1,1) \\
 S(*,0) &= S(0,0) + S(1,0) \\
 S(*,1) &= S(0,1) + S(1,1) \\
 S(*,*) &= S(0,0) + S(0,1) + S(1,0) + S(1,1)
 \end{aligned}$$

one may derive that:

$$\begin{aligned}
 K_a^{SA} &= \frac{[S(1,0)]}{[S(0,0)][A]} = \frac{[S(1,1)]}{[S(0,1)][A]} = \frac{[S(1,*)]}{[S(0,*)][A]} & (4.49) \\
 K_a^{SB} &= \frac{[S(0,1)]}{[S(0,0)][B]} = \frac{[S(1,1)]}{[S(1,0)][B]} = \frac{[S(*,1)]}{[S(*,0)][B]}
 \end{aligned}$$

these two equations can be simplified to produce a molar fractional expression, given by:

$$\frac{[S(1,1)]}{[S(*,*)]} = \frac{[S(*,1)]}{[S(*,*)]} \frac{[S(1,*)]}{[S(*,*)]} \quad (4.50)$$

Problem 2. For a network consisting of a scaffold protein S with two partners, A and B, with A consists of a modification state and B consists of two modification states, the binding equilibrium produced the following equations:

$$\begin{aligned}
K_a^{SA} &= \frac{[S(1,0)]}{[S(0,0)][A]} = \frac{[S(1,1)]}{[S(0,1)][A]} = \frac{[S(1,2)]}{[S(0,2)][A]} & (4.51) \\
K_a^{SB} &= \frac{[S(0,1)]}{[S(0,0)][B]} = \frac{[S(1,1)]}{[S(1,0)][B]} \\
K_a^{SBp} &= \frac{[S(0,2)]}{[S(0,0)][B_p]} = \frac{[S(1,2)]}{[S(1,0)][B_p]}
\end{aligned}$$

Using the DN following lumped definition,

$$\begin{aligned}
S(1,*) &= S(1,0) + S(1,1) + S(1,2) & (4.52) \\
S(*,1) &= S(0,1) + S(1,1) \\
S(*,2) &= S(0,2) + S(1,2) \\
S(*,*) &= S(0,0) + S(0,1) + S(0,2) + S(1,0) + S(1,1) + S(1,2)
\end{aligned}$$

one may obtain the following relationship:

$$K_a^{SA} = \frac{[S(1,0)]}{[S(0,0)][A]} = \frac{[S(1,1)]}{[S(0,1)][A]} = \frac{[S(1,2)]}{[S(0,2)][A]} = \frac{[S(1,*)]}{[S(0,*)][A]} \quad (4.53)$$

$$K_a^{SB} = \frac{[S(0,1)]}{[S(0,0)][B]} = \frac{[S(1,1)]}{[S(1,0)][B]} = \frac{[S(*,1)]}{[S(*,0)][B]} \quad (4.54)$$

$$K_a^{SBp} = \frac{[S(0,2)]}{[S(0,0)][B_p]} = \frac{[S(1,2)]}{[S(1,0)][B_p]} = \frac{[S(*,2)]}{[S(*,0)][B_p]} \quad (4.55)$$

Because the forms of Equations 4.53– 4.54 are similar to Problem 1, then the combinatorial binding complex follows the molar fractional expression similar to Problem 1:

$$\frac{[S(1,1)]}{[S(*,*)]} = \frac{[S(*,1)]}{[S(*,*)]} \frac{[S(1,*)]}{[S(*,*)]} \quad (4.56)$$

Similarly, the forms of Equations 4.53, 4.55 are similar to Problem 1, and we also obtained molar fractional expression of the combinatorial binding complex:

$$\frac{[S(1,2)]}{[S(*,*)]} = \frac{[S(*,2)]}{[S(*,*)]} \frac{[S(1,*)]}{[S(*,*)]} \quad (4.57)$$

Using induction, we can extend this proof to any network containing scaffold protein partners with more modification states.

Problem 3. For a network consisting of a scaffold protein S with three partners, A, B, and C, the binding equilibrium produced the following equations:

$$\begin{aligned}
K_a^{SA} &= \frac{[S(1,0,0)]}{[S(0,0,0)][A]} = \frac{[S(1,1,0)]}{[S(0,1,0)][A]} = \frac{[S(1,1,1)]}{[S(0,1,1)][A]} & (4.58) \\
K_a^{SB} &= \frac{[S(0,1,0)]}{[S(0,0,0)][B]} = \frac{[S(1,1,0)]}{[S(1,0,0)][B]} = \frac{[S(1,1,1)]}{[S(1,0,1)][B]} \\
K_a^{SC} &= \frac{[S(0,0,1)]}{[S(0,0,0)][C]} = \frac{[S(1,0,1)]}{[S(1,0,0)][C]} = \frac{[S(1,1,1)]}{[S(1,1,0)][C]}
\end{aligned}$$

First, using the following lumping definition,

$$\begin{aligned}
S(1,*,*) &= S(1,0,0) + S(1,1,0) + S(1,0,1) + S(1,1,1) & (4.59) \\
S(*,1,*) &= S(0,1,0) + S(1,1,0) + S(0,1,1) + S(1,1,1) \\
S(1,1,*) &= S(1,1,0) + S(1,1,1) \\
S(*,*,*) &= S(0,0,0) + S(0,1,0) + S(1,0,0) + S(1,1,0) \\
&\quad + S(0,0,1) + S(0,1,1) + S(1,0,1) + S(1,1,1)
\end{aligned}$$

we obtained expressions similar to Equation 4.49, and thus, molar fractional expression also holds:

$$\frac{[S(1, 1, *)]}{[S(*, *, *)]} = \frac{[S(*, 1, *)]}{[S(*, *, *)]} \frac{[S(1, *, *)]}{[S(*, *, *)]} \quad (4.60)$$

