Characterizing Signal Transduction Networks and Biological Responses Using Computer Simulations and Machine Learning

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Aaron Matthew Prescott

May 2018

© by Aaron Matthew Prescott, 2018
All Rights Reserved.

Abstract

The use of computer simulations in biology is often limited due to the lack of experimentally measured parameters. In these scenarios, parameter exploration can be used to probe biological systems and refine understanding of biological mechanisms. For systems with few unknown parameters, parameter sweeps that concurrently vary all unknown parameters are tractable. In complex systems with many unknown parameters, supervised machine learning algorithms can be used to discover parameters leading to targeted system responses. In this thesis, we study three biological problems in which we use parameter exploration methods to gain mechanistic insights. We first explore the role of altered metabolism in cancer cells that reside in heterogeneous tumor microenvironments. We use a multiscale, hybrid cellular automaton model to evaluate tumor progression while varying malignant cell traits using a systematic parameter sweep. The results reveal distinct growth regimes associated with varied malignant cell traits. We then study kinetic mechanisms governing fixed-topology signal transduction networks and use evolutionary algorithms to discover kinetic parameters that produce specified network responses. We analyze the growth-response network in *Arabidopsis* with this supervised machine learning approach. This allows us to identify constraints on kinetic parameters that govern the observed responses. The evolved parameters are used to calculate the responses of individual network components, which are used to generate hypotheses that can be tested *in vivo* to help determine the network topology. We finally apply a similar approach to redesign signal transduction networks. We demonstrate that the T cell receptor network and an oscillator network show remarkable flexibility in generating altered responses to input, and we further use a nonlinear clustering method to identify design criteria for the underlying kinetic parameters. For each project,

observations produced from *in silico* simulations lead to the formation of hypotheses that are experimentally testable.

Table of Contents

1 Introduction **1**

1.1 The Role of Computational Biology 1

1.2 Parameter Exploration Methodologies 2

1.2.1 Machine learning for intelligent parameter exploration 3

1.2.2 Machine learning in biology and soft matter 7

1.3 Outline of Thesis 8

1.3.1 Exploring cancer growth and altered metabolism in heterogeneous environments 9

1.3.2 Analyzing signal transduction networks underlying ethylene growth response kinetics 10

1.3.3 Redesigning signal transduction network response dynamics *in silico* . 10

2 Exploring Cancer Growth and Altered Metabolism in Heterogeneous Environments **12**

2.1 Introduction 12

2.2 Methods 16

2.2.1 Hybrid Cellular Automata (HCA) model of cancer 16

2.2.2 Analyzing simulation results 23

2.3 Results 24

2.3.1 Tumor growth patterns 25

2.3.2 The influence of glucose on tumor progression 28

2.3.3 Varying cancer hallmarks in HCA simulations 30

2.3.4 Effects of angiogenesis 35

2.4 Discussion . . . 36
2.4.1 Isolated clustering . . . 39
2.4.2 Continuous growth . . . 40
2.4.3 Tumor trapping . . . 40
2.4.4 Domain seeding . . . 41
2.4.5 Angiogenesis . . . 41
2.5 Conclusions . . . 42

3 Analysis of network topologies underlying ethylene growth response kinetics in *Arabidopsis* 45
3.1 Introduction . . . 45
3.2 Methods . . . 51
3.2.1 Kinetics of the ethylene response network . . . 51
3.2.2 Evolutionary algorithm . . . 52
3.2.3 Screening for targeted responses . . . 55
3.2.4 Ethylene dose response . . . 57
3.3 Results . . . 58
3.3.1 Coherent feedforward/negative feedback network . . . 58
3.3.2 EIN2 cleavage with positive feedback network . . . 69
3.4 Conclusions . . . 72

4 Resdesigning Signaling Network Responses *In Silico* 77
4.1 Introduction . . . 77
4.1.1 Early T cell receptor (TCR) signaling network . . . 79
4.1.2 Robust oscillator network . . . 81
4.2 Methods . . . 81
4.2.1 Kinetics of the TCR network . . . 81
4.2.2 Kinetics of the robust oscillator network . . . 84
4.2.3 *In silico* evolution of the networks . . . 85
4.2.4 Dimensionality reduction . . . 86
4.2.5 Identification of groups . . . 88

4.3 Results . . . 89
4.3.1 Evolution and analysis of the TCR network . . . 89
4.3.2 Evolution and analysis of the robust oscillator network . . . 101
4.4 Discussion . . . 103
4.5 Conclusions . . . 105

5 Conclusions **106**
5.1 Exploring cancer growth and altered metabolism in heterogeneous environments 107
5.2 Analyzing signal transduction networks underlying ethylene growth response kinetics . . . 108
5.3 Redesigning signal transduction network response dynamics *in silico* . . . 108
5.4 Closing . . . 109

Bibliography **110**

Appendices **125**
A Summary of Signal Transduction Ordinary Differential Equations . . . 126
A.1 Ethylene Signaling in Arabidopsis . . . 126
A.2 Early T Cell Receptor Network . . . 133

Vita **138**

List of Tables

2.1 **Cancer hybrid cellullar automata model parameters** 17

3.1 **Fitness function: Target growth values and scaling factors (wildtype conditions)** 53

3.2 **Fitness function: Target growth values and scaling factors (*ein3;eil1* mutant conditions)** 54

3.3 **Parameter range constraints for ethylene signaling evolutionary algorithms** 55

3.4 **Screening results for ethylene growth responses of simplified coherent feedforward/negative feedback (CFF/NFB) networks** 67

4.1 **Chemical reactions and ranges of kinetic rates for the TCR network** 82

4.2 **Parameters and parameter ranges for the robust oscillator network** 84

List of Figures

2.1 Cancer hybrid cellullar automata cell decision tree 20
2.2 Cancer model assumption: cell displacement vs. replacement upon proliferation 22
2.3 Emergent growth patterns in cancer simulations 26
2.4 Malignant density progression without glucose 28
2.5 Malignant density progression with glucose consumption 29
2.6 Progression of the number of vascular sites without angiogenesis . 30
2.7 Mean density of malignant cells and mean number of vascular sites 31
2.8 Vascular site displacement 32
2.9 Proliferation domain count 33
2.10 Discrete compactness of malignant cells 34
2.11 Malignant density for simulations with angiogenesis 36
2.12 Tumor progression with angiogenesis 37
2.13 Number of bridging events 38

3.1 Basic linear ethylene signaling network 46
3.2 Growth response kinetics of dark-grown *Arabidopsis* seedling hypocotyls 48
3.3 Diagrams illustrating proposed network topologies 49
3.4 Logic diagram for two-phase growth inhibition (2-PGI) screening . 56
3.5 Logic diagram for mutant partial growth recovery (MPGR) screening 57
3.6 Characteristic growth responses at saturating ethylene doses 59

3.7 Time evolution of coherent feedforward/negative feedback (CFF/NFB) network components (wildtype and *ein3/eil1* mutant conditions) . 60
3.8 Time evolution of coherent feedforward/negative feedback (CFF/NFB) network components (wildtype conditions) 60
3.9 Time evolution of coherent feedforward/negative feedback (CFF/NFB) network components (*ein3;eil1* mutant conditions) 61
3.10 Distributions of parameters from evolved sets of coherent feedforward/negative feedback (CFF/NFB) network parameters 63
3.11 Complete parameter distributions for the coherent feedforward/negative feedback (CFF/NFB) network 64
3.12 Characteristic growth responses at sub-saturating ethylene doses . 67
3.13 Distributions of parameters from the coherent feedforward/negative feedback (CFF/NFB) network screened for sub-saturating growth recovery (SSGR) behavior 67
3.14 Parameter distributions for the coherent feedforward/negative feedback (CFF/NFB) network based on sub-saturating growth recovery (SSGR) screening 68
3.15 Simplified coherent feedforward/negative feedback (CFF/NFB) networks 69
3.16 Time evolution of positive feedback (PFB) network components . . 71
3.17 Simplified positive feedback (PFB) network 72
3.18 Time evolution of coherent feedforward/negative feedback (CFF/NFB) and positive feedback (PFB) network components (wildtype conditions) 76

4.1 Schematic overview of method for kinetic optimization of signal transduction response 79
4.2 Schematic of T cell response network species 82
4.3 Robust oscillator network 83
4.4 Shifting the T cell response threshold 90

4.5 Generating graded and inverted T cell responses 91
4.6 T cell response behavior outside of the target range 93
4.7 Dimensionality reduction of native and shifted T cell response results 95
4.8 Dimensionality reduction of native and inverted T cell responses . 97
4.9 Distributions of parameters for groups identified in native and shifted T cell response results 98
4.10 Distributions of parameters for groups identified in native and inverted T cell response results 100
4.11 Evolving the robust oscillator network 101
4.12 Dimensionality reduction of robust oscillator results 102
4.13 Distributions of parameters for groups identified in the robust oscillator results 103

Chapter 1

Introduction

1.1 The Role of Computational Biology

Computational biology is a field of research with a growing footprint in the scientific community. Traditionally, the role of computational biology has been classified into one of two roles: knowledge discovery through data mining of experimental results or *in silico* simulations for hypothesis testing.[83] The separation of these roles is blurred when biological simulations are used to generate large quantities of data that necessitate mining. This combined tactic is particularly useful for analyzing poorly defined biological systems. In these cases, model parameters and simulation rules can be systematically modified to gain insight into system operations. Properly performed, this methodology can be used to transform vague biological concepts into testable hypotheses.[92]

Despite the best efforts of researchers, many experimental models cannot fully elucidate system functionality. Whether experimental shortcomings arise from an inability to recreate *in situ* conditions, a failure to isolate the system from external noise, or the impracticality of taking reliable measurements, results from *in vivo* and *in vitro* systems can be complemented by *in silico* simulations. For computational models to produce useful results, they must remain rooted in biological principles derived from *in vivo*/*in vitro* experimental results. Failure to do so comes with the risk of producing results that merely reflect model design and do not reflect biological phenomena. The feedback between *in silico* models and experimental systems used by computational and molecular biologists, respectively, produces an iterative

loop allowing for successive hypothesis refinement that can lead to accelerated knowledge discovery.

In order to create computational models, biological systems must be given quantitative descriptions. Depending on the model scale, numerical values are assigned to parameters describing such conditions as molecular concentrations, chemical reaction rates, cellular proliferation rates, etc. Frequently, these quantities are unknown. In these situations, systematic parameter exploration can elucidate the possible range of responses a system may produce. Data mining these results can illuminate mechanisms driving system functionality. These insights prove valuable for producing refined, testable hypotheses.

To demonstrate the utility of data mining results produced from *in silico* parameter exploration, we applied the outlined principles to three projects:

1. Examining tumor growth within vascularized tissue for cancer cells with various characteristics (Ch. 2).

2. Testing kinetic constraints of proposed signal transduction networks describing the growth of *Arabidopsis* in response to ethylene exposure (Ch. 3).

3. Exploring the potential to redesign fixed-topology signal transduction networks to produce novel input-output respones (Ch. 4).

1.2 Parameter Exploration Methodologies

For simulations with few unknown parameters, it is sufficient to use a systematic parameter sweep based upon a combinatoric approach for evaluating parameter values. For example, given a system with 3 unknown parameters each having 4 potential values, a total of 64 (4^3) parameter combinations need to be evaluated. This style of parameter sweep is utilized in the cancer research discussed in Ch. 2. When the unknown parameter space grows, this type of exhaustive parameter sweep becomes computationally intractable. In cases such as these, supervised machine learning algorithms can used to determine parameter combinations that produce experimentally-observed system responses. For the signal transduction research

described in Ch. 3 and 4, an evolutionary algorithm serves the role of the supervised machine learning algorithm.

1.2.1 Machine learning for intelligent parameter exploration

Machine learning (ML) is a class of algorithms that seeks to improve performance as measured by a user-defined metric. The algorithms accomplish this by learning from data.[5] There are two major classes of machine learning used in the following research: supervised and unsupervised ML. Semi-supervised and reinforcement learning are two additional machine learning classes, but they are not used in this research. These ML algorithm classes differ in how they learn from data. Both algorithm classes, supervised and unsupervised learning, seek to find parameters that map observed features to predictor variables. In supervised ML algorithms, the predictor variable is a measured feature taken for each observation of a data set that the algorithm attempts to predict using a predictor function and learned parameters. The predictor function uses the learned parameters to transform measured features from each observation into a predicted value for the measured predictor variable. Supervised ML algorithms used to predict continuous values are termed regression algorithms while algorithms that predict discrete categories are termed classifiers. Common supervised ML algorithms include linear regression, logistic regression, and linear discriminant analysis.[51] Linear regression is discussed below to articulate the role of regression analysis and to compare with nonlinear regression methods such as evolutionary algorithms that we use in this thesis. Logistic regression, despite its name, is primarily used as a classifier. Linear discriminant analysis is also a classifier. In contrast to supervised ML, unsupervised ML algorithms use a predictor function to map observed features to novel predictor variables that were not previously measured. Common unsupervised ML algorithms include dimensionality reduction algorithms such as principle component analysis and clustering algorithms such as k-means clustering.[74, 51]

Supervised machine learning (ML) example - Linear regression

All supervised ML algorithms attempt to map a set of features ($\vec{x}_i$) for observation i to a measured predictor variable (y_i). This mapping is accomplished through the use of a predictor function. In the case of linear regression, the predictor function is:

$$\hat{y}_i = \hat{m}\, x_i + \hat{b} \tag{1.1}$$

In this example, the feature vector $\vec{x}_i$ is a single scalar value (x_i) for each observation i and $\hat{y}_i$ is the predicted value for the predictor variable (y_i) that is calculated by the predictor function (Eqn. 1.1). The goal of linear regression is to discover values for $\hat{m}$ and $\hat{b}$ that minimize the difference between the predicted and measured values of y_i across all observations i. That is, given N observations, we seek values for $\hat{m}$ and $\hat{b}$ such that:

$$\min_{\hat{m},\hat{b}} \sum_{i=1}^{N} \left(y_i - (\hat{m} x_i + \hat{b}) \right)^2 \tag{1.2}$$

This function is commonly referred to as the cost function. In this case, the cost function is equivalent to minimizing the residual sum of squares. Due to the convex nature of linear regression minimization problems, the optimization can be solved using linear algorithms such as gradient descent.[74, 51] To optimize more complex cost functions, other optimization algorithms are required.