Second, when we defined this lumping definition:

$$S(*, *, 1) = S(0, 0, 1) + S(0, 1, 1) + S(1, 0, 1) + S(1, 1, 1) \quad (4.61)$$

we obtained this simplified dynamic equation:

$$\begin{aligned} \frac{([S(*, *, *)] - [S(*, *, 1)])[C]}{[S(*, *, 1)]} &= \frac{([S(1, 1, *)] - [S(1, 1, 1)])[C]}{[S(1, 1, 1)]} & (4.62) \\ \frac{[S(1, 1, *)]}{[S(*, *, 1)]} &= \frac{[S(1, 1, *)]}{[S(*, *, *)]} \\ \frac{[S(1, 1, 1)]}{[S(*, *, *)]} &= \frac{[S(1, 1, *)]}{[S(*, *, *)]} \frac{[S(*, *, 1)]}{[S(*, *, *)]} \\ \frac{[S(1, 1, *)]}{[S(*, *, *)]} &= \frac{[S(1, *, *)]}{[S(*, *, *)]} \frac{[S(*, 1, *)]}{[S(*, *, *)]} \frac{[S(*, *, 1)]}{[S(*, *, *)]} \end{aligned}$$

Using induction, we can extend the proof for a network containing scaffold proteins with more domain sites.

Problem 4. For a network containing a scaffold protein that consists of more modification states, the different modification states of the scaffold protein are connected by modification fluxes. Because the binding equilibrium approximation assumes negligible modification fluxes, then the network reduces to separate networks with multiple protein scaffolds (Problem 5).

Problem 5. For a network containing two scaffold proteins, R and S, each with two domains, with interaction networks defined as the first domain of R interacting with the first domain of S, and the second domain of R interacting with protein A, and the second domain of S interacting with protein B, we defined the following mass-action kinetics variables: $A, B, R(i, j), S(i, j), RS(i, j)$, where $R(i), i \in \{0, 1\}$ is for the complex between R and A when R does not bind S, $S(i), i \in \{0, 1\}$ is for the complex between S and B when R does not bind S, $RS(i, j), i, j \in \{0, 1\}$ is for the complex between R and S bound to A (i) and/or B (j). In this network, the binding equilibrium applied to the MA model produces:

$$\begin{aligned} k_{\text{on}}^{RA} AR(0) &= k_{\text{off}}^{RA} R(1) & (4.63) \\ k_{\text{on}}^{RA} ARS(0, 0) &= k_{\text{off}}^{RA} RS(1, 0) \\ k_{\text{on}}^{RA} ARS(0, 1) &= k_{\text{off}}^{RA} RS(1, 1) \\ k_{\text{on}}^{SB} BS(0) &= k_{\text{off}}^{SB} S(1) \\ k_{\text{on}}^{SB} BRS(0, 0) &= k_{\text{off}}^{SB} RS(0, 1) \\ k_{\text{on}}^{SB} BRS(1, 0) &= k_{\text{off}}^{SB} RS(1, 1) \\ k_{\text{on}}^{RS} R(0)S(0) &= k_{\text{off}}^{RS} RS(0, 0) \\ k_{\text{on}}^{RS} R(0)S(1) &= k_{\text{off}}^{RS} RS(0, 1) \\ k_{\text{on}}^{RS} R(1)S(0) &= k_{\text{off}}^{RS} RS(1, 0) \\ k_{\text{on}}^{RS} R(1)S(1) &= k_{\text{off}}^{RS} RS(1, 1) \end{aligned}$$

When we applied the following lumping:

$$\begin{aligned}
R(0, *) &= R(0) + R(1) & (4.64) \\
R(*, 0) &= R(0) + RS(0, 0) + RS(0, 1) \\
R(*, 1) &= R(1) + RS(1, 0) + RS(1, 1) \\
R(*, *) &= R(0) + R(1) + RS(0, 0) + RS(0, 1) + RS(1, 0) + RS(1, 1) \\
S(0, *) &= S(0) + S(1) \\
S(*, 0) &= S(0) + RS(0, 0) + RS(1, 0) \\
S(*, 1) &= S(1) + RS(0, 1) + RS(1, 1) \\
S(*, *) &= S(0) + S(1) + RS(0, 0) + RS(1, 0) + RS(0, 1) + RS(1, 1) \\
RS(*, *) &= RS(0, 0) + RS(0, 1) + RS(1, 0) + RS(1, 1) \\
A &= A \\
B &= B
\end{aligned}$$

we obtained the following relationship:

$$K_a^{RA} = \frac{[R(1)]}{[R(0)][A]} = \frac{[RS(1, 0)]}{[RS(0, 0)][A]} = \frac{[RS(1, 1)]}{[RS(0, 1)][A]} = \frac{[R(*, 1)]}{[R(*, 0)][A]} \quad (4.65)$$

$$K_a^{SB} = \frac{[S(1)]}{[S(0)][B]} = \frac{[RS(0, 1)]}{[RS(0, 0)][B]} = \frac{[RS(1, 1)]}{[RS(1, 0)][B]} = \frac{[S(*, 1)]}{[S(*, 0)][B]} \quad (4.66)$$

$$K_a^{RS} = \frac{[RS(0, 0)]}{[R(0)][S(0)]} = \frac{[RS(0, 1)]}{[R(0)][S(1)]} = \frac{[RS(1, 0)]}{[R(1)][S(0)]} = \frac{[RS(1, 1)]}{[R(1)][S(1)]} \quad (4.67)$$

which can be simplified to produce molar fractional expression:

$$\frac{RS(1, 1)}{R(*, *)} = \frac{RS(*, *)}{R(*, *)} \frac{R(*, 1)}{R(*, *)} \frac{S(*, 1)}{S(*, *)} \quad (4.68)$$

By induction, we can prove that the molar fractional expression can be held for independent binding equilibrium of a network with more scaffold proteins.

4.5.5 Proof for the nonlinear reverse lumping

For a mass-action kinetic model given by:

$$\dot{x} = f(x, k) \quad (4.69)$$

$(x(t) \in \mathbb{R}^n)$, we proposed a lumping approach, given by:

$$\hat{x} = Mx \quad (4.70)$$

where $\hat{x} \in \mathbb{R}^m, m < n$ and $M \in \mathbb{R}^{m \times n}$. The form of M is motivated by biological arguments. Assume further that M has full rank. Let $\hat{M} \in \mathbb{R}^{n \times m}$ be any $\{1\}$ -inverse of M [2] so that by definition:

$$M\hat{M}M = M. \quad (4.71)$$

If, in addition, M has full rank, then:

$$M\hat{M} = I. \quad (4.72)$$

Assumes that we know by an independent physical argument that all solution $x(t)$

of Eq 4.69 satisfy:

$$x(t) = \hat{M}Mx(t) \quad (4.73)$$

or equivalently

$$(I - \hat{M}M)x(t) = 0 \quad (4.74)$$

i.e., $x(t) \in \mathcal{N}(I - \hat{M}M)$ (i.e., the solutions of Equation 4.69 lie in a subspace of \mathbb{R}^n). The construction is such that under these assumptions M provides an invertible linear mapping from this subspace to \mathbb{R}^m and \hat{M} provides the inverse mapping.

Suppose that $x(t)$ is a solution of Equation 4.69 and define:

$$\hat{x} = Mx \Rightarrow \dot{\hat{x}}(t) = M\dot{x}(t). \quad (4.75)$$

From $x(t)$ a solution of ODE and this definition, it follows that:

$$\dot{\hat{x}} = M\dot{x} = Mf(x, k)\dot{\hat{x}} = M\hat{f}(\hat{x}, k). \quad (4.76)$$

This shows that for every $x(t)$ satisfying $x(t) = \hat{M}Mx(t)$ and ODE, there is a corresponding \hat{x} that satisfies Eq 4.76 and $\hat{x}(t) = Mx(t)$.

On the other hand, suppose that $\hat{x}(t)$ is a solution of Eq 4.76 and define

$$x(t) \equiv \hat{M}\hat{x} \Rightarrow \dot{x}(t) = \hat{M}\dot{\hat{x}}(t). \quad (4.77)$$

From $\hat{x}(t)$ a solution of Eq 4.76 and this definition, it follows that:

$$\dot{x} = \hat{M}\hat{f}(\hat{x}, k) \quad (4.78)$$

$$\dot{x} = \hat{M}Mf(x, k) \quad (4.79)$$

$$\dot{x} = f(x, k). \quad (4.80)$$

Assume that we know by an independent physical argument that all solutions $x(t)$ of ODE satisfy:

$$x(t) = g_{\hat{M}}(Mx(t))Mx(t) \quad (4.81)$$

where $g_{\hat{M}} : \mathbb{R}^m \rightarrow \mathbb{R}^{n \times m}$ and

$$Mg_{\hat{M}}M = M, \quad \forall \hat{x} \in \mathbb{R}^m. \quad (4.82)$$

If M has full rank then it can be shown that a similar equivalence exists between full and reduced systems.

Table 4.5: Canonical parameter values for System I, II and III

Parameter	Values	Unit	Note
System I and II			
k_{on}^{SE}	1.1e1	$\mu M^{-1} s^{-1}$	Typical value for proteins
k_{off}^{SE}	0.085	s^{-1}	obtained from half time measurement in Ste4-Ste5 from [130]
k_{on}^{SK}	1	$\mu M^{-1} s^{-1}$	Typical value for proteins
k_{off}^{SK}	0.1	s^{-1}	computed based on K_d of Ste5-Fus3 obtained from [10]
$k_{on}^{SK_p}$	1	$\mu M^{-1} s^{-1}$	Typical value for proteins
$k_{off}^{SK_p}$	0.1	s^{-1}	assumed equal to k_{off}^{SK}
k_{cat}	3	s^{-1}	based loosely on kcat of raf and MEK
E_0	0.0332	μM	[110]
S_0	0.0358	μM	[110]
K_0	0.0358	μM	[110]
System III			
k_{on}^{SS}	1e2	$\mu M^{-1} s^{-1}$	made up
k_{off}^{SS}	0.5	s^{-1}	made up
k_{on}^{SA}	10	$\mu M^{-1} s^{-1}$	Typical value for proteins
k_{off}^{SA}	0.085	s^{-1}	obtained from half time measurement in Ste4-Ste5 from [130]
k_{on}^{SB}	1	$\mu M^{-1} s^{-1}$	Typical value for proteins
k_{off}^{SB}	0.1	s^{-1}	computed based on K_d of Ste5-Fus3 obtained from [10]
$k_{on}^{SB_p}$	1	$\mu M^{-1} s^{-1}$	Typical value for proteins
$k_{off}^{SB_p}$	0.1	s^{-1}	computed based on K_d of Ste5-Fus3 obtained from [10]
k_{cat}^B	3	s^{-1}	based loosely on kcat of raf and MEK
k_{on}^{SC}	1	$\mu M^{-1} s^{-1}$	Typical value for proteins
k_{off}^{SC}	0.1	s^{-1}	computed based on K_d of Ste5-Fus3 obtained from [10]
$k_{on}^{SC_p}$	1	$\mu M^{-1} s^{-1}$	Typical value for proteins
$k_{off}^{SC_p}$	0.1	s^{-1}	computed based on K_d of Ste5-Fus3 obtained from [10]
k_{cat}^C	3	s^{-1}	based loosely on kcat of raf and MEK
S_0	0.0358	μM	[110]
A_0	0.0332	μM	[110]
B_0	0.0358	μM	[110]
C_0	0.0217	μM	[110]

Chapter 5

Conclusion and Future Directions

In Chapter 2, we have developed a set of computational tools to study design principles of biological networks. The study utilized dynamic optimization to obtain theoretically optimal networks, a multi-objective optimization to examine trade-offs, and finally, applied the framework of “trade-offs” to compare the function of different network structures. This approach allowed us to utilize further biological network models to examine theoretical limits of operational strategies that a network may use when subjected to a specified objective.

Three major areas may improve further the tools that have been developed. First, the development of rigorous global optimization technique that have stronger elimination power than what is currently available. A rigorous global optimization tool with stronger bounding will include guarantees of the results, and open the door toward systematically varying network structure in the context of a mixed integer dynamic optimization formulation. Second, an improved post-optimality analysis that accounts for multi-dimensions. While the one-dimensional-sensitivity-based method that we constructed may allow us to extract knowledge from optimal results, a more rigorous multi-dimensional system may extend the analysis for more complex systems. And finally, a more efficient parametric optimization method. In our approach, the

parametric optimization is performed by adding a constraint for one of the objectives and re-optimizing; however, many optimizations may benefit from runs of adjacent objective values, and thus, a more efficient approach may be fruitful for application to larger systems.