Supervised machine learning examples - Linear classifiers

Classifier algorithms are used for supervised machine learning when predictor variables are categorical. There has been an extensive number of classifier algorithms developed. Two common linear classifiers are logistic regression and linear discriminant analysis. Logistic regression is an algorithm that seeks to estimate the parameters $\hat{\beta}_i$ such that the logistic equation (below) predicts the probability that an observation with the measured values X_i belongs to a discrete group. The probability that an observation with M measured features

belongs to a specified group is given by:

$$\mathrm{p}(X_1, X_2, ...X_M) = \frac{exp(\beta_0 + \sum_{i=1}^{M} \beta_i X_i)}{1 + exp(\beta_0 + \sum_{i=1}^{M} \beta_i X_i)}$$

When $\mathrm{p}(X_1, X_2, ...X_M)$ exceeds a user-defined probability threshold, the observation is classified as belonging to the group.[51] Linear discriminant analysis operates in a similar manner but shows superior performance when there classifying involves more than two groups. Linear discriminant analysis estimates the parameters $\hat{\mu_k}$ and $\hat{\sigma}$ such that

$$\hat{\delta_k}(x) = x^T \hat{\Sigma}^{-1} \hat{\mu}_k - \frac{1}{2} \hat{u}_k^T \hat{\Sigma}^{-1} \hat{\mu}_k + log\ \hat{\pi}_k,$$

where

- $\hat{\pi}_k$ is the estimate for the prior probability that an observation belongs to group k
- $\hat{\mu}_k$ is the estimate for the mean value of feature X_i for all observations belonging to group k
- $\hat{\Sigma}$ is an estimate of the covariance matrix of feature vector X across all observations

The previous equation holds when all features are drawn from independent Guassian distributions. Linear discriminate analysis is based upon the Bayesian classifier. Each of the estimated parameters has an explicit formula that uses sample data in order to predict group membership for each observation in X. For each possible classification group, $\hat{\delta_k}(x)$ is calculated. Each observation is assigned to the group for which $\hat{\delta_k}(x)$ is greatest.[51]

Unsupervised machine learning examples - Dimensionality reduction and clustering

Dimensionality reduction algorithms seek to reduce the number of descriptive features required to describe observed events while preserving valuable information. A common linear dimensionality reduction method, principle component analysis seeks to find orthogonal basis of the covariance matrix of all measured features for a sample set. Each basis is comprised of linear combinations of measured values from the initial feature space. This procedure acts to

find new dimensions that preserve the variance from the higher-dimensional data set. Bases that describe minimal variance are considered uninformative and discarded.[51]

Clustering algorithms are used to identify groups of observations that exhibit common features. K-means clustering is a linear algorithm that seeks to find distinct clusters formed from similar observations. The number of clusters, k, must be decided *a priori*. The cost function can take one of several forms but often seeks to minimize the total squared distance between each observation and the center of mass for the cluster to which it has been assigned. Both principle component analysis and k-means clustering discover new information about previously measured observations by performing transformations on existing features.[51] Results obtained from unsupervised ML require careful interpretation by the user. Details about unsupervised machine learning algorithms used in this thesis are discussed in Chapter 4.

Evolutionary algorithms for supervised machine learning

Evolutionary algorithms (EA) describe a broad class of metaheuristic optimization procedures that includes evolutionary strategies, genetic algorithms, and evolutionary programming.[13] Initially, each distinct subclassification was developed independently. However, the procedural details for each EA category have since converged such that differences between each are subtle. To avoid unnecessary confusion, we will discuss the common traits shared by all EAs without mentioning minor differences between each implementation.

Within the context of supervised ML, EAs are use to optimize difficult regression problems. With this in mind, it is useful to draw parallels between linear regression and the EAs used in our research, where we seek to discover kinetic parameters that lead to a desired system response. The response is determined by evaluating kinetic parameters within a system of coupled ordinary differential equations (ODEs) describing a signal transduction network. To discover optimal parameter values, the EA seeks to minimize a user-defined fitness function. The fitness function measures the deviation between a desired system response and the response calculated from ODEs evaluated using a specific set of kinetic parameters. In comparison to linear regression, the coupled ODE system serves the same role as the linear predictor function (Eqn. 1.1). The kinetic parameters of the ODE are

optimized instead of the slope ($\hat{m}$) and y-intercept ($\hat{b}$) in linear regression. The cost function minimized in linear regression (Eqn. 1.2) is now replaced by the fitness function of our EAs. The primary difference between linear regression and EAs is the optimization method used to minimize the cost/fitness function.

In stark contrast to linear optimization procedures, EAs are stochastic optimization methods. That is to say, repeated optimizations of the same system with an EA may return different parameter values each time. This is a consequence of the heuristic nature of this optimization method. Unlike deterministic optimization algorithms, where specific actions are prescribed to every set of conditions, EAs have a degree of randomness during each step of the optimization procedure. EAs have the potential to exhibit increased computational efficiency at the risk of failing to find the globally optimum parameter set. However, EAs are useful for finding multiple local optima with equal fitness. Mimicking the biological principle of inherited fitness, EAs find optimal parameter values by applying crossover and mutation operators on previously evaluated parameter sets. The top-performing parameter sets are assigned to the parent population and new parameter sets are generated from those in the parent population. The crossover operator acts on two parameter sets selected from the parent population. From these parameter sets, a portion of one set is combined with a portion of the other to form a new parameter set. The mutation operator acts to randomly perturb the value of a parameter in a parameter set from the parent population or in one formed from crossover. The rates and precise details for performing crossover and mutation are specific to each EA implementation. After being generated, new parameter sets are evaluated using the fitness function and the parent population is updated. Repeated cycles of parameter set evaluation, parent population selection, and new parameter generation are repeated until termination criteria are met.

1.2.2 Machine learning in biology and soft matter

The development of machine learning algorithms has been intertwined with the growth of bioinformatics. The production of large volumes of experimental data from genome sequencing and microarray experiments has required specialized algorithms to distill usable knowledge. Clustering algorithms have long been used on gene expression data in order to

classify various malignant cells and their sources.[17] Supervised machine learning classifiers have also been used to analyze a diverse range of biological experiment results. This includes analyzing sub-cellular structures from fluorescent microscopy images, identifying and diagnosing cancers from gene expression data, classification of higher-order protein structures, and individual polymorphisms in specified genes from sequence data.[17]

In more recent years, researchers have applied machine learning algorithms to investigate protein folding mechanisms.[46] Simulations involving protein-folding utilize high-dimensional data to describe multi-body interactions with many degrees of freedom. The cooperativity occurring between various protein elements leads to the production of well-defined conformations produced from a limited number of assembly paths. Dimensionality reduction algorithms have proven useful for describing the time-evolution of these assembly paths in a smaller set of collective coordinates. Principle component analysis was an early method utilized for this application.[46] As the field of machine learning has advanced, nonlinear manifold folding techniques have been developed and applied to the analysis of protein folding. Diffusion mapping is a dimensionality reduction technique that utilizes pair-wise distances in high dimensions to estimate a density that is reproduced in lower dimensions. It has recently been applied to analyzing the folding trajectory of short antimicrobial peptides.[46] Diffusion mapping, and other advanced, nonlinear machine learning algorithms, continue to be applied in analysis of high-dimension and high-volume data and natural lend themselves to exploration of results produced from molecular dynamic simulations.

1.3 Outline of Thesis

We demonstrate the utility of applying data mining techniques to results obtained from computer simulations in three areas of computational biology. In each project, parameter exploration is used to generate data that examines the range of responses the system may exhibit. Various data exploration techniques are then used to develop insight into mechanisms and constraints that influence system response. Observations made are used to develop hypothesis that can be further tested and refined.

1.3.1 Exploring cancer growth and altered metabolism in heterogeneous environments (Chapter 2)

Cancer occurs as a consequence of healthy cells accumulating somatic mutations that lead to loss of homeostatic regulation. Cellular traits leading to this loss of regulation are well documented and succinctly enumerated in landmark reviews by Hanahan and Weinberg.[67, 68] Among these traits is the deregulation of cellular metabolism. Healthy cells rely on glycolysis and oxidative phosphorylation to generate ATP. Under conditions with low oxygen (hypoxia), healthy cells increase their reliance on glycolysis and produce lactate as metabolic byproduct. Cancer cells frequently exhibit a reliance on glycolysis even in the presence of sufficient oxygen. This effect is known as either aerobic glycolysis or the Warburg effect. This metabolism is approximately 18 times less efficient at producing ATP as compared to oxidative phosphorylation.[127] It is still not understood why cancer cells undergo the metabolic shift to aerobic glycolysis when it is less efficient at energy production. It has been hypothesized that it provides an advantage to rapidly proliferating malignant cells under conditions present within the tumor microenvironment.[88]

In vivo tumor progression occurs within a heterogeneous microenvironment. This heterogeneity takes the form of spatial variability of metabolite concentrations and cellular populations.[122, 68] *In vivo* and *in vitro* cancer models each present challenges that limit the ability to study the influence of altered metabolism. For *in vivo* systems, controlling environmental conditions is prohibitively difficult. For most *in vitro* systems, heterogeneous conditions are lost. The use of computer simulations allows systematic probing of the effects from altered malignant metabolism on tumor progression in heterogeneous environments with well-defined conditions.

Utilizing hybrid cellular automota (HCA) models to simulate cancer growth, we systematically explore tumor progression within vascularized tissue using computer simulations. We perform a systematic parameter sweep to explore the relative effects of varying cancer hallmark traits, including altered cellular metabolism. These simulations reveal four distinct growth patterns that correlate with specific malignant cell traits. Furthermore, the results suggest potential advantages conferred to malignant cells by the switch to aerobic glycolysis.

1.3.2 Analyzing signal transduction networks underlying ethylene growth response kinetics (Chapter 3)

Most models for ethylene signaling in the flowering plant *Arabidopsis* involve a linear pathway. However, measurements of seedling growth kinetics when ethylene is applied and removed have resulted in more complex network models that include coherent feedforward, negative feedback, and positive feedback motifs. However, the dynamical responses of the proposed networks have not been explored in a quantitative manner. In this chapter, we explore (i) whether any of the proposed models are capable of producing growth-response behaviors consistent with experimental observations and (ii) what mechanistic roles various parts of the network topologies play in ethylene signaling. To address this, we use computational methods to explore two general network topologies: The first contains a coherent feedforward loop that inhibits growth and a negative feedback from growth onto itself (CFF/NFB). In the second, ethylene promotes the cleavage of EIN2, with the product of the cleavage inhibiting growth and promoting the production of EIN2 through a positive feedback loop (PFB). Since few network parameters for ethylene signaling are known in detail, we use an EA to explore sets of parameters that produce behaviors similar to experimental growth response kinetics of both wildtype and mutant seedlings. We screen these parameter sets for growth recovery in the presence of sub-saturating ethylene doses, which is an experimentally-observed property that emerges in some of the evolved parameter sets. Finally, we probe simplified networks maintaining key features of the CFF/NFB and PFB topologies. From this, we verify observations drawn from the larger networks about mechanisms underlying ethylene signaling. Analysis of each network topology results in predictions about changes that occur in network components that can be experimentally tested to give insights into which, if either, network underlies ethylene responses.

1.3.3 Redesigning signal transduction network response dynamics *in silico* (Chapter 4)

The rational design of network behavior is a central goal of synthetic biology. In this chapter, we combine *in silico* evolution with nonlinear dimensionality reduction to redesign the

responses of fixed-topology signaling networks and to characterize sets of kinetic parameters that underlie various input-output relations. We first consider the earliest part of the T cell receptor (TCR) signaling network and demonstrate that it can produce a variety of input-output relations (quantified as the level of TCR phosphorylation as a function of the characteristic TCR binding time). We utilize an evolutionary algorithm to identify sets of kinetic parameters that give rise to: (i) sigmoidal responses with the activation threshold varied over 6 orders of magnitude, (ii) a graded response, and (iii) an inverted response in which short TCR binding times lead to activation. We also consider a network with both positive and negative feedback and use the EA to evolve oscillatory responses with different periods in response to a change in input. For each targeted input-output relation, we conduct many independent runs of the EA and use nonlinear dimensionality reduction to embed the resulting data for each network in two dimensions. We then partition the results into groups and characterize constraints placed on the parameters by the different targeted response curves. Our approach provides a way (i) to guide the design of kinetic parameters of fixed-topology networks to generate novel input-output relations and (ii) to constrain ranges of biological parameters using experimental data. In the cases considered, the network topologies exhibit significant flexibility in generating alternative responses, with distinct patterns of kinetic rates emerging for different targeted responses.

Chapter 2

Exploring Cancer Growth and Altered Metabolism in Heterogeneous Environments

2.1 Introduction

Cancer has long been understood as a disease that emerges due to the accumulation of somatic mutations. These mutations accumulate and confer hallmark cellular traits that allow an individual cell to escape healthy, homeostatic regulation. These hallmarks have been enumerated in landmark reviews by Hannahan and Weinberg and include proliferative growth signaling, escape from growth inhibition, evasion of apoptosis, inducing angiogenesis, and invasion/mestastasis.[67, 68] Many of these traits confer an obvious fitness advantage to malignant cells. However, altered metabolism is an emerging cancer hallmark that has drawn speculation as to whether it has a beneficial effect.[88]

Healthy cells in oxygenated conditions derive a minor fraction of cellular ATP from the glycolysis of glucose. A much greater proportion of cellular ATP produced from glucose occurs through oxidative phosphorylation. Utilizing oxidative phosphorylation requires oxygen as a terminal electron acceptor leading to the production of carbon dioxide and approximately 18 times more ATP than through glycolysis alone.[127] When oxygen

concentrations reach limited concentrations, healthy cells rely on glycolysis to produce ATP. The metabolic switch requires increased glucose consumption to compensate for decreased production of oxidative phosphorylation-derived ATP. Additionally, the terminal electron acceptor now becomes pyruvate which leads to the production of lactate as a byproduct. In comparison to healthy cells, cancer cells exhibit increased glycolytic metabolism, even in the presence of oxygen.[88] This metabolism, originally identified by Otto Warburg in the 1920s, has been termed aerobic glycolysis. Increased glucose uptake and lactate acid production rates produce conditions that negatively impact healthy cells in the tumor microenvironment. In order for malignant cells to proliferate under these conditions, it has been reasoned that malignant cells must derive a fitness advantage by switching to aerobic glycolysis.[71] The specific benefits conferred by the metabolic transition continue to be debated.[88]

The contribution to tumor progression from other cancer hallmarks is readily apparent. These beneficial effects include increased proliferation rates, reduced susceptibility to cell death, increased metabolite supply, and invasion of new tissue. Mutated signaling pathways leading to expression of these malignant cell traits frequently lead to traits associated with aerobic glycolysis. Mutations to cell cycle regulatory pathways frequently result in increased expression of the transcription factor HIF-1 (hypoxia inducing factor-1). HIF-1 leads to increased transcription of glycolytic enzymes and an associated increase in glucose consumption. In healthy cells, HIF-1 is expressed in response to low oxygen (hypoxic) conditions. Vascular endothelial growth factor (VEGF) is also expressed in response to elevated HIF-1 concentrations. VEGF expression leads to the formation of new vascular tissue, a process known as angiogenesis.[37] New vascular tissue provides an additional source of nutrients to actively proliferating malignant cell. Cancer mutations resulting in altered HIF-1 expression pathways are caused by alterations to healthy mTORC-1 activity.[130, 37] Part of the PI3K/Akt/mTOR pathway, mTORC-1 plays a critical role in cell cycle progression through the regulation of amino acid production.[37] Mutations increasing the activity of the PI3K/Akt/mTOR pathway are also commonly observed in cancer lines.[68] Beyond increasing glycolysis and glucose uptake, this pathway plays a role in proliferation, adhesion, migration, and invasion.[78, 128] Mutations to Ras, a GTPase in the MAPK pathway, have been shown to increase HIF-1 expression by activating the

PI3K/Akt/mTOR pathway.[130] Through its regulatory action on Rac and the MAPK pathway, mutations to Ras influence cell proliferation, motility, and invasion.[91] Furthmore, approximately 50% of tumors have mutations to p53. This mutation is associated with malignant resistance to apoptosis and increased rates of glycolysis.[37] Considering the number of common molecular pathways, the emergence of aerobic glycolysis could be a side effect of acquiring critical cancer hallmark traits.

An alternative hypothesis to explain the emergence of aerobic glycolysis suggests that a selective advantage is gained by malignant cells within the tumor microenvironment. The reliance on glycolysis observed in aerobic glycolysis leads to increased lactate secretion and decreased pH.[57, 88, 60] Reduced intracellular pH in the range of 7.3 to 6.5 has been shown to induce apoptosis in cultured neurons.[40] Malignant cells adapt to acidic conditions by exhibiting increased activity of various proton pumps including vacuolar H^+-ATPases, Na^+-H^+ antiporters, and H^+-linked monocarboxylate transporters.[57] In addition to demonstrating selective toxicity to healthy cells from acidic conditions, Gatenby *et al.* found evidence that reduced pH leads to extracellular matrix degradation that supports malignant migration.[56]

Physical conditions within heterogeneous tumor microenvironments may be necessary for malignant cells to benefit from aerobic glycolysis. Mixed cell populations and spatial variations in metabolite/growth factor concentrations are principle sources of heterogeneity in the tumor microenvironment.[122, 68] We use our *in silico* cancer model to simulate heterogeneous, microenvironment conditions that are difficult to produce or measure in other experimental settings. The model allows us to study the effects that occur from controlled variation of several malignant cell responses. The varied responses include producing acidic microenvironments, malignant migration, angiogenesis, and altered metabolism.

Computational models have been developed using different mathematical frameworks and include continuum, discrete, and hybrid models. Continuum models have been used to study the influence of hypoxic conditions on cell cycle progression,[2] hypoxic triggering of angiogenesis,[59] and the role mechanical stress has in tumor growth.[6] Continuum models are useful for describing systems of well-mixed cells and diffusing molecules but are limited in their ability to model heterogeneity. Discrete models represent cells as distinct entities

that interact with one another and external factors.[42] Simulating discrete cell responses can illuminate emergent tissue-level responses in a heterogeneous environment. Discrete computational cancer models have been used to study tumor spheroid growth patterning,[77] extracellular matrix influence on migration and invasion,[75] and the role of angiogenesis and vascular organization on tumor progression.[3]

Hybrid cellular automata (HCA) models are a common class of hybrid models utilized in cancer modeling. These models combine continuum and discrete approaches in studying tumor microenvironment heterogeneity. Common HCA models combine continuous descriptions of diffusing molecules with discrete cells. Molecules are typically ligands or metabolites that are consumed, produced, and/or elicits cellular responses based upon local concentrations. HCA models are well-suited for simulations examining emergent phenomena resulting from cellular- and molecular-scale events. HCA models have been used to study the influence of acidosis[118, 117], somatic mutation rate[55], interactions between cancer hallmarks,[113, 29] and chemotherapy/radiation treatment regimes.[116, 114]

HCA models have been valuable for studying the influence of the tumor microenvironment. However, assumptions made in previous models have shortcomings when exploring cancer metabolism within vascularized tissues. In this study we assess the effects of altered metabolism by developing an HCA model that expands on previous models. As many HCA models have done before, we simulate oxygen that diffuses from vascular sites that are homogeneously distributed among the simulated region.[114] In addition, we consider the effects of glucose consumption in a subset of models. We are unaware of HCA models that simulate the diffusion of glucose and oxygen within a vascularized field of cells. The inclusion of both metabolites is critical when analyzing the benefits of aerobic glycolysis. In healthy tissue, oxygen concentrations influence rates of cellular glucose consumption. Increased glycolytic flux in healthy cells is a response to low oxygen pressure during anaerobic glycolysis. Additionally, there are differences in diffusion constants and metabolic requirements for each metabolite.[98, 62] Lastly, our model accounts for space required by rapidly dividing malignant cells. If all available extracellular space is occupied by other cells, cellular rearrangement becomes necessary. Unlike previous HCA models, we allow rapidly dividing malignant cells to displace other cells. This can lead to rearrangement that alters the

microenvironment heterogeneity. Using these HCA modifications, we explore how aerobic glycolysis may potentially benefit cancer cells within the context of other cancer hallmarks.

2.2 Methods

2.2.1 Hybrid Cellular Automata model of cancer

We use a Hybrid Cellular Automata (HCA) framework to simulate tumor progression within the context of a heterogeneous microenvironment. The model is comprised of a two-dimensional lattice in which each site is occupied by a cellular automaton that responds to local metabolite concentrations and the identities of neighboring automata. Local metabolite concentrations are determined by solving associated reaction diffusion equations described below. The lattice size is 100 by 100 with each automaton exhibiting the characteristics of one of four basic site types: vascular, empty, healthy cell, or malignant cell. Parameter values were selected to reflect primary brain glioblastoma tumor progression.

Vascular sites act as sources for diffusing metabolites (glucose and oxygen). Vascular sites do not consume any metabolites and do not respond to local metabolite concentrations. However, these sites occupy physical space within the lattice and can be displaced by proliferating cells. Simulations are initialized with 25 regularly-spaced vascular sites with metabolite concentration at each lattice site equal to the local steady state concentration in a field of healthy cells. Steady state concentrations are calculated using a Jacobi iteration procedure.[120] The total number and arrangement of vascular sites is not held fixed and may change over time. If cellular division displaces vascular sites beyond the boundaries of the simulation lattice, the vascular site is removed and the total number decreases. The number of vascular sites can increase due to angiogenesis.

In addition to vascular sites, the lattice is initially occupied by one malignant cell. All remaining lattice sites contain healthy cells. Malignant and healthy cells have basal consumption rates for each metabolite and make cellular decisions based upon local metabolite concentrations and the identity of neighboring cells. Possible cell decisions include apoptosis, proliferation, migration, and angiogenesis. Conditions associated with

Table 2.1: HCA simulation parameters

Parameter	Meaning	Value (Original Units)	Non-dim. Value	Ref.
τ	Average cell cycle time	16 h	1	[114]
Δx	Average cell diameter (lattice site length)	50 μm	1	[114]
D_{O_2}	Oxygen diffusion coefficient	$10^{-5}\frac{cm^2}{s}$	2.304×10^4	[98, 114]
D_{Glc}	Glucose diffusion coefficient	$5.000 \times 10^{-6}\frac{cm^2}{s}$	1.152×10^4	[58, 62]
Δt	Metabolite update time step	0.2500 s	$230{,}400^{-1}$	-
K_{m,O_2}	Michaelis-Menten constant for oxygen autoregulation	0.8 mmHg	0.01	[114, 95]
V_{max,O_2}	Maximum oxygen consumption by healthy cells	$4.6 \times 10^{-16}\frac{mol}{cell \cdot s}$	51.54	[114, 50]
$K^{O_2}_{m,Glc}$	Michaelis-Menten constant for glucose inhibition by oxygen	8 mmHg	0.1	-
$K^{Glc}_{m,Glc}$	Michaelis-Menten constant for glucose autoregulation	940 nM	0.20	-
$V_{max,Glc}$	Maximum glucose consumption by healthy cells	$3.06 \times 10^{-15}\frac{mol}{cell \cdot s}$	300	-
σ_{prol}	Proliferative metabolite consumption scalar	NA	5	[114, 50]
σ_{O_2}	Malignant cell oxygen consumption scalar	NA	Varied	-
σ_{Glc}	Malignant cell glucose consumption scalar	NA	Varied	-
c_{max,O_2}	Vascular oxygen partial pressure	80 mmHg	1	[114, 30]
c_{p,O_2}	Oxygen proliferation threshold	8.8 mmHg	0.11	[114]
c_{ap,O_2}	Oxygen apoptosis threshold	8 mmHg	0.1	[114, 30]
$c_{max,Glc}$	Vascular glucose concentration	4.7 mM	1.0	[90]
$c_{p,Glc}$	Glucose proliferation threshold	1.41 mM	0.300	-
$c_{ap,Glc}$	Glucose apoptotis threshold	0.94 mM	0.20	-
p_H	Probability of healthy cell proliferation	NA	0.005	[114]
p_M	Probability of malignant cell proliferation	NA	0.01	[114]
p_{ap}	Probability of apoptosis when [Glc] or [O_2] falls below threshold values	NA	0.25	[114]
p_{tox}	Probability that healthy cell undergoes apoptosis when adajacent to one malignant cell	NA	Varied	-
p_{migr}	Probability that malignant cell swaps position with adjacent apoptotic cell	NA	Varied	-
p_{ang}	Probability that a healthy cell with local oxygen below c_{ap,O_2} or any malignant cell leads to formation of new vascular site	NA	Varied	-

each decision are different for healthy and malignant cells and are designed to mimic cancer.[68] Healthy and malignant cells that undergo apoptotis result in empty sites. Empty sites do not consume metabolites and are available for placement of new cells.

Layered on top of the cellular automata model are fields of diffusing metabolites. Updates to metabolite concentration occur at a much greater frequency than cellular response decisions and are described in detail below.

Calculating metabolite concentration

Oxygen and glucose are supplied by vascular sites and act as diffusing metabolites. Local concentrations are updated using a finite difference approximations of the modified reaction-diffusion partial differential equations shown below. Numerical approximations utilize time-forward and central difference approximations with no flux (Neumann) boundary conditions. Metabolite concentrations are updated 2.304×10^5 times per cell cycle to satisfy stability

criterion based on oxygen diffusion. Oxygen concentration, c_{O_2}, follows:

$$\frac{\partial c_{O_2}}{\partial t} = D_{O_2}\nabla^2 c_{O_2} - \sigma_{O_2}\, V_{max,O_2}\left(\frac{c_{O_2}}{K_{m,O_2} + c_{O_2}}\right)$$

Glucose concentration, c_{Glc}, follows:

$$\frac{\partial c_{Glc}}{\partial t} = D_{Glc}\nabla^2 c_{Glc} - \sigma_{Glc}\, V_{max,Glc}\left(\frac{K^{O_2}_{m,Glc}}{K^{O_2}_{m,Glc} + c_{O_2}}\right)\left(\frac{c_{Glc}}{K^{Glc}_{m,Glc} + c_{Glc}}\right)$$

The cell metabolite scalars (σ_{O_2} and σ_{Glc}) are functions of position, with values specific to the cell type at each lattice site. The values is 0 for empty sites, 1 for healthy cells, and varied for malignant cells. Varying these parameters for malignant cells allow us to examine the influence of altered metabolism on tumor progression. Vascular sites act as sources of metabolites with fixed concentrations of oxygen and glucose. The consumption of metabolites by healthy and malignant cells follows Michaelis-Menten kinetics with increased local concentration leading to increased consumption rates.[95] Additionally, oxygen has an inhibitory effect on glucose consumption that mimics a switch to anaerobic glycolysis as observed in healthy cells experiencing hypoxic stress. As oxygen concentration decreases, cells consume more glucose and less oxygen. Parameters were chosen based on experimentally measured values when available (see Table 2.1).

Parameter values without experimentally measured values were estimated or varied within a systematic parameter sweep. Parameter values that were estimated include glucose consumption parameters and glucose concentration threshold values for proliferation and apoptosis. To estimate glucose consumption parameters, we performed a parameter search in which different values for maximum glucose consumption ($V_{max,Glc}$), and different values of Michaelis-Menten constants for glucose comsumption ($K^{Glc}_{m,Glc}$ and $K^{O_2}_{m,Glc}$) were evaluated. For each set of tested parameters, the steady state concentrations of oxygen and glucose were determined in field of healthy cells with 25 vascular sites. Then, using the same metabolite consumption parameters, the local glucose and oxygen concentrations were used to determine individual cell glucose consumption. The calculated mean glucose concentration and consumption rate were compared to experimentally measured values in brain tissue (1.8

mM [63] and 0.08-0.51 $\mu\text{mol} \cdot \text{g}^{-1} \cdot \text{min}^{-1}$).[129] The estimated parameter values used in the simulation are reported in Table 2.1. With these values, the simulation with healthy cells and 25 vascular sites produces a steady-state mean glucose concentration of 1.96 mM and a mean glucose consumption rate of 0.155 $\mu\text{mol} \cdot \text{g}^{-1} \cdot \text{min}^{-1}$.

Cellular response rules

Healthy and malignant cells undergo cellular processes including metabolite consumption, proliferation, migration, apoptosis, and angiogenesis. Cellular responses in the HCA simulation are governed by the logic outlined in Fig. 2.1. Cellular decisions for all healthy and malignant cell are made 100 times per cell cycle (τ = 16h). To prevent any ordering bias, the sequence in which cell responses are evaluated is randomized each decision cycle. The conditions leading to each decision reflect cell behaviors associated with specific cancer hallmarks. The response rules and cancer biology are outlined in this section.

Apoptosis Cells that are sufficiently stressed may initiate the cellular death program known as apoptosis. Within our simulations, apoptosis occurs when healthy and malignant cells encounter reduced availability of critical metabolites or when local toxicity causes healthy cell apoptosis when neighboring malignant cells. When local concentrations of oxygen or glucose fall below the associated apoptosis concentration threshold, c_{ap,O_2} and $c_{ap,Glc}$, the cell enters apoptosis with probability p_{ap}.

Unregulated cellular division observed in tumors can produce toxic conditions including acidosis. Malignant cells exhibit mutations to evade apoptosis triggering under these unfavorable conditions. Healthy cells do not exhibit such apoptosis evasion mechanism. Our model captures this difference in apoptosis triggering by including local toxicity effect.

Local toxicity effect. Within the tumor microenvironment, cells can be subjected to additional stresses that lead to apoptosis. Uncontrolled cellular division can result in increased pressure on a cell as the number of cells increase within a restricted volume. Aerobic glycolysis from malignant cells leads to decreased pH from increased lactic acid production.[88] It has been observed that malignant cells can adjust to these harsher conditions with such adaptations as increased expression of transmembrane proton