In Chapters 2 and 3, we applied tools developed here to study design principles in MAP kinase networks, in particular, how dynamical properties such as responsiveness, transient noise filtering, and amplification are affected by the construction of the networks. Results indicate that networks with multiple activation sites may have advantages in filtering transient noise better while modulating output amplitude across wider ranges. The methods may be readily extended to investigate design principles of other network motifs [104].

In Chapter 4, we proposed a general compact modeling methodology applicable to problems with a combinatorial set of components that can be numerically unmanageable, and we applied it to model signal transduction networks by representing the problem as a network of protein domains. This change of representation produced a compact model whose number of species scaled quadratically to the number of protein binding sites and modification states and produced a tremendous saving over the combinatorial scaling in the more standard mass-action kinetic approach. When compared to the more established model reduction methods, this method avoids construction of a fully enumerated mass-action kinetic model yet is relatively accurate across a wider range of parameterizations.

We expect further applications of the proposed method to networks with scaffold proteins and to other large, combinatorial networks. Recent work [106] indicates that Ste5, a scaffold protein in yeast mating pathway, plays a significant role in modulation of mating response through allosteric interactions, instead of being only a passive tethering molecule. The scaffold protein, although difficult to detect because of low sequence conservation, has also been found in other MAP kinase pathways of

different organisms [3]. Thus, incorporation of scaffold proteins into the network of signal transduction pathways should have increasing importance as more becomes known about scaffold proteins. In other but related applications, many of the surface receptors in signal transduction networks consist of multiple splice variants and cross-dimerization species; these systems typically consist of too large number of species when modeled with mass-action kinetics. A complete network that consists all of these proteins may be interesting to study in detail and compared to simplified models. Application of the domain network modeling method to these systems may allow incorporation of many known molecular details and hopefully will provide greater utilization of models in the study of design principles in biological networks.

Appendix A

Tables

Table A.1: Canonical parameter values in a unit network of one-step activation-deactivation reactions.

Parameter	Canonical value	References
<i>Rate constants</i>		
$k_{\text{on}}^{[F:A]}$	$0.1 \mu M^{-1} s^{-1}$	[51, 22]
$k_{\text{off}}^{[F:A]}$	$0.033 s^{-1}$	[51, 22]
$k_{\text{cat}}^{[F:A]}$	$16 s^{-1}$	[51, 22]
$k_{\text{on}}^{[R:A']}$	$5 \mu M^{-1} s^{-1}$	[28]
$k_{\text{off}}^{[R:A']}$	$0.5 s^{-1}$	[28]
$k_{\text{cat}}^{[R:A']}$	$0.3 s^{-1}$	[28]
<i>Initial conditions</i>		
$[F]_0$	$20 \mu M$	[22]
$[A]_0$	$34 \mu M$	[22]
$[R]_0$	$16 \mu M$	[22]
$[F : A]_0$	$0 \mu M$	problem definition
$[A']_0$	$0 \mu M$	problem definition
$[R : A']_0$	$0 \mu M$	problem definition

Table A.2: Differential equations governing a one-step enzymatic activation-deactivation reaction network

One-step enzymatic activation-deactivation reactions			
$\frac{d[A]}{dt} = -w_1 + w_2$		$+w_6$	$w_1 = \mu_1 k_{\text{on}}^{[F:A]} [F][A]$
$\frac{d[F:A]}{dt} = +w_1 - w_2$	$+w_3$		$w_2 = \mu_2 k_{\text{off}}^{[F:A]} [F:A]$
$\frac{d[A']}{dt} =$	$+w_3 - w_4 + w_5$		$w_3 = \mu_3 k_{\text{cat}}^{[F:A]} [F:A]$
$\frac{d[R]}{dt} =$	$-w_4 + w_5$	$+w_6$	$w_4 = \mu_4 k_{\text{on}}^{[R:A']} [R][A']$
$\frac{d[R:A']}{dt} =$	$+w_4 - w_5$	$-w_6$	$w_5 = \mu_5 k_{\text{off}}^{[R:A']} [R:A']$
			$w_6 = \mu_6 k_{\text{cat}}^{[R:A']} [R:A']$

Table A.3: Differential equation governing a two-step enzymatic activation-deactivation reaction network

Two-step enzymatic activation-deactivation reactions	
$\frac{[A]}{dt} =$	$-w_1 + w_2 \qquad \qquad \qquad + w_{12}$
$\frac{[F : A]}{dt} =$	$w_1 - w_2 - w_3$
$\frac{[A']}{dt} =$	$\qquad + w_3 \quad -w_4 + w_5 \qquad \qquad + w_9 \quad -w_{10} + w_{11}$
$\frac{[F : A']}{dt} =$	$\qquad \qquad + w_4 - w_5 - w_6$
$\frac{[A'']}{dt} =$	$\qquad \qquad \qquad + w_6 \quad -w_7 + w_8$
$\frac{[R]}{dt} =$	$\qquad \qquad \qquad -w_7 + w_8 + w_9 \quad -w_{10} + w_{11} + w_{12}$
$\frac{[R : A'']}{dt} =$	$\qquad \qquad \qquad + w_7 - w_8 - w_9$
$\frac{[R : A']}{dt} =$	$\qquad \qquad \qquad \qquad \qquad + w_{10} - w_{11} - w_{12}$
$w_1 =$	$\mu_1 k_{\text{on}}^{[F:A]} [F] [A]$
$w_2 =$	$\mu_2 k_{\text{off}}^{[F:A]} [F : A]$
$w_3 =$	$\mu_3 k_{\text{mod}}^{[F:A]} [F : A]$
$w_4 =$	$\mu_4 k_{\text{on}}^{[F:A']} [A'] [F]$
$w_5 =$	$\mu_5 k_{\text{off}}^{[F:A']} [F : A']$
$w_6 =$	$\mu_6 k_{\text{mod}}^{[F:A']} [F : A']$
$w_7 =$	$\mu_7 k_{\text{on}}^{[R:A'']} [A''] [R]$
$w_8 =$	$\mu_8 k_{\text{off}}^{[R:A'']} [R : A'']$
$w_9 =$	$\mu_9 k_{\text{mod}}^{[R:A'']} [R : A'']$
$w_{10} =$	$\mu_{10} k_{\text{on}}^{[R:A']} [A'] [R]$
$w_{11} =$	$\mu_{11} k_{\text{off}}^{[R:A']} [R : A']$
$w_{12} =$	$\mu_{12} k_{\text{mod}}^{[R:A']} [R : A']$