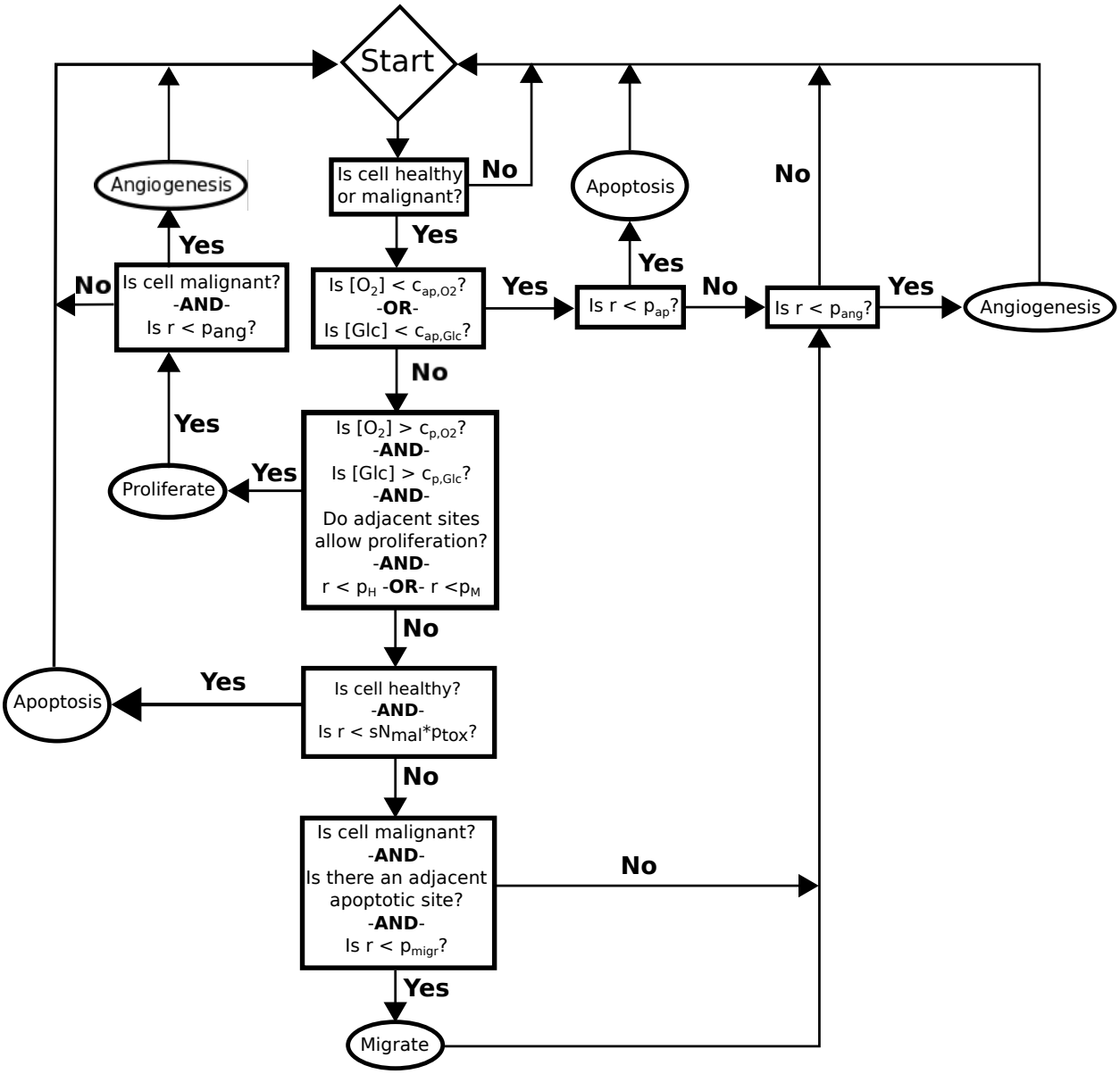


Figure 2.1: HCA cell decision tree. Cells undergo cell fate decisions in response to local metabolite concentration and neighboring cell types. Parameters in the cellular decision tree are explained in Table 2.1. Exceptions include N_{mal}, which is the number of nearest neighbor cells that are malignant, and r, which is a unit random value drawn from a uniform distribution.

pumps.[71] Healthy cells do not have such adaptations and are more susceptible to apoptosis. While some HCA models include an explicit proton concentration to partially account for these effects [118, 55, 116, 58], we generalize these effects into a single parameter p_{tox} that accounts for microenvironment stressors that provide a fitness advantage to malignant cells within a toxic tumor microenvironment. This reduces the computational expense associated with simulating a third field of diffusing molecules. A contact toxicity effect is used as an approximation to the more complex conditions occurring within the microenvironment. Operating under the assumptions that toxic conditions are produced from unregulated growth of malignant cells and that toxicity increases in proportion to the local density of malignant cells, the probability that a healthy cell undergoes apoptosis from toxic effects is equal to the product of the probability of apoptosis from local toxicity, p_{tox}, and the number of neighboring malignant cells, N_{mal}. This contact-based effect is similar to contact degradation of the extracellular matrix associated with glycolytic phenotype utilized in another HCA model. [58]

Proliferation Healthy and malignant cells supplied with sufficient concentrations of metabolites can undergo proliferation. In the simulations, local glucose and oxygen concentrations must exceed threshold values of $c_{p,Glc}$ and c_{p,O_2} for proliferation to occur. Additionally, healthy cells are subject to contact inhibition, which is a form of growth suppression that stops healthy cells from dividing when in physical contact with other cells. Mutations in cadherin-E have been observed in malignant cells and are linked to loss of contact inhibition.[68] Within our simulations, we capture this feature by requiring that healthy cells have an empty neighboring site in order to proliferate. In contrast, malignant cells can also proliferate when adjacent sites are occupied by healthy cells. When the previous conditions are met, there is a probability that a healthy cell (p_H) or malignant cell (p_M) will produce an identical daughter cell. For healthy cell proliferation, the location of daughter cell placement is chosen at random from neighboring empty sites. For malignant cell proliferation, daughter cell placement is randomly chosen from adjacent sites that are either empty or healthy. When the site selected for daughter cell placement is occupied by a healthy cell, a conflict is produced as each lattice site can be occupied by only one cell. In

previous models, this conflict was resolved by having the malignant daughter cell replace the the healthy cell.[114] In our model, the malignant daughter cell displaces the healthy cell by a distance of once lattice site. Other cells within the same row/column are also displaced. Fig. 2.2 depicts the consequences of each choice. Cell displacement can lead to displacement of vascular sites that is not observed when the replacement assumption is used. Changes in the configuration of vascular sites due to displacement can lead to vastly different patterns of tumor progression.

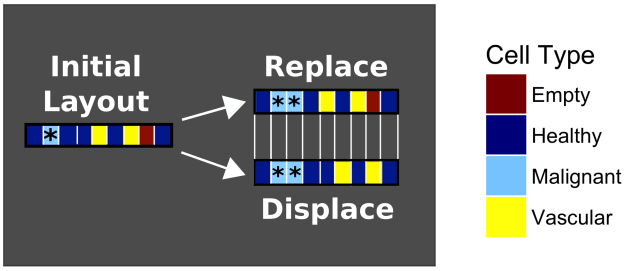


Figure 2.2: Cellular replacement vs displacement upon proliferation. Following proliferation of a malignant cell, daughter cell placement requires an empty site to occupy. This can be produced by removing and replacing the adjacent healthy cell or displacing all cells in the same row until an empty site is encountered.

Estimating glucose threshold concentrations. The glucose concentration thresholds utilized in our model, $c_{ap,Glc}$ and $c_{p,Glc}$, are below values used in similar models [58]. Mean glucose concentration in brain tissue is below $c_{ap,Glc}$ value used in these models. Since metabolite concentration thresholds are tissue specific and other models were simulated in different tissues, we adjust the theshold values used. By decreasing the value of $c_{ap,Glc}$ and $c_{p,Glc}$, we adapt the apoptotic and proliferative glucose thresholds to reflect conditions present in brain tissue.

Migration Migration of malignant cells is influenced by many properties of the tumor microenvironment. Of particular interest is increased migration associated with extracellular matrix remodeling that results from reduced pH.[88] To explore the effects of cellular migration on tumor progression, we include an additional parameter for the probability that a malignant cell will move to an adjacent empty site, p_{migr}. Malignant cells that do not undergo apoptosis or proliferation move to a neighboring empty site with probability p_{migr}.

The previous position occupied by the migrating malignant cell becomes an empty site. The value of p_{migr} is systematically varied throughout the simulations.

Angiogenesis In our simulations, angiogenesis leads to the placement of new vascular sites. Healthy cells in hypoxic conditions and many cancer cells, especially those exhibiting aerobic glycolysis, secrete vascular endothelial growth factor (VEGF), which is a growth factor that stimulates angiogenesis. The parameter p_{ang} sets the probability that a healthy cell with local oxygen concentration in the hypoxic range or malignant cells under any oxygen condition will lead to the formation of a new vascular site. All cells satisfying these conditions are checked for angiogenesis following all other cellular decisions. When a new vascular site is created, it is placed in a site neighboring the cell. Cells are displaced the same as in proliferation. The probability of angiogenesis is systematically varied to determine the level of impact on overall tumor progression. Many of our simulations were performed with $p_{ang} = 0$.

2.2.2 Analyzing simulation results

In our analysis, we often refer to "malignant density", which is the fraction of sites occupied by malignant cells. We also consider the measures described below.

Measuring vascular displacement

We characterize vascular displacement as the Euclidian distance between the initial position and its position at some later time point. When reporting mean displacement for a simulation at a given time point, vascular sites displaced beyond the boundary of the lattice are excluded from the calculation. This reduces the reported mean vascular displacement. However, for simulations that include glucose consumption, loss of vascular sites is rare and this minimally effects reported values.

Determining proliferation domains

Proliferation domains are regions in which lattice sites are identified as having sufficient metabolite concentrations to support proliferation. For lattice sites above proliferative thresholds of glucose and oxygen ($c_{p,Glc}$ and c_{p,O_2}), the sites are designated as "Proliferative".

For lattice sites below apoptotic thresholds of glucose or oxygen ($c_{ap,Glc}$ or c_{ap,O_2}), the sites are designated as "Apoptotic". All other sites are designated "Quiescent". Proliferation domains are then identified by finding clusters of adjacent "Proliferative" sites using the "clump()" function from the "raster" package in R.[69] This function assigns unique cluster numbers to distinct proliferation domains to allow for tracking individual domain progression.

Measuring compactness

The compactness of a two-dimensional shape is defined as a measure comparing the squared perimeter to the area.[26] Discrete compactness calculates the compactness of objects comprised of pixels. For our system, each lattice site serves as a pixel. Discrete compactness (C_d) is given by:

$$C_d = \frac{n - P/4}{n - \sqrt{n}}$$

where n is the total number of malignant cells in the entire lattice and P is the total number of malignant cell edges that are not adjacent to another malignant cell edge. That is, for a malignant cell without neighboring malignant cells, the contribution to P is 4. A malignant cell with just one neighboring malignant cell contribultes 3 while a malignant cell with 4 neighboring malignant cells contributes 0. Tumors exhibit values close to 1 when they are compact, i.e. they are no gaps between malignant cells and tumors have regular edges. Decreasing discrete compactness values indicate increased dispersion and/or irregular tumor borders.

2.3 Results

The HCA cancer simulations utilized parameters based on experimental values measured in brain tissue. For parameters describing malignant cell traits, namely p_{tox}, p_{migr}, σ_{Glc}, and σ_{O_2}, a broad parameter sweep was performed. All simulations included oxygen as a consumable metabolite. A subset of simulations also included glucose. For simulations in which only oxygen was included, cellular decisions were made according to the cellular

decision graph (Fig. 2.1) assuming glucose exceeded the proliferative concentration threshold at every lattice site.

In all, 288 parameter combinations were screened in simulations in which cells consumed glucose and oxygen. Another 96 parameter combinations were screened in simulations in which only oxygen was included. 10 trajectories were simulated under each set of conditions. The resulting simulations revealed four distinct patterns of tumor progression. These growth patterns are shown in Fig. 2.3 and are termed isolated clustering, continuous growth, tumor trapping, and domain seeding. Each growth pattern shows a distinctive distribution of cells and metabolites as the simulation progresses.

In Fig. 2.3, snapshots showing cell progression are displayed on the left. To the right are corresponding snapshots that show regions where metabolite concentrations give rise to proliferation, quiescence, and apoptosis. In many of the metabolite zone maps, regions that support proliferation (green), which we call "proliferation domains", are surrounded by regions of cell quiescence (yellow). Each growth pattern exhibits a distinct pattern of proliferating malignant cells and distinct patterns of proliferation domains. Growth pattern characteristics are discussed below.

2.3.1 Tumor growth patterns

Isolated clustering

As malignant cells divide, daughter cells are placed in adjacent lattice sites. When an empty site is available, surrounding cells are left undisturbed. However, when daughter cell placement occurs in a lattice site that is occupied, surrounding cells are displaced as described in the Methods section. This displacement can lead to a rearrangement of vascular sites. In extreme cases, vascular sites can be displaced beyond the lattice boundaries resulting in reduced vascular site density. Isolated clustering occurs when malignant cell growth results in reduced vascular site density and a reduction in the number of metabolite sources within the system. The overall reduction in metabolites leads to an overall shift towards malignant apoptosis and empty site formation. This leads to disparate domains of malignant cells

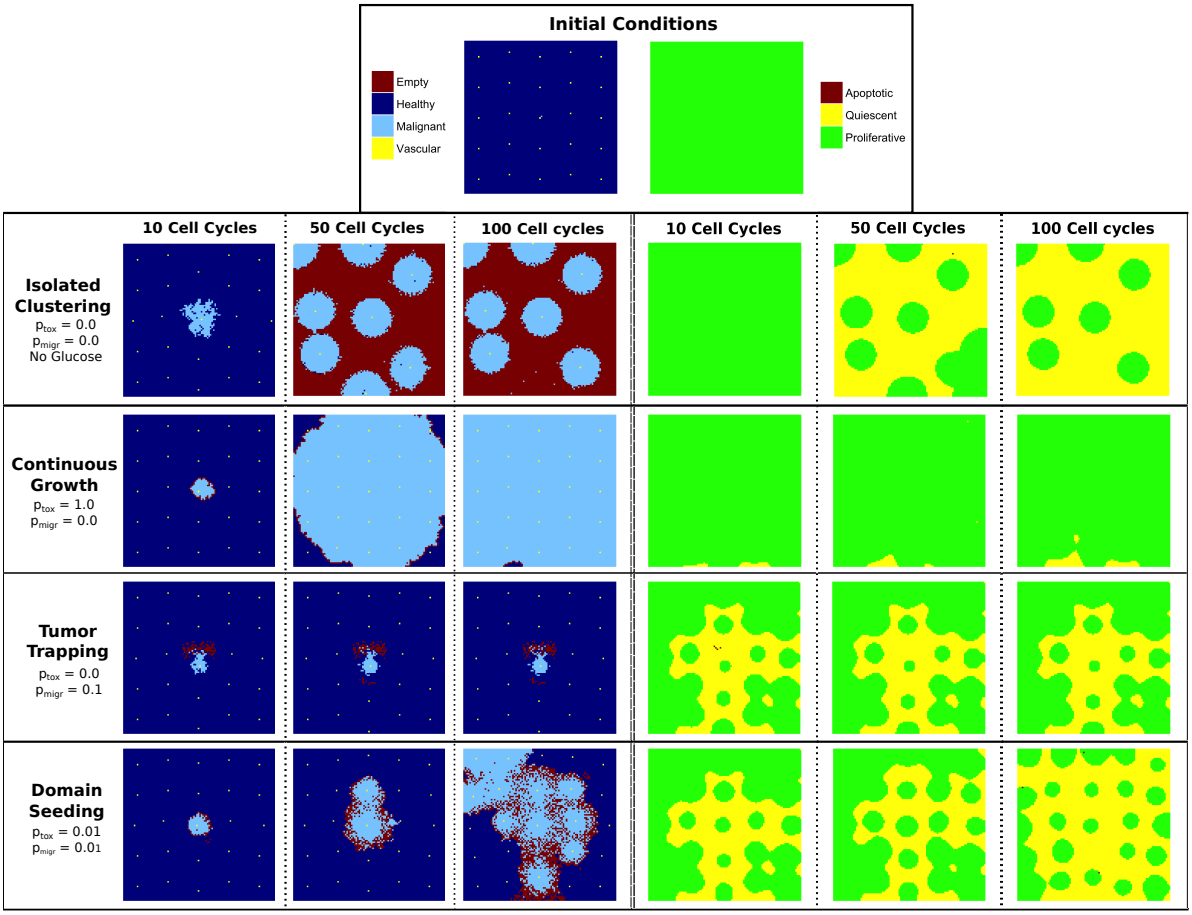


Figure 2.3: Emergent growth patterns in cancer simulations. Snapshots from simulations with each characteristic growth pattern are shown. On the left are snapshots showing the spatial distributions of cells for trajectories with different parameters that start from identical initial conditions (at top). On the right are spatial maps indicating the cellular proliferation decisions possible given the concentration of metabolites.

centered around the remaining vascular sites. Remaining vascular sites are located at the center of each proliferation domain.