Bibliography

- [1] Arkin A, Ross J, and McAdams HH. Stochastic kinetic analysis of developmental pathway bifurcation. *Genetics*, 149:1633–1648, 1998.
- [2] Ben-Israel A. and Greville T.N.E. *Generalized Inverses: Theory and Applications*. Springer-Verlag, New York, second edition, 2003.
- [3] Breitkreutz A and Tyers M. A sophisticated scaffold wields a new trick. *Science*, 311.
- [4] Breitkreutz A and Tyers M. MAPK signaling specificity: It takes two to tango. *Trends Cell Biol.*, 12:254–257, 2002.
- [5] Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Gustavo Pesce C, and Brent R. Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437:699–706, 2005.
- [6] Goldbeter A, Stock JB, and Koshland DE Jr. Amplification and adaptation in regulatory and sensory systems. *Science*, 217:220–225, 1982.
- [7] Hoffmann A, Levchenko A, Scott ML, and Baltimore D. The I κ B-NF- κ B signaling module: Temporal control and selective gene activation. *Science*, 298(5596):1241–5, 2002.
- [8] Levchenko A, Bruck J, and Sternberg PW. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl. Acad. Sci. U. S. A.*, 97:5818–5823, 2000.
- [9] Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, and Alon U. Just-in-time transcription program in metabolic pathways. *Nature genetics*, 36(5):486–491, 2004.
- [10] Kusari AB, Molina DM, Sabbagh W, Lau CS, and Bardwell L. A conserved protein interaction network involving the yeast map kinases fus3 and kss1. *J. Cell. Biol.*, 164:267–277, 2004.
- [11] Singer AB, Taylor JW, Barton PI, and Green WH. Global dynamic optimization for parameter estimation in chemical kinetics. *J. Phys. Chem. A*, 110:971–976, 2006.

- [12] Singer AB and Barton PI. Global solution of non linear ordinary differential equations. *Journal of Global Optimization*, 34:159–190, 2006.
- [13] Sharrocks AD, Yang SH, and Galanis A. Docking domains and substrate-specificity determination for MAP kinases. *Trends Biochem. Sci.*, 25:448–453, 2000.
- [14] Whitmarch AJ, Cavanagh J, Tournier C, Yausda J, and Davis RJ. A mammalian scaffold complex that selectively mediates MAP kinase activation. *Science*, 281:1671–1674, 1999.
- [15] Asthagiri AR and Lauffenburger DA. A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (MAPK) pathway model. *Biotechnol. Progr.*, 17:227–239, 2001.
- [16] Kansal AR and Trimmer J. Application of predictive biosimulation within pharmaceutical clinical development: examples of significance for translational medicine and clinical trial design. *IEE Proc. Systems Biol*, 152:214–220, 2005.
- [17] Dhillon AS and Kolch W. Untying the regulation of the Raf-1 kinase. *Arch. of Biochem. Biophys.*, 404:3–9, 2002.
- [18] Chachuat B, Singer AB, and Barton PI. Global mixed integer dynamic optimization. *AIChE J.*, 51(8):2235–2253, 2005.
- [19] Goldstein B, Faeder JR, Hlavacek WS, Blinov ML, Redondo A, and Wofsy C. Modeling the early signaling events mediated by Fc ϵ RI. *Mol. Immunol.*, 38:1213–1219, 2002.
- [20] Houchmandzadeh B, Wieschaus E, and Leibler S. Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature*, 415:798–802, 2002.
- [21] McCright B, Rivers AM, Audlin S, and Virshup DM. The B56 family of protein phosphatase 2A (PP2A) regulatory subunits encodes differentiation-induced phosphoproteins that target PP2A to both nucleus and cytoplasm. *J. Biol. Chem.*, 271:22081–22089, 1996.
- [22] Schoeberl B, Eichler-Jonsson C, Gilles ED, and Muller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.*, 20:370–375, 2002.
- [23] Dahiyat BI and Mayo SL. De novo protein design: Fully automated sequence selection. *Science*, 278:82–87, 1997.
- [24] Adiwijaya BS, Barton PI, and Tidor B. Biological network design principle: Discovery through optimization. *in preparation*.

- [25] Adiwijaya BS, Barton PI, and Tidor B. Differentiating signal from transient noise in non-linear modules of signal transduction networks. *in preparation*.
- [26] Mattison CP, Spencer SS, Kresge KA, Lee J, and Ota IM. Differential regulation of the cell wall integrity mitogen-activated protein kinase pathway in budding yeast by the protein tyrosine phosphatases ptp2 and ptp3. *Mol. Cell. Biol.*, 19:7651–7660, 1999.
- [27] Rao CV, Kirby JR, and Arkin AP. Design and diversity in bacterial chemotaxis: A comparative study in *Escherichia coli* and *Bacillus subtilis*. *PLoS Biol*, 2(2):0239–0252, 2004.
- [28] Huang CY and Ferrell JE Jr. Ultra-sensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. U. S. A.*, 93:10078–10083, 1996.
- [29] Bokemeyer D, Sorokin A, Yan M, Ahn NG, Templeton DJ, and Dunn MJ. Induction of mitogen-activated protein kinase phosphatase 1 by the stress-activated protein kinase signaling pathway but not by extracellular signal-regulated kinase in fibroblasts. *J. Biol. Chem.*, 271:639–642, 1996.
- [30] Endy D, Kong D, and Yin J. Intracellular kinetics of a growing virus: A genetically structured simulation for bacteriophage $\tau 7$. *Biotechnol. Bioeng.*, 55:375–389, 1997.
- [31] Endy D, You L, Yin J, and Molineux IJ. Computational, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl. Acad. Sci. U. S. A.*, 97:5375–5380, 2000.
- [32] Voet D, Voet JG, and Pratt CW. *Fundamentals of Biochemistry*. Wiley, New York, 1999.
- [33] Schwer DA, Tolsma JE, Green Jr. WH, and Barton PI. On upgrading the numerics in combustion chemistry codes. *Combustion and Flame*, 128(3):270–291, 2002.
- [34] Zak DE, Stelling J, and Doyle FJ. Sensitivity analysis of oscillatory (bio)chemical systems. *Comput. Chem. Eng.*, 29:663–673, 2005.
- [35] Gillespie DT. Exact stochastic simulation of coupled chemical-reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.
- [36] Dekel E and Alon U. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436:588–592, 2005.
- [37] Kangas E and Tidor B. Electrostatic specificity in molecular ligand design. *J. Chem. Phys.*, 112:9120–9131, 2000.
- [38] Korobkova E, Emonet T, Vilar J M G, Shimizu T S, and Cluzel P. From molecular noise to behavioural variability in a single bacterium. *Nature*, 428:574–578, 2004.