Continuous growth

Continuous growth describes simulations in which malignant cells display rapid, unhindered growth until they surround all vascular sites. In these simulations, the majority of healthy cells undergo apoptosis, leaving empty sites available for placing daughter cells of proliferating malignant cells. Apoptosis of healthy cells and proliferation of malignant cells occur at rates that allow malignant cells to reach adjacent proliferation domains without slowing malignant growth rates or significantly displacing vascular sites. Under certain conditions, continuous growth occurs within a lattice that consists of a single, contiguous proliferation domain.

Tumor trapping

Proliferation of malignant cells surrounded by healthy cells can lead to displacement of vascular sites. In certain parameter regimes, this results in the formation of proliferation domains that trap malignant cells, with tumor progression limited to small regions surrounding a single vascular site.

Domain seeding

Domain seeding dynamics are associated with malignant densities that exceed levels observed in tumor trapping. Compared to continuous growth, malignant density increases more slowly in this regime. As domain seeding occurs, increasing malignant density leads to vascular site displacement and formation of proliferation domains. However, malignant cells do not remain confined to a single proliferation domain. Instead, they eventually reach new proliferation domains.

2.3.2 The influence of glucose on tumor progression

For our simulations, we selected parameters to describe glucose and oxygen consumption dynamics in glial cells. Consumption rates relative to supplied concentrations from vascular tissue are similar for glucose and oxygen within glial cells. However, the diffusion constant of glucose in tissue is approximately half the diffusion constant of oxygen. Additionally, the threshold concentrations promoting proliferation and apoptosis are much higher for glucose (see Table 2.1). These differences result in tumor growth patterns that depend on whether glucose consumption is included in the HCA model.

Fig. 2.4 shows the density of malignant cells for simulations without glucose consumption and Fig. 2.5 shows analogous results for simulations with glucose consumption. Two malignant cell metabolite consumption scalar combinations (σ_{O_2} and σ_{Glc}) are shown of each. Within each subplot, 10 trajectories for p_{tox} and p_{migr} are plotted together. These are the probability of local toxicity causing healthy cell apoptosis and the probability of migration by malignant cells, respectively. When comparing simulations with identical parameters, increased malignant density is observed in simulations without glucose.

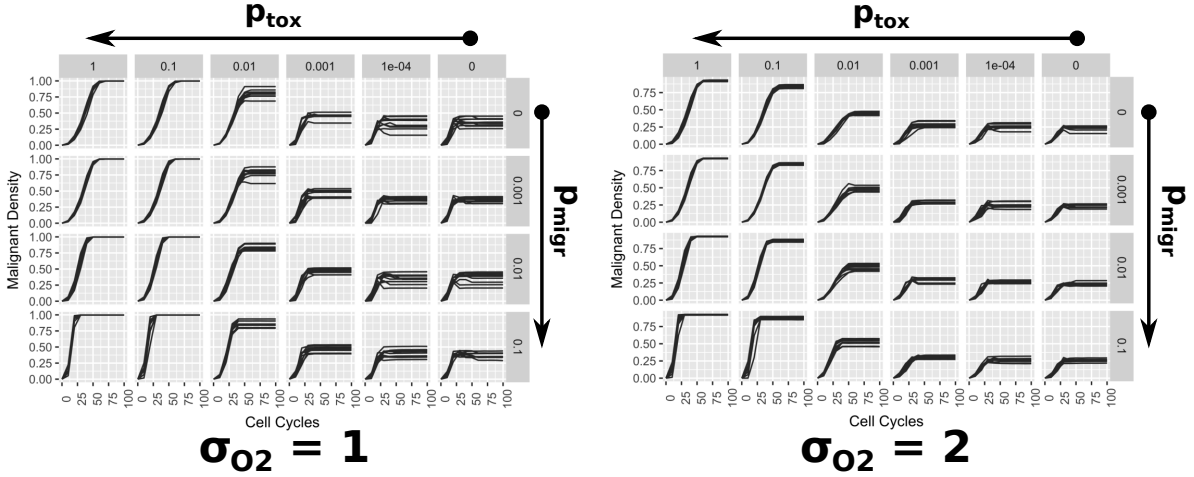


Figure 2.4: Malignant density progression without glucose consumption. The time progression of malignant density for simulations without glucose consumption and two values of malignant oxygen consumption scalar (σ_{O_2}). Local toxicity effect (p_{tox}) increases from right to left. Migration (p_{migr}) increases from top to bottom. The probability of angiogenesis for results shown is 0 ($p_{ang} = 0$). The results for 10 trajectories are shown for each parameter combination.

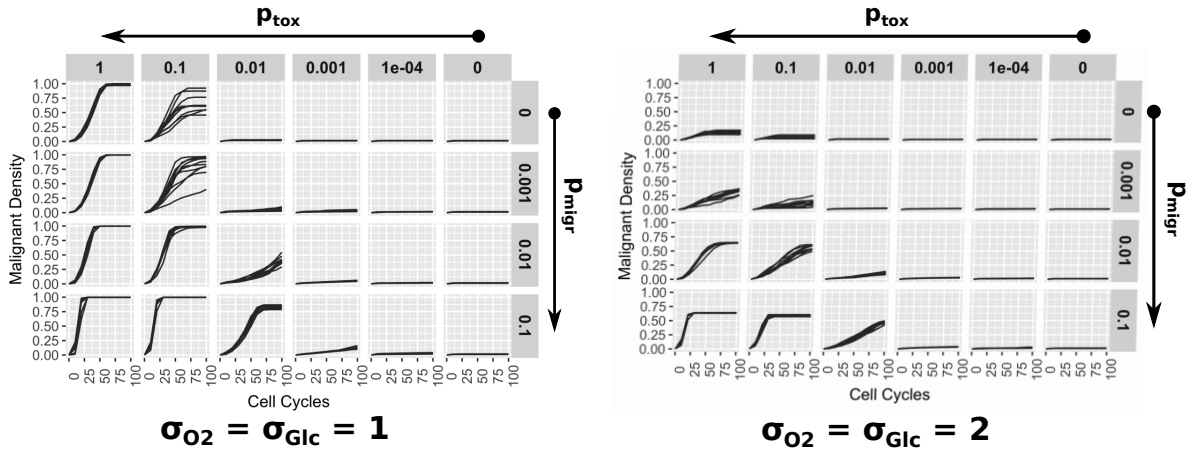


Figure 2.5: Malignant density progression with glucose consumption. The time progression of malignant density for simulations with $\sigma_{O_2} = \sigma_{Glc} = 1$ (left) and $\sigma_{O_2} = \sigma_{Glc} = 2$ (right). Local toxicity effect (p_{tox}) increases from right to left. Migration (p_{migr}) increases from top to bottom. The probability of angiogenesis for results shown is 0 ($p_{ang} = 0$). The results for 10 trajectories are shown for each parameter combination.

In addition to differences in malignant density, the inclusion of glucose consumption has a significant effect on vascular site density. For simulations with $\sigma_{O_2} = 1$ (and $\sigma_{Glc} = 1$ for simulations with glucose consumption), Fig. 2.6 highlights that reduced vascular density is observed in the absence of glucose consumption. Exceptions to this arise when local toxicity effects are elevated. In contrast, simulations with glucose consumption rarely lead to a decrease in the number of vascular sites as a result of displacement.

A snapshot from final simulation configurations can be used to compare tumor progression across all varied parameters. Fig. 2.7 shows mean malignant and vascular site densities at 100 cell cycles for all screened parameter combinations without angiogenesis. These results reinforce the observation that inclusion of glucose consumption leads to reduced malignant densities and less vascular site loss from displacement. When comparing malignant cell density between simulations with and without glucose consumption, we see similar trends in response to local toxicity effect (p_{tox}), migration (p_{migr}), and metabolite consumption scalars (σ_{O_2}, σ_{Glc}). Below is detailed discussion of each effect within the context of simulations that include glucose consumption and no angiogenesis.

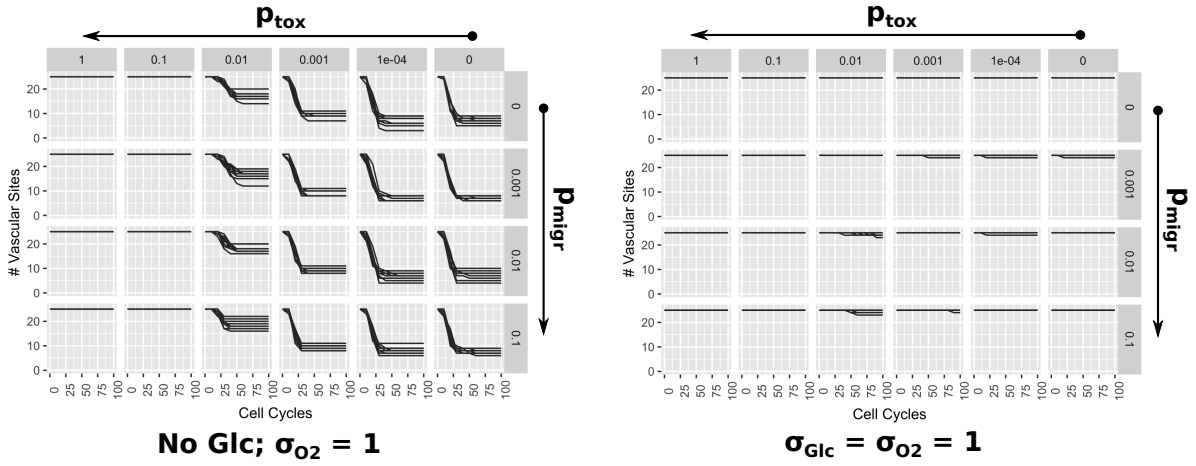


Figure 2.6: Progression of the number of vascular sites without angiogenesis. The time progression of vascular density for multiple parameter combinations comparing simulations with (right) and without (left) glucose consumption. The results for 10 trajectories are shown for each parameter combination. Vascular site count decreases when displacement from cell proliferation pushes vascular sites beyond the boundary of the lattice. The probability of angiogenesis for results shown is 0 ($p_{ang} = 0$).

2.3.3 Varying cancer hallmarks in HCA simulations

Among simulations that including glucose consumption, only three growth patterns are observed. Isolated clustering is absent from these simulations. Simulations without glucose consumption exhibit isolated clustering but do not exhibit tumor trapping. Despite these differences, the influence of the parameters that control certain malignant cell traits generally apply to both modeling regimes. With this in mind, we focus on analyzing the results produced from parameter sweeps for simulations including glucose consumption. From this data, we identify tumor progression trends produced from varying the local toxicity effect (p_{tox}), migration (p_{migr}), and metabolite consumption scalars (σ_{O_2}, σ_{Glc}).

The influence of local toxicity

The magnitude of local toxicity effects is controlled by the parameter p_{tox}. Increasing its value promotes the death of healthy cells next to malignant cells and increases the likelihood of malignant cells being adjacent to empty sites when they divide. Simulations with low values of p_{tox} (< 0.001) have a low malignant density (< 0.05) after 100 cell cycles (Fig. 2.7). While the values of p_{migr}, σ_{O_2}, and σ_{Glc} all influence malignant density, Fig. 2.7 shows a

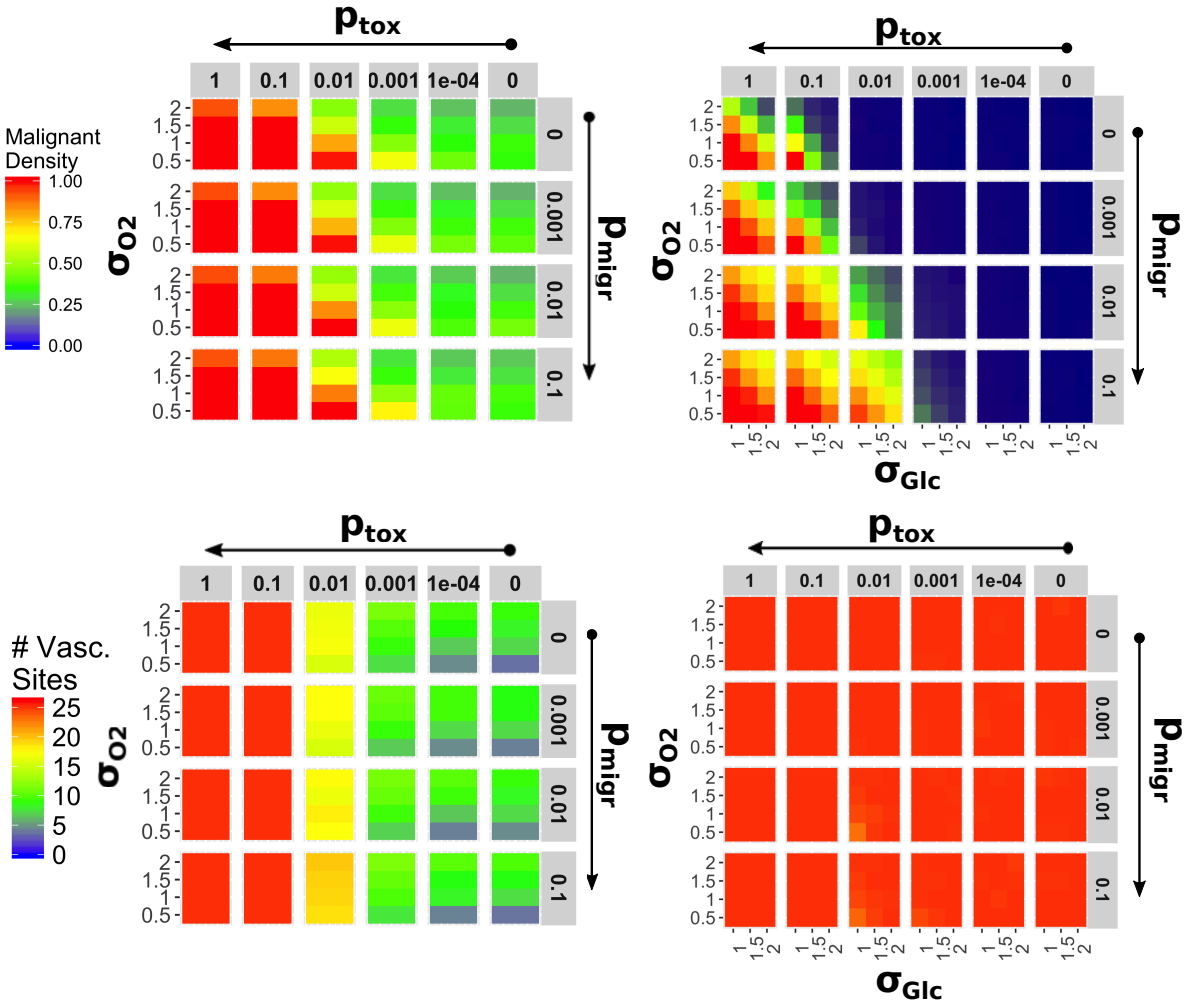


Figure 2.7: Mean density of malignant cells and mean number of vascular sites. The malignant cell density (top) and vascular site count (bottom) are shown for each parameter combination. The mean value for 10 trajectories is reported for each. Simulations without glucose consumption are shown on the left; simulations including glucose consumption are shown on the right. The probability of angiogenesis for results shown is 0 ($p_{ang} = 0$).

positive correlation between local toxicity effects and malignant density at intermediate to high values of p_{tox} (≥ 0.001).

The presence of empty sites adjacent to dividing cells reduces the extent of cellular displacement during a simulation. Since increased local toxicity leads to increased apoptosis of healthy cells, it enhances the presence of empty sites and decreases cell displacement. Fig. 2.8 shows the mean vascular site displacement averaged across the 10 trajectories for each parameter combination. At low p_{tox}, vascular displacement is relatively low due to minimal proliferation of malignant cells. Vascular displacement is at an absolute minimum at high values of p_{tox}. However, at some intermediate values of p_{tox}, there is a large increase in vascular displacement.

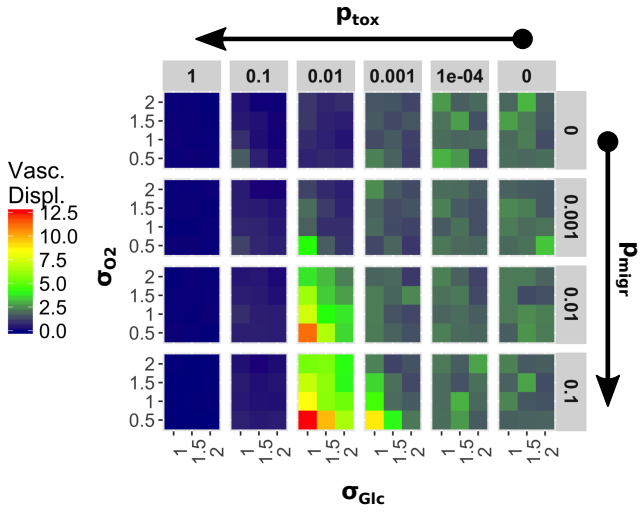


Figure 2.8: Vascular site displacement. The mean Euclidian distance (reported in units of lattice site length) for each vascular site from its initial position are calculated as described in Methods. Simulations include glucose consumption and are evaluated at 100 cell cycles. The mean displacement is averaged across the 10 trajectories for each parameter combination. The probability of angiogenesis for results shown is 0 ($p_{ang} = 0$).

Simulations reveal a correlation between vascular site displacement and the total proliferation domains at intermediate and low values p_{tox} (< 0.01). This is evident by comparing data from Fig. 2.8 and Fig. 2.9. In contrast, when $p_{tox} \geq 0.01$, vascular site displacement is at a minimum while proliferation domains formation is at a maximum. Exceptions to this occur at low values of metabolite consumption scalars (σ_{Glc} and σ_{O_2}), where the lowest number of proliferation domains are observed. In these cases, low

metabolite consumption rates limit the number of cells that experience concentrations below proliferation thresholds. When these metabolite consumption scalar values are excluded, simulations with $p_{tox} \geq 0.01$ show strong correlation between malignant density (Fig. 2.7) and proliferation domain formation (Fig. 2.9).

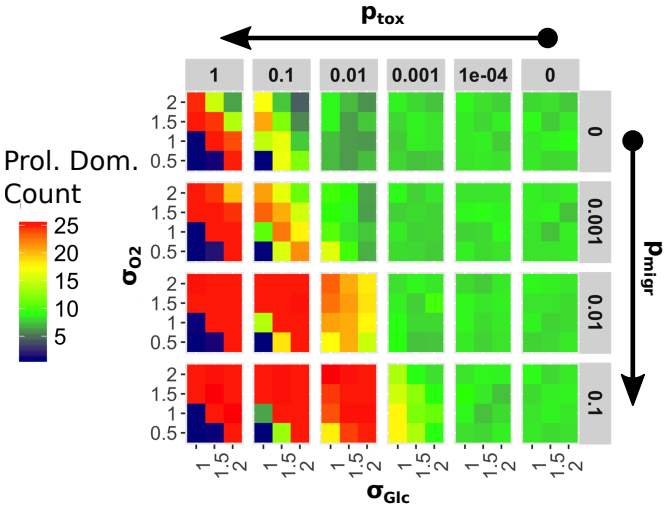


Figure 2.9: Proliferation domain count. Proliferation domains were identified after simulating 100 cell cycles in simulations including glucose consumption. The mean count is averaged across 10 trajectories for each parameter combination. The probability of angiogenesis for results shown is 0 ($p_{ang} = 0$).

The influence of migration

Migration allows malignant cells to move into adjacent empty sites. When simulations do not include angiogenesis, this is the only means by which malignant cells can traverse quiescent regions. To reach new proliferation domains, Fig. 2.7 shows that increased migration of malignant cells (p_{migr}) leads to increased malignant densities at intermediate to high levels of local toxicity. The influence of migration on vascular displacement is marginal, except at intermediate values of p_{tox}. At these values, Fig. 2.8 shows a positive correlation between vascular displacement and p_{migr}. Migrating malignant cells have an increased probability of contacting healthy cells. Consequently, there is an increase in healthy cells that undergo apoptosis from local toxicity effects leading to additional empty sites. The availability of empty sites reduces the total cellular displacement that occurs in response to proliferation in

a crowded region. A positive correlation between migration and proliferation domain count is weak and only present at intermediate to high values of local toxicity effect (Fig. 2.9).

Migration allows malignant cells to detach from clusters of malignant cells. At increased values of p_{migr}, malignant cells can be observed diffusing between empty sites in quiescent regions near the primary malignant cell mass. This phenomenon is observed in all growth regimes and can be quantified using discrete compactness. Discrete compactness is used to compare the total perimeter contribution of malignant cells to their total area. Values close to 1 indicate compact structures (minimum perimeter to area) while values near zero indicate disperse malignant cells. Fig. 2.10 shows discrete compactness scores for final configurations of simulations with glucose consumption. In general, increased migration leads to lower discrete compactness due to neighborless malignant cells that contribute the maximum perimeter to area ratio possible for a pixel.

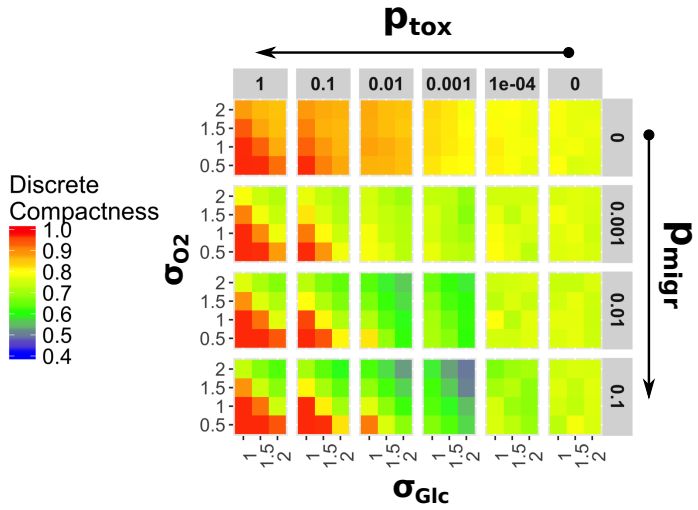


Figure 2.10: Discrete compactness of malignant cells. The discrete compactness of tumors measured using malignant cells to determine tumor shape and area after 100 cell cycles. The heat map visualizes the mean compactness averaged across the 10 trajectories performed for simulations that include glucose consumption. The probability of angiogenesis for results shown is 0 ($p_{ang} = 0$).

The influence of altered metabolism

Increasing the metabolic consumption scalars (σ_{O_2} and σ_{Glc}) leads to lower malignant densities. This is illustrated in Fig. 2.7. Exceptions to this occur for parameter combinations

subject to tumor trapping. As malignant cells divide and proliferation domains form, the radii of proliferation domains are influenced by the amount of metabolites being consumed. When either metabolite concentration falls below its respective proliferation threshold (c_{p,O_2} and $c_{p,Glc}$), a lattice site can no longer support proliferation. Comparison of glucose and oxygen reveal that quiescent regions form because glucose, but not oxygen, falls below its proliferation threshold. By increasing σ_{Glc}, more glucose is consumed and the penetration depth of glucose is decreased. This directly leads to smaller proliferation domains.

Varying σ_{O_2} indirectly influences proliferation domain size. Oxygen acts as an inhibitor of glucose consumption leading to increased glucose consumption under hypoxic conditions. This mimics normal, healthy anaerobic glycolysis that is observed *in vivo*. By increasing σ_{O_2}, additional oxygen is consumed by malignant cells, decreasing its concentration. This reduces the inhibitory effect on glucose consumption. Glucose consumption increases, leading to additional quiescent regions and smaller proliferation domains.

2.3.4 Effects of angiogenesis

In our HCA model, angiogenesis leads to the formation of new vascular sites. In comparison to cell division, angiogenesis occurs on a longer time scale. The formation of new vascular sites occurs infrequently and requires longer simulation times to assess. Thus, we performed longer simulations for a reduced set of parameter combinations to study the effects of angiogenesis. Fig. 2.11 compares moderate levels of angiogenesis ($p_{ang} = 10^{-7}$) to no angiogenesis ($p_{ang} = 0$) for simulations with $\sigma_{Glc} = \sigma_{O_2} = 1$. Angiogenesis leads to higher malignant density by 300 cell cycles in some simulations. Higher variability is also observed in simulations with angiogenesis.

Increased malignant density can be partially attributed to the addition of new metabolite sources from new vascular sites. However, we also observe "bridging events" which are depicted in Fig. 2.12. When a new vascular site is formed near the outer edge of a cluster of malignant cells in a quiescent region, the additional nutrients may form new proliferation zones that form bridges between pre-existing proliferation domains. As malignant cell proliferation proceeds, vascular site displacement and the additional metabolite consumption from new malignant cells may result in previously separate proliferation domains reforming.

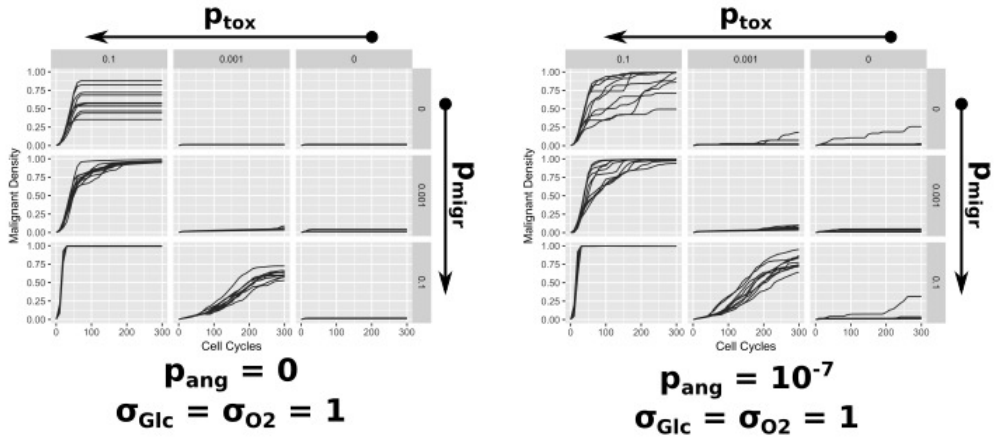


Figure 2.11: Malignant density for simulations with angiogenesis. The time progression for malignant density across multiple parameter combinations is plotted for simulations with $\sigma_{O_2} = \sigma_{Glc} = 1$. The results for 10 trajectories are shown for each parameter combination. Local toxicity effect (p_{tox}) increases from right to left. Migration (p_{migr}) increases from top to bottom. The probability of angiogenesis (p_{ang}) is 0 on the left and 10^{-7} on the right.

The transient existence of bridges between proliferation domains allow malignant cells to expand into new regions.

All simulations that included angiogenesis were screened to determine the frequency of new vascular site formation and the occurrence of bridging events. Fig. 2.13 shows that new vascular site formation is correlated with parameter combinations leading to increased malignant density. Furthermore, bridging events occur more frequently in simulations with larger numbers of vascular sites.

2.4 Discussion

The four observed growth patterns can be mapped to parameters (p_{tox}, p_{migr}, σ_{Glc}, and σ_{O_2}) in a manner that illuminates mechanisms underlying each response pattern.

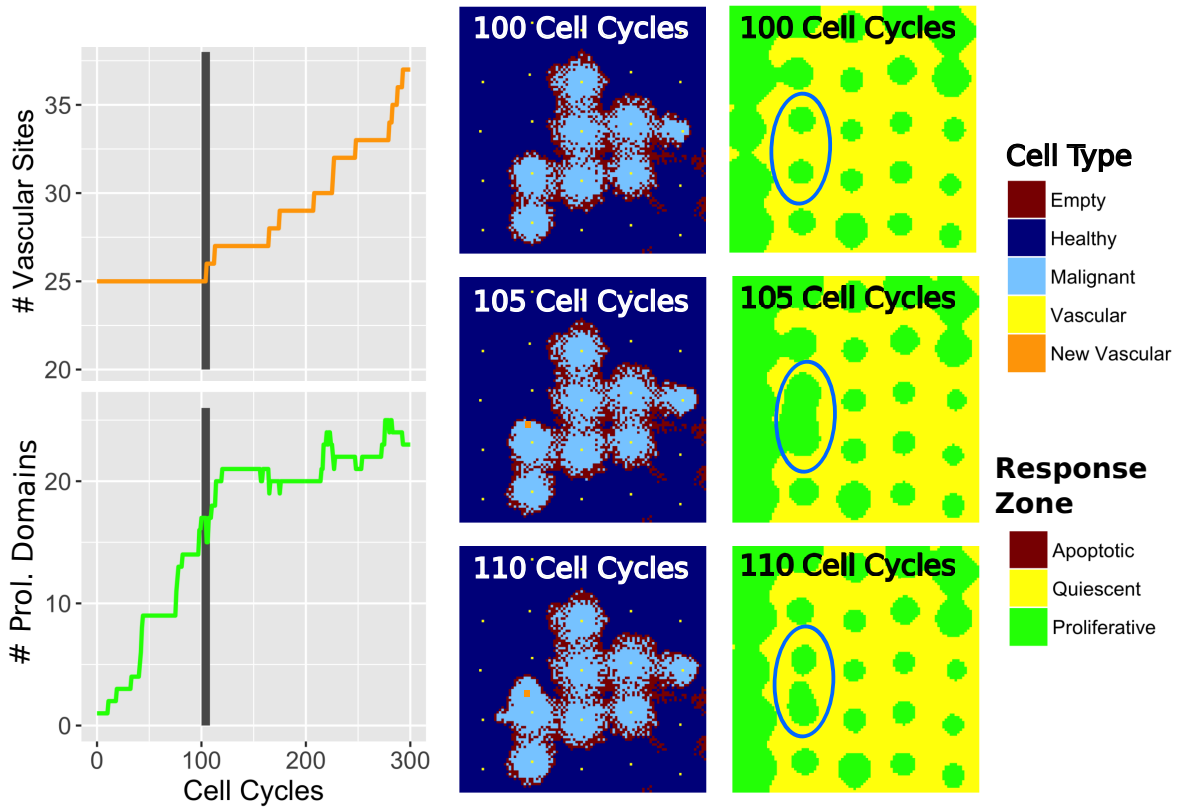


Figure 2.12: Tumor progression with angiogenesis. The time progression for the number of vascular sites and proliferation domains are shown on the left for a single simulation. Parameters used in the simulation are $p_{ang} = 10^{-7}$, $p_{tox} = 0.1$, $p_{migr} = 0.001$, and $\sigma_{Glc} = \sigma_{O_2} = 2$. The shaded region corresponds to appearance of the first new vascular site. To the right are cell progression snapshots shown with spatial maps indicating the cellular proliferation decisions possible given the concentrations of metabolites. Snapshots highlight simulation conditions prior, during, and after new vascular site formation.