- [39] Martin-Blanco E, Gampel A, Ring J, Virdee K, Kirov N, Tolkovsky AM, and Martinez Arias A. puckerred encodes a phosphatase that mediates a feedback loop regulating JNK activity during dorsal closure in *Drosophila*. *Genes Dev.*, 12:557–570, 1998.
- [40] Elion EA. The Ste5p scaffold. *J. Cell Sci.*, 114:3967–3978, 2001.
- [41] Davidson EH, Rast JP, Oliveri P, et al. A genomic regulatory network for development. *Science*, 295:1669–1678, 2002.
- [42] Grimme EJ. *Krylov Projection methods for model reduction*. PhD thesis, University of Illinois, Urbana-Champaign, IL, 1997.
- [43] Roy F and Therrien M. MAP kinase module: The KSR connection. *Curr. Biol.*, 12:R325–R327, 2002.
- [44] Altan-Bonnet G and Germain RN. Modeling t cell antigen discrimination based on feedback control of digital erk responses. *PLoS Biol.*, 3:1925–1938, 2005.
- [45] Li G. Lumping analysis in mono- or/and bimolecular reaction systems. *Chem. Eng. Sci.*, 29:1261–1270, 1984.
- [46] Li G and Rabitz H. A general-analysis of exact lumping in chemical-kinetics. *Chem. Eng. Sci.*, 44:1413–1430, 1989.
- [47] Pearson G, Robinson F, Gibson TB, Karandikar M, Berman K, and Cobb MH. Mitogen-activated protein (MAP) kinase pathways: Regulation and physiological functions. *Endocrine Rev.*, 22:153–183, 2001.
- [48] Von Dassow G, Meir E, Munro EM, and Odell GM. The segment polarity network is a robust developmental module. *Nature*, 406:188–192, 2000.
- [49] Conzelmann H, Saez-Rodriguez J, Sauter T, Kholodenko BN, and Gilles ED. A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinfo.*, 7:1–15, 2006.
- [50] Sun H, Charles CH, Lau LF, and Tonks NK. MKP-1 (3CH134), an immediate-early gene-product, is a dual-specificity phosphatase that dephosphorylates map kinase in-vivo. *Cell*, 75:487– 493, 1993.
- [51] El-Masri HA and Portier CJ. Replication potential of cells via the protein kinase C-MAPK pathway: Application of a mathematical model. *Bull. Math. Biol.*, 61:379–398, 1999.
- [52] Madhani HD, Styles CA, and Fink GR. MAP kinases with distinct inhibitory functions impart signaling specificity during yeast differentiation. *Cell*, 91:673–684, 1997.
- [53] Madhani HD and Fink GR. The riddle of MAP kinase signaling specificity. *Trends Genet.*, 14:151–155, 1998.

- [54] Dohlman HG and Thorner JW. Regulation of G protein-initiated signal transduction in yeast: Paradigms and principles. *Annu. Rev. Biochem.*, 70:703–754, 2001.
- [55] McAdams HH and Shapiro L. Circuit simulation of genetic networks. *Science*, 269:650–656, 1995.
- [56] Wiley HS, Shvartsman SY, and Lauffenburger DA. Computational modeling of the egf-receptor system: A paradigm for systems biology. *Trends Cell Biol.*, 13(1):43–50, 2003.
- [57] Kuhn HW and Tucker AW. Nonlinear programming. In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, California, 1951.
- [58] Duff IS and Reid JK. A FORTRAN code for direct solution of sparse unsymmetric linear systems of equations. Technical Report RAL-93-072, Rutherford Appleton Laboratory, Oxon, UK, 1993.
- [59] Andersson J, Simpson DM, Qi M, Wang Y, and Elion EA. Differential input by Ste5 scaffold and Msg5 phosphatase route a MAPK cascade to multiple outcomes. *EMBO J.*, 23:2564–2576, 2004.
- [60] Gunawardena J. Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proc. Natl. Acad. Sci. U. S. A.*, 102:14617–22, 2005.
- [61] Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell*, 103:211–225, 2000.
- [62] Stelling J, Gilles ED, and Doyle FJ III. Robustness properties of circadian clock architecture. *Proc. Natl. Acad. Sci. U. S. A.*, 101(36):13210–13215, 2004.
- [63] Wei J and Kuo JCW. A lumping analysis in monomolecular reaction systems — analysis of exactly lumpable system. *Ind. Eng. Chem. Fundam.*, 8:114–123, 1969.
- [64] Zallen JA and Wieschaus E. Patterned gene expression directs bipolar planar polarity in *Drosophila*. *Dev. Cell.*, 6:343–355, 2004.
- [65] Stock JB and Surette MG. In Neidhardt FC, editor, *E. coli and Salmonella, Cellular and Molecular Biology*, pages 1103–1129, Washington DC, 1996. American Society of Microbiology.
- [66] Cuthrell JE and Biegler LT. On the optimization of differential-algebraic process systems. *AIChE J.*, 33:1257–1270, 1987.
- [67] Tolsma JE, Clabaugh JA, and Barton PI. Symbolic incorporation of external procedures into process modeling environments. *Ind. Eng. Chem. Res.*, 41(16):3867–3876, 2002.

- [68] Falke JJ, Bass RB, Butler SL, Chervitz SA, and Danielson MA. The two-component signaling pathway of bacterial chemotaxis: A molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu. Rev. Cell Dev. Biol.*, 13:457–512, 1997.
- [69] Tyson JJ. Modeling the cell-division cycle Cdc2 and Cyclin interactions. *Proc. Natl. Acad. Sci. U. S. A.*, 88:7328–7332, 1991.
- [70] Brondello JM, Pouyssegur J, and McKenzie FR. The dual specificity mitogen-activated protein kinase phosphatase-1 and -2 are induced by the p42/p44^{MAPK} cascade. *J. Biol. Chem.*, 272:1368–1376, 1997.
- [71] Brondello JM, Pouyssegur J, and McKenzie FR. Reduced MAP kinase phosphatase-1 degradation after p42/p44MAPK-dependent phosphorylation. *Science*, 286:2514–2517, 1999.
- [72] Faeder JR, Blinov ML, Goldstein B, and Hlavacek WS. Combinatorial complexity and dynamical restriction of network flows in signal transduction. *IEE Systems Biology*, 2:5–15, 2005.
- [73] Ferrell JE Jr and Machleder EM. The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science*, 280:895–898, 1998.
- [74] Teo K, Goh G, and Wong K. *A Unified Computational Approach to Optimal Control Problems*. Pitman Monographs and Surveys in Pure and Applied Mathematics. John Wiley and Sons, Inc., New York, 1991.
- [75] Brenan KE, Campbell SL, and Petzold LR. *Numerical Solution of Initial Value Problems in Differential-Algebraic Equations*. SIAM, Philadelphia,PA, 1996.
- [76] Lok L and Brent R. Automatic generation of cellular reaction networks with Molecuizer 1.0. *Nat. Biotechnol.*, 23:131–136, 2005.
- [77] Groom LA, Sneddon AA, Alessi DR, Dowd S, and Keyse SM. Differential regulation of the MAP, SAP and ERK/p38 kinases by pyst1, a novel cytosolic dual-specificity phosphatase. *EMBO Journal*, 15:3621–3632, 1996.
- [78] Leong CC Defranoux N Holgate ST Stokes CL Lewis AK, Paterson T. The roles of cells and mediators in a computer model of chronic asthma. *Int. Arch. Allergy Immunol.*, 124:282–286, 2001.
- [79] Lee LP and Tidor B. Barstar is electrostatically optimized for tight binding to barnase. *Nat. Struct. Biol.*, 8:73–76, 2001.
- [80] Chong LT, Dempster SE, Hendsch ZS, Lee LP, and Tidor B. Computation of electrostatic complements to proteins: A case of charge stabilized binding. *Prot. Sci*, 7:206–210, 1998.

- [81] Camps M et al. Catalytic activation of the phosphatase MKP-3 by ERK2 mitogen-activated protein kinase. *Science*, 280:1262–1265, 1998.
- [82] Hammer M, Mages J, Dietrich H, Schmitz F, Striebel F, Murray PJ, Wagner H, and Lang R. Control of dual-specificity phosphatase-1 expression in activated macrophages by IL-10. *Euro J. Immunol.*, 35:2991– 3001, 2005.
- [83] Samoilov M, Plyasunov S, and Arkin AP. Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. *Proc. Natl. Acad. Sci. U. S. A.*, 102(2):2310–2315, 2005.
- [84] Thattai M and van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.*, 98:8614–8619, 2001.
- [85] Schwartz MA and Madhani HD. Principles of MAP kinase signaling specificity in *Saccharomyces cerevisiae*. *Annu. Rev. Genetics*, 38:725–748, 2004.
- [86] Elowitz MB, Levine AJ, Siggia ED, and Swain PS. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.
- [87] Wishart MJ and Dixon JE. Gathering styx: Phosphatase like form predicts functions for unique protein-interaction domains. *Trends Biochem. Sci.*, 23:301–306, 1998.
- [88] Blinov ML, Faeder JR, Goldstein B, and Hlavacek WS. BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20:3289–3291, 2004.
- [89] Maurya MR, Bornheimer SJ, Venkatasubramanian V, and Subramaniam S. Reduced-order modelling of biochemical networks: application to the gtpase-cycle signalling module. *IEE Proc.-Syst. Biol.*, 152(4):229–242, 2005.
- [90] Bazaraa MS, Sherali HD, and Shetty CM. *Nonlinear Programming*. John Wiley and Sons, Inc., second edition, 1993.
- [91] Okino MS and Mavrovouniotis ML. Simplification of mathematical models of chemical reaction systems. *Chem. Rev.*, 98:391–408, 1998.
- [92] Springer MS, Goy MF, and Adler J. Protein methylation in behavioural control mechanism and in signal transduction. *Nature*, 280:279–284, 1979.
- [93] Bao MZ, Schwartz MA, Cantin GT, Yates JR 3rd, and Madhani HD. Pheromone-dependent destruction of the Tec1 transcription factor is required for MAP kinase signaling specificity in yeast. *Cell*, 119:991–1000, 2004.
- [94] Barkai N and Leibler S. Robustness in simple biochemical networks. *Nature*, 387:913–917, 1997.
- [95] Tonks NK and Neel BG. Combinatorial control of the specificity of protein tyrosine phosphatases. *Curr. Opin. Cell. Biol.*, 13:128–195, 2001.