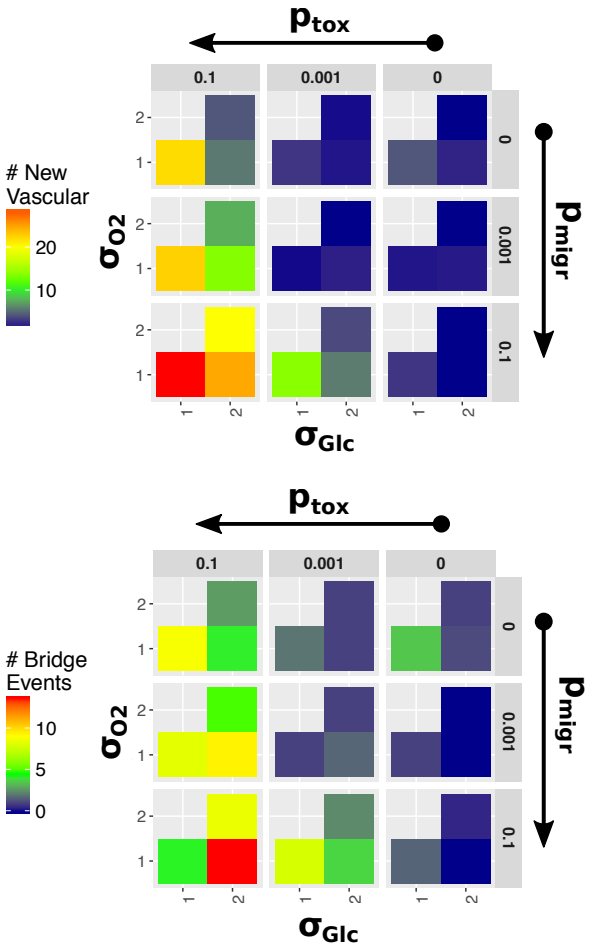


Figure 2.13: Number of bridging events. The mean number of new vascular sites is shown (top) for each screened parameter combination after 300 cell cycles for simulations with $p_{ang} = 10^{-7}$. Bridging events occur when the total number of proliferation domains transiently decreases following the formation of a new vascular site. Each simulation trajectory was screened for bridging events over the entire 300 cell cycles. The mean number of bridging events for each set of screened parameter combination are shown (bottom).

2.4.1 Isolated clustering

Isolated clustering is only observed in simulations with low values of p_{tox} that do not include glucose consumption. Under these conditions, healthy cells experience lower rates of apoptosis and fewer empty sites are produced. Without empty sites, proliferation of malignant cells results in displacement of vascular sites and the formation of proliferation domains. Furthermore, oxygen has a higher diffusion constant and a lower proliferation threshold than glucose ($c_{p,O_2} = 0.11$ versus $c_{p,Glc} = 0.3$). These features lead to larger proliferation domains than those observed in simulations including glucose consumption. Larger proliferation domains allow for increased proliferation of malignant cells. Combined with the reduced availability of empty sites, malignant cell proliferation displaces vascular sites beyond the edge of the simulation lattice. When this occurs, the vascular site is lost for the remainder of the simulation. The reduced vascular density leads to proliferation domains that are each infected by disjoint masses of malignant cells and the eventual death of most healthy cells.

While isolated clustering is an interesting growth phenomena, it is difficult to envision this as a biologically relevant regime. Vascular tissue does not simply disappear from an organism when displaced by rapidly growing cells. In reality, metabolite sources would still exist but metabolites would diffuse a greater distance from displaced vascular tissue. Our HCA model does not effectively account for vascular sites displaced beyond lattice boundaries. Conclusions from simulations exhibiting isolated clustering must be carefully considered within this context. However, the drastic difference in growth response due to the inclusion of glucose consumption is stark and emphasizes the importance of accounting for it when modeling tumor progression in vascularized tissue.

The absence of isolated clustering among simulations including glucose is an interesting phenomenon. For simulations without glucose consumption, oxygen acts as the limiting metabolite. Under these conditions and prior to extensive vascular site displacement, isolated clustering mimics continuous tumor growth. Model modifications to account for vascular site removal has the potential to reveal insights into tumor progression from malignant cells that exert minimal local toxicity effects where oxygen acts as the limiting metabolite.

2.4.2 Continuous growth

Continuous growth is the only tumor progression pattern observed in simulations both with and without glucose consumption. It is characteristic of simulations with high levels of local toxicity (p_{tox}), the probability that a malignant cell kills adjacent healthy cells. This subsequently increases the probability that malignant daughter cells are placed in empty sites, reducing overall levels of vascular site displacement. Reducing vascular site displacement decreases the formation of proliferation domains. For low glucose and oxygen consumption scalars (σ_{Glc}, and σ_{O_2}, respectively), proliferation domains may not form at all. For increased values of σ_{Glc}, and σ_{O_2}, migration allows malignant cells to reach new proliferation domains.

Of all observed growth responses, continuous growth exhibits the greatest malignant cell density. This growth regime occurs only when malignant cells exert high levels of local toxicity on healthy cells. For malignant cells with increased metabolic consumption, an additional requirement for continuous growth is increased malignant cell migration (p_{migr}). The degree to which these traits must be present in simulations that exhibit continuous growth may exceed biologically relevant values. It seems unreasonable to find a malignant cell that would guarantee cell death of neighboring healthy cells or discover a malignant cell that exhibit motion which mimics random diffusion. Each of these phenomena are observed in simulations with parameters producing continuous growth ($p_{tox} = 1$ and $p_{migr} = 0.1$). However, it should be noted that local toxicity and migration are tissue-dependent factors that have not been measured *in vivo*.

2.4.3 Tumor trapping

Tumor trapping is the result of healthy cells impeding malignant cell progression across quiescent regions to proliferation domains. It is only observed in simulations with glucose consumption and low values of p_{tox}. Initial proliferation of malignant cells leads to displacement of vascular sites as placement of daughter cells forces the relocation of adjacent cells. Proliferation domains form as a result of displaced vascular sites and there is a subsequent formation of quiescent regions. Malignant cells cannot proliferate within

quiescent regions and must rely on migration to reach nearby proliferation domains. The presence of healthy cells within quiescent regions acts as a barrier to migration. Tumor trapping can also be observed in simulations with intermediate values of p_{tox} and larger values of σ_{Glc} and σ_{O_2}. In this regime, proliferation domains decrease in size, increasing the chances of tumor trapping at intermediate p_{tox} values. Under these conditions, increasing p_{migr} increases the chance of malignant cells reaching new proliferation domains. For this to occur, p_{tox} must be high enough to kill healthy cells that are in quiescent regions between adjacent proliferation domains.

2.4.4 Domain seeding

Domain seeding is a response that appears in simulations with glucose consumption at intermediate values of p_{tox}, high values of p_{migr}, and high metabolic consumption scalars (σ_{Glc}, σ_{O_2}). In comparison to continuous growth and tumor trapping, these simulations show increased vascular displacement, increased numbers of proliferation domains, and the lowest discrete compactness scores. These measurements indicate that proliferation of malignant cells slightly outpaces the production of empty sites from healthy cell death. When this occurs, proliferating malignant cells frequently displace neighboring healthy cells when placing daughter cells and this results in vascular site displacement. Vascular site displacement is greatly subdued as compared to what is observed in isolated clustering and allows for proliferation domain formation. Unlike in tumor trapping cases, local toxicity is sufficiently high to kill healthy cells, thus clearing paths through quiescent regions separating proliferation domains. This allows malignant cells to migrate through empty sites in quiescent zones until they reach new proliferation domains. Growing tumors take on irregular shapes with low discrete compactness due to the stochastic nature in which new proliferation domains are seeded with malignant cells.

2.4.5 Angiogenesis

Including angiogenesis in simulations with glucose consumption provides an additional path for the growth of tumors subject to trapping and domain seeding. For cases of tumor

trapping, the formation of additional vascular sites increases available nutrients and formerly quiescent regions occupied by malignant cells can then support proliferation. Due to low values of p_{tox} associated with tumor trapping, proliferating malignant cells cause increased displacement of vascular sites while the continued presence of healthy cells block paths of migration to uninfected proliferation domains. Under these conditions, angiogenesis increases the number of overall malignant cells, but the tumor remains trapped in a single proliferation domain.

The dynamics of angiogenesis in domain seeding regimes include a similar effect and also allow for bridging events between proliferation domains. Bridging events create zones of proliferation between proliferation domains previously separated by quiescent zones. Without angiogenesis, malignant cells must migrate across quiescent regions for tumor growth to proceed. However, in the presence of angiogenesis, malignant cells can cross proliferation bridges by dividing, thus reaching proliferation domains that were previously free of malignant cells.

2.5 Conclusions

The pattern of tumor growth within the microenvironment is influenced by the relative magnitude of malignant traits. Our model indicates that vascular displacement caused by proliferation of malignant cells leads to the formation of proliferation domains. In the absence of toxic effects by malignant cells, simulated tumor progression is halted. Except in simulations where local toxicity is high enough to produce continuous growth in its absence, migration becomes a critical parameter for allowing escape from tumor trapping. A moderate degree of migration is required to transition to domain seeding from tumor trapping. Conversely, increasing metabolic consumption scalars, in a manner reflecting deregulated energetics in cancer, increases the likelihood of tumor trapping. Computer simulations allow us to decouple migration and local toxicity effects from deregulated energetics. Within the context of molecular biology, however, mutations associated with increased migration and reduced pH are linked those that produce aerobic glycolysis.

The transition from tumor trapping to domain seeding occurs at increased levels of local toxicity and migration. *In vivo* local toxicity and migration are associated with mutations that produce aerobic glycolysis and increased glucose consumption.[68] This includes increased activity of pathways such as PI3K/Akt/mTOR and increased H^+ concentration from elevated lactate production. Our simulation results suggest that increased glucose consumption from aerobic glycolysis creates barriers to tumor progression. These barriers take the form of regions with low glucose concentration that are unable to support proliferation. Additionally, healthy cells within quiescent regions obstruct paths by which malignant cells migrate to additional metabolite sources. The drawbacks from increased glucose consumption are offset by the increased ability for migration of malignant cells and the production of harsh microenvironemntal conditions that result in apoptosis of healthy cells. Furthermore, VEGF expression associated with aerobic glycolysis leads to angiogenesis. The formation of smaller proliferation domains caused by increased glucose consumption leads to reduced proliferation of malignant cells. This is offset by the increased availability of metabolites and merging of previously separated proliferation domains when new vascular sites form. Our simulations suggest that aerobic glycolysis leads to a fitness tradeoff for malignant cells. However, consequences of these molecular mechanisms and microenvironmental conditions associated with aerobic glycolysis lead to an increase in overall tumor invasiveness.

An interesting growth regime produced by our simulations is tumor trapping. The initial conditions for the simulations place a single malignant cell adjacent to a vascular site. This arrangement mirrors a critical step in the metastatic process. Specifically, following extravasation, the point at which a malignant cell exits vascular tissue and invades new tissue, and prior to the formation of small, malignant modules (micrometastases).[68] Tumor trapping mimics tumors that remained trapped as micrometastatic lesions. It has been suggested that, in order to progress to a macroscopic tumor, an "angiogenic switch" must be activated.[47] Our simulations suggest that, under certain environmental conditions, micrometastatic lesions can progress to macroscopic tumors without angiogenesis. Moderate levels of selective toxicity and migration facilitate domain seeding. This growth regime allows tumor progression beyond micrometastatic lesions in the absence of angiogenesis.

It has been observed that secondary tumors preferentially metastasize to specific tissues in the body. This tissue preference is unique to the tissue source from which the primary tumor originated.[32] Our simulations focused exclusively on conditions occurring within the brain. An examination of different vascular arrangements and cellular metabolite consumptions may reveal conditions that facilitate conversion of micrometastatic lesions to macroscopic tumors and offer an explanation for the tissue bias exhibited by secondary tumors.

Chapter 3

Analysis of network topologies underlying ethylene growth response kinetics in *Arabidopsis*

This chapter was reproduced with some modifications from the following publication: Prescott, A. M., McCollough, F. W., Eldreth, B. L., Binder, B. M., and Abel, S. M. Analysis of network topologies underlying ethylene growth response kinetics. *Front. Plant Sci.*, 7: 1308 (2016).

3.1 Introduction

Ethylene is the simplest of olefin gases and functions as a plant hormone, affecting many processes throughout the lifetime of a plant including seed germination, growth, formation of the apical hook, senescence, fruit ripening, abscission, and responses to various stresses.[93, 1] Ethylene inhibits the growth of dark-grown eudicot seedlings,[1] and sustained exposure to ethylene leads to a growth-inhibition response that has been used to screen for mutants and to provide information about the ethylene signaling network.[22, 65] Most proposed models of ethylene signaling consist of a linear pathway (Fig. 3.1), where in air ethylene receptors signal to the CONSTITUTIVE RESPONSE1 (CTR1) protein kinase which functions as a negative regulator of ethylene signaling.[79] CTR1 prevents ethylene

signaling by phosphorylating the ETHYLENE INSENSITIVE2 (EIN2) protein leading to its ubiquitination and proteolysis.[33, 76, 106] The binding of ethylene to ethylene receptors reduces the activity of the receptors, leading to reduced activity of CTR1 kinase and reduced phosphorylation of EIN2 protein.[33, 76, 106] The reduction in EIN2 phosphorylation leads to a decrease in ubiquitination of EIN2, causing a rise in EIN2 protein levels and allowing for proteolytic release of the C-terminal portion (EIN2-C) of the protein.[105, 76, 106, 132] EIN2-C affects the levels of two transcription factors, EIN3 and EIN3-Like1 (EIL1), in part by regulating their ubiquitination via S-PHASE KINASE-ASSOCIATED1-CULLIN-F-BOX (SCF) ubiquitin ligase complexes containing EIN3-BINDING F-BOX1 and 2 (EBF1 and 2) F-box proteins.[64, 102, 135, 52, 21, 8] In the presence of ethylene, ubiquitination of EIN3 and EIL1 is reduced, leading to accumulation of these transcription factors causing most ethylene responses.[64, 102, 135, 52, 19, 21, 8]



Figure 3.1: Basic linear ethylene signaling network

The above model was developed based on end-point analyses. Even though end-point analysis of ethylene responses continues to be an instructive bioassay, it is limited because transient events are overlooked. Time-lapse imaging has provided information about the kinetics of ethylene growth responses. The kinetics of ethylene responses have been studied for several plant species [84, 131, 28, 61, 109, 73] and most extensively studied in the model flowering plant, *Arabidopsis thaliana*.[19, 20, 103, 21, 54, 36, 125, 126, 136, 81, 82, 94, 15, 96, 108]

Studies of ethylene growth kinetics in Arabidopsis have revealed two phases of growth inhibition at saturating ethylene levels (1 ppm or above, see Fig. 3.2A).[20] The first phase is rapid, with a decrease in growth rate beginning approximately 10 min after the application of ethylene and lasting approximately 10 min, at which point the growth rate reaches a plateau. This plateau lasts approximately 30 min when a second, slower phase of growth inhibition is observed. Approximately 90 min after the application of ethylene, the growth rate reaches a minimum that lasts for as long as saturating levels of ethylene are present. If ethylene is

removed after 2 h, seedlings recover to pre-treatment growth rates in approximately 90 min (Fig. 3.2A).