- [96] Borisov NM, Markevich NI, Hoek JB, and Kholodenko B. Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity. *Biophys. J.*, 89:951–966, 2005.
- [97] Jorgensen P, Nishikawa JL, Breitskreutz BJ, and Tyers M. Systematic identification of pathways that couple cell growth and division in yeast. *Science*, 297:395–400, 2002.
- [98] Clark PA and Westerberg AW. Optimization for design problems having more than one objective. *Comput. Chem. Eng.*, 7(4):259–278, 1983.
- [99] Gill PE, Murray W, Saunders MA, and Wright MH. User’s guide for NPSOL 5.0: A Fortran package for Nonlinear Programming. *Report SOL 86-1, Department of Operations Research, Stanford University, Stanford, CA*, 1998.
- [100] Li QJ, Dinner AR, Qi S, Irvine DJ, Huppa JB, Davis MM, and Chakraborty AK. CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse. *Nat. Immunol.*, 5(8):791–799, 2004.
- [101] Brent R and Endy D. Modelling cellular behaviour. *Nature*, 409:391–395, 2001.
- [102] Gunawan R, Cao Y, Petzold L, and Doyle FJ. Sensitivity analysis of discrete stochastic systems. *Biophys. J.*, 88:2530–2540, 2005.
- [103] Heinrich R, Neel BG, and Rapoport TA. Mathematical models of protein kinase signal transduction. *Mol. Cell*, 9:957–970, 2002.
- [104] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, and Alon U. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002.
- [105] Davis RJ. Signal transduction by the JNK group of MAP kinases. *Cell*, 103:239–252, 2000.
- [106] Bhattacharyya RP, Remenyi A, Good MC, Bashor CJ, Falick AM, and Lim WA. The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science*, 311:822–826, 2006.
- [107] Chapman S and Asthagiri AR. Resistance to signal activation governs design features of the MAP kinase signaling module. *Biotechnol. Bioeng.*, 85(3):311–322, 2004.
- [108] Chou S, Huang L, and Liu H. Fus3-regulated Tec1 degradation through SCF Cdc4 determines MAPK signaling specificity during mating in yeast. *Cell*, 119:981–990, 2004.
- [109] Erdman S and Snyder M. A filamentous growth response mediated by the yeast mating pathway. *Genetics*, 159(3):919–928, 2001.

- [110] Ghaemmaghami S, Huh W, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, and Weissman JS. Global analysis of protein expression in yeast. *Nature*, 425:737–741, 2003.
- [111] Vajda S, Valko P, and Turanyi T. Principal component analysis of kinetic models. *Int. J. Chem. Kinet.*, 17:55–81, 1985.
- [112] Elena SF and Lenski RE. Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nature Rev. Genet.*, 4:457–469, 2003.
- [113] Bornheimer SJ, Maurya MR, Farquhar MG, and Subramaniam S. Computational modeling reveals how interplay between components of a GTPase-cycle module regulates signal transduction. *Proc. Natl. Acad. Sci. U. S. A.*, 101:15899–15904, 2004.
- [114] Keyse SM. Protein phosphatases and the regulation of MAP kinase activity. *Semin. Cell Dev. Biol.*, 9:143–152, 1998.
- [115] Keyse SM. Protein phosphatases and the regulation of mitogen-activated protein kinase signalling. *Curr. Opin. Cell. Biol.*, 12:186–192, 2000.
- [116] O'Rourke SM, Herskowitz I, and O'Shea EK. Yeast go the whole hog for the hyperosmotic response. *Trends Genet.*, 18:405–412, 2002.
- [117] Wurgler-Murphy SM, Maeda T, Witten EA, and Saito H. Regulation of the *saccharomyces cerevisiae* HOG1 mitogen-activated protein kinase by the PTP2 and PTP3 protein tyrosine phosphatases. *Mol. Cell. Biol.*, 17:1289–1297, 1997.
- [118] Hunter T. Signaling-2000 and beyond. *Cell*, 100:113–127, 2000.
- [119] Ideker T and Lauffenburger D. Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol.*, 21:255–262, 2003.
- [120] Ideker T, Galitski T, and Hood L. A new approach to decoding life: Systems biology. *Annu. Rev. Genomics Hum. Genet.*, 2:343–372, 2001.
- [121] Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, and Hood L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292:929–934, 2001.
- [122] Jacoby T, Flanagan H, Faykin A, Seto AG, Mattison C, and Ota I. Two protein-tyrosine phosphatases inactivate the osmotic stress response pathway in yeast by targeting the mitogen-activated protein kinase, Hog1. *J. Biol. Chem.*, 272:17749–17755, 1997.
- [123] Tanoue T, Adachi M, Moriguchi T, and Nishida E. A conserved docking motif in MAP kinases common to substrate, activators and regulators. *Nat. Cell Biol.*, 2:110–116, 2000.

- [124] Turanyi T, Berces T, and Vajda S. Reaction-rate analysis of complex kinetic systems. *Int. J. Chem. Kinet.*, 21:83–99, 1989.
- [125] Haystead TAJ, Dent P, Wu J, Haystead CM, and Sturgill TW. Ordered phosphorylation of p42MAPKmapk by MAP kinase kinase. *FEBS Lett.*, 306:17–22, 1992.
- [126] Yi TM, Huang Y, Simon MI, and Doyle J. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci. U. S. A.*, 97:46–49, 2000.
- [127] Bhalla US, Ram PT, and Iyengar R. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science*, 297:1018–1023, 2002.
- [128] Bhalla US and Iyengar R. Emergent properties of networks of biological signaling pathways. *Science*, 283:381–387, 1999.
- [129] Sakizlis V, Dua V, Perkins JD, and Pistikopoulos EN. Robust model-based tracking control using parametric programming. *Comput. Chem. Eng.*, 28:195–207, 2004.
- [130] van Drogen F, Stucke VM, Jorritsma G, and Peter M. MAP kinase dynamics in response to pheromones in budding yeast. *Nat. Cell Biol.*, 3:1051–1059, 2001.
- [131] Feehery WF, Tolsma JE, and Barton PI. Efficient sensitivity analysis of large-scale differential-algebraic systems. *Appl. Numer. Math.*, 25:41–54, 1997.
- [132] Esposito WR and Floudas CA. Global optimization for the parameter estimation of differential-algebraic systems. *Ind. Eng. Chem. Res.*, 39:1291–1310, 2000.
- [133] Schulze WX, Deng L, and Mann M. Phosphotyrosine interactome of the ErbB-receptor kinase family. *Nat Molecular Systems Biology*, 2005.
- [134] Yarden Y and Sliwkowski MX. Untangling the ErbB signaling network. *Nat. Rev. Mol. Cell Biol.*, 2:127–137, 2001.