The two phases of growth inhibition are genetically separable.[19]. Mutants lacking EIN3 and EIL1 (*ein3;eil1*) have a normal first phase of growth inhibition but fail to have a second phase response and over time return to pre-treatment growth rates in the continued presence of ethylene (Fig. 3.2B). This demonstrates that the first phase of growth inhibition is EIN3/EIL1-independent. Mutants lacking EIN2 (*ein2*) have no response to ethylene. Additionally, in wild type plants, adaptation is observed at sub-saturating levels of ethylene. At intermediate to high sub-saturating levels (e.g., 100 ppb), seedlings initially show both phases of growth inhibition but then have a partial recovery to an intermediate growth rate. At lower levels of ethylene (e.g., 2 and 10 ppb), only the first phase of growth inhibition occurs and is followed by recovery of the growth rate (Fig.3.2C).[19]

The experiments described above and other observations that indicate possible alternative pathways and feedback control [79, 110, 86, 66, 107, 108] suggest that ethylene signal transduction is not simply a linear pathway. Several network models have been proposed that involve more complicated topologies.[19, 54, 82] These and related networks are the focus of this paper.

A study comparing the time-dependent growth responses of several plant species led to a proposed network that included both coherent feedforward and negative feedback (CFF/NFB) signaling motifs (Fig. 3.3A).[82] In the CFF/NFB model, in air (i.e., without ethylene) the receptors signal to CTR1, which in turn inhibits downstream signaling. This leads to fast growth with feedback on growth occurring via the modulation of gibberellin (GA), a hormone known to stimulate growth. Application of ethylene inhibits the receptors, leading to a reduction of CTR1 activity and hence an increase in EIN2 levels. EIN2 is predicted to cause an initial growth inhibition response independently of EIN3 and EIL1. EIN2 also inhibits the EBF1 and EBF2 F-box proteins, leading to increases in EIN3 and EIL1. In the CFF/NFB network model, EIN3 and EIL1 inhibit growth via a GA-dependent and GA-independent pathway. The indirect inhibition of growth from EIN2 through EIN3 and EIL1 is hypothesized to be responsible for the second phase of growth inhibition. Qualitatively, this network topology provides a framework to understand the molecular basis

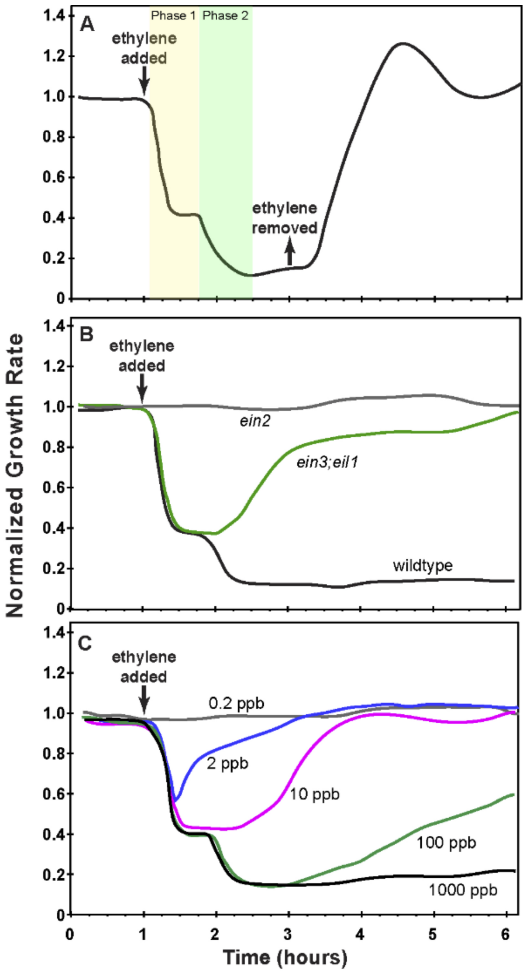


Figure 3.2: Growth response kinetics of dark-grown *Arabidopsis* seedling hypocotyls. (A) The normalized growth rate of wildtype Arabidopsis. Seedlings were grown in air for 1 h at which time 10 ppm ethylene was applied for 2 h at which time ethylene-free air was used to replace the ethylene. (B) The normalized growth rate of wildtype *Arabidopsis* compared to *ein3;eil1* and *ein2* mutants. Seedlings were grown in air for 1 h at which time 10 ppm ethylene was applied. (C) The normalized growth rate of wildtype Arabidopsis treated with varying concentrations of ethylene as indicated. Seedlings were grown in air for 1 h prior to application of ethylene. In all panels, the growth rate was normalized to the growth rate in air prior to treatment with ethylene. Based on data from [19, 20].

for phase 1 and phase 2 growth inhibition and its regulation by the coherent feedforward signal. It also provides a mechanism for transient growth inhibition in the absence of EIN3 and EIL1 that is regulated by negative feedback components.

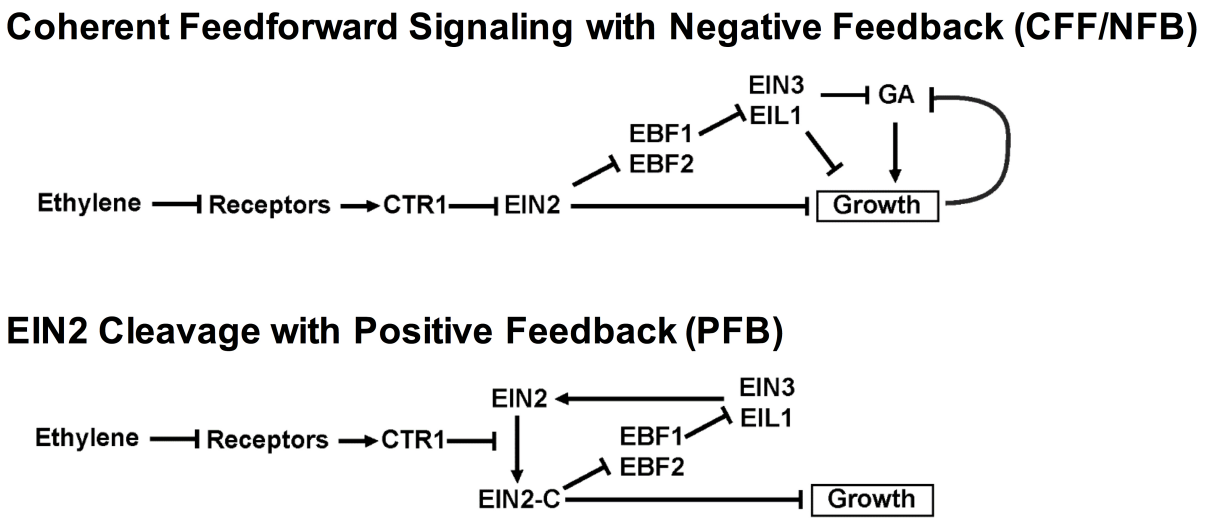


Figure 3.3: Diagrams illustrating proposed network topologies.

We were curious to determine whether other network topologies would yield response kinetics similar to experiments. Our collaborators previously proposed a model where one function of EIN3 and EIL1 is to provide feedback to regulate growth.[19] More recent data has shown that ethylene signaling requires the accumulation of EIN2 followed by the proteolytic cleavage of the EIN2 C-terminal tail.[4, 76, 106, 132] This accumulation of EIN2 protein is reduced in *ein3;eil1* double mutants [105] suggesting a mechanism for feedback by EIN3 and EIL1. We therefore developed a network topology where the function of EIN3 and EIL1 is to provide feedback stimulation in the form of increased synthesis of EIN2 (Fig. 3.3B). EIN2-C feeds back through EBF1/2 and EIN3 to promote EIN2. Without accumulation of EIN2, the prediction is that the levels of EIN2-C would decrease, resulting in growth reversal. We refer to this network as the cleavage with positive feedback (PFB) network.

Few computational models of ethylene signaling have been published,[38, 39] and none take into account the dynamical information obtained from kinetic studies of ethylene response. Considering experimental results of this nature provides an opportunity to identify network features underlying the observed growth responses. Therefore, we developed

computational models that account for proposed interactions within the CFF/NFB and PFB networks described above. These networks are modelled as sets of coupled ordinary differential equations (ODEs) describing the time evolution of network components. We are interested in (i) whether the proposed networks are capable of producing dynamical growth-response behavior consistent with experiments, and (ii) mechanisms underlying the network response when it does recapitulate experimental results. The CFF/NFB and PFB networks we consider have relatively high-dimensional parameter spaces (35 and 26 parameters, respectively). The parameters regulate numerous coupled, nonlinear ODEs describing the dynamics of the network, and changing the value of one parameter can have unexpected effects on network responses. Few *in vivo* measurements of network parameters are available. As such, we used an evolutionary algorithm (EA) for supervised machine learning in order to search for sets of parameters that produce network behavior consistent with experimental growth response kinetics. EAs are a class of numerical optimization techniques and have been used to investigate networks in a variety of biochemical applications.[14, 24, 48, 101, 12, 121, 119, 45] For example, time-course data has been used to determine parameters of small genetic networks.[80] and parameters associated with signal transduction in neurons.[10]

We used an EA to evolve parameter sets that produce ethylene growth responses similar to those observed in experiments. In particular, we focused on evolving two-phase growth inhibition (2-PGI) and *ein3;eil1* mutant partial growth recovery (MPGR). By repeatedly performing independent runs of our EA, we created libraries of evolved parameter sets. We gain insight into mechanisms underlying network responses by analyzing the dynamics of individual network components and the distributions of parameters governing the network. We additionally screen the parameter sets for partial growth recovery in the presence of sub-saturating ethylene doses, which is a property that emerges in some of the evolved parameter sets. We further explore each network by identifying simplified networks producing both 2-PGI and MPGR.

3.2 Methods

3.2.1 Kinetics of the ethylene response network

We model the ethylene growth-response networks proposed above using systems of coupled ordinary differential equations (ODEs). Ethylene concentration is treated as an input variable that is varied to mimic experimental conditions. We treat the growth rate, denoted by [Growth], as a concentration-like variable that measures the fraction of maximal growth rate. The unbound ethylene receptor concentration is described by a production term and a mass-action binding term representing ethylene binding. The time-dependence of all other components is described by production and degradation terms. As an example, the ODE describing CTR1 dynamics is written

$$\frac{d[\mathrm{CTR1}]}{dt} = \left(k_{\mathrm{prod}} \frac{[\mathrm{R}]^N}{K_{\mathrm{prod}}^N + [\mathrm{R}]^N} \right)_{\mathrm{R}} (1 - [\mathrm{CTR1}]) - k_{\mathrm{degr}}^{\mathrm{CTR1}} [\mathrm{CTR1}]$$

Concentrations are denoted by square brackets, k denotes a reaction rate, K denotes an activation coefficient in a Hill equation, and N is the associated Hill coefficient. All concentrations are restricted to the range of 0 to 1 and can be interpreted as the fraction of the maximum concentration for each species. The concentration range is constrained by describing the production as a logistic production term, with production vanishing when the concentration approaches 1. Interactions between network components are described with Hill-like kinetics. For example, in the above equation, CTR1 is promoted by receptors (R), which is captured by the Hill equation in the production term on the right-hand side. Inhibitory interactions promote the rate of degradation (e.g., EBF increases the degradation rate of EIN3 and EIL1). Stimulatory interactions promote the rate of production. The complete sets of ODEs for the networks studied are included in Appendix A. We initially equilibrate the system with no ethylene present, allowing it to reach steady state. We then introduce a step-change in ethylene to mimic experimental conditions. We treat EIN3 and EIL1 as a single entity, and therefore, to model the *ein3;eil1* mutant, the production rate for EIN3 is set equal to zero.

3.2.2 Evolutionary algorithm

Given the system of ODEs describing network dynamics, we perform supervised machine learning by using an evolutionary algorithm (EA) to identify sets of parameters that produce growth-response behavior similar to that observed experimentally. EAs are a class of optimization techniques that utilize the principle of inherited fitness to optimize parameters. A population of parameter sets is evolved over multiple generations. At each generation, each parameter set is evaluated by a fitness function and ranked by its fitness. Parameter sets with better rankings are modified in order to produce a new population of parameter sets for evaluation. The modifications consist of mutation and crossover operations. Mutations change a parameter value within a set to a new, randomly sampled value. Crossovers are events in which subsets from two high-performing parameter sets are recombined to form a new parameter set. Details of the mutations and crossovers depend on the specific implementation of the EA. Each iteration in which the population of parameter sets is updated is termed a generation.

We used the following fitness function for all proposed ethylene signaling networks:

$$\begin{aligned}\text{fitness} = &\sum_{i=1}^{N_{\text{wt}}} \alpha_i \left([\text{Growth}]_{\text{calc}}(t_i) - [\text{Growth}]_{\text{targ}}(t_i)\right)^2_{\text{wt}} \\ &+ \sum_{i=1}^{N_{\text{mt}}} \beta_i \left([\text{Growth}]_{\text{calc}}(t_i) - [\text{Growth}]_{\text{targ}}(t_i)\right)^2_{\text{mt}}\end{aligned}$$

Calculated growth values ($[\text{Growth}]_{\text{calc}}(t_i)$]) are determined by numerically solving the system of ODEs in MATLAB using the ode45 numerical solver. Target values ($[\text{Growth}]_{\text{targ}}(t_i)$]) were obtained using experimentally-determined growth rates at select times (Figs. 3.2A and 3.2B). Target values were selected from wildtype (wt) and *ein3;eil1* mutant (mt) experiments, with the index i in each sum indexing the target values (there are N_{wt} target values for the wildtype response and N_{mt} target values for the mutant response). For each time point, the squared deviation of the calculated growth rate from the target value is multiplied by a weighting factor (α_i, β_i) that emphasizes important regimes of the growth response. Specifically, we emphasize pre-ethylene steady state values, the wildtype two-phase growth inhibition response, minimal growth rate following ethylene introduction, and maximum

Table 3.1: Fitness function: Target growth values and scaling factors (wildtype conditions).

Elapsed Time (h)	Targ. Growth Rate	α_i
0.42	0.750	3
0.75	0.750	3
1.00	0.750	3
1.17	0.611	2
1.25	0.441	2
1.33	0.339	2
1.42	0.295	2
1.58	0.278	5
1.83	0.268	5
2.25	0.119	5
2.58	0.095	7
3.17	0.119	5
3.50	0.220	2
3.75	0.509	2
4.00	0.754	2
4.17	0.877	2
4.5	0.971	2

growth recovery levels. Target values and weighting factors are provided in Tables 3.1 and 3.2. Target growth rates were scaled by dividing all experimental growth rates by the maximum observed growth rate under both wildtype and mutant conditions. This gives a pre-ethylene growth rate target value that is less than one, in contrast with Fig. 3.2 in which growth rates were scaled so that pre-ethylene growth rates were unity. The evolutionary algorithm was designed to minimize the fitness function.

In our EA, we use a population size of 200 parameters sets at each generation. The initial parameter sets are generated by selecting uniformly distributed random values for each parameter from allowed parameter ranges (see Supplemental Table 3.3). Hill coefficients (N) are restricted to integer values. After evaluating each parameter set, the 50 parameter sets with the lowest fitness scores are selected as source parameter sets. These source sets are used to generate the population of parameter sets for the next generation. New parameter sets are produced by performing a two-point crossover followed by mutations. The crossover events and mutations allow a balance of global and local exploration of parameter space, and the

Table 3.2: Fitness function: Target growth values and scaling factors (*ein3;eil1* mutant conditions).

Elapsed Time (h)	**Exp. Growth Rate**	β_i
0.17	0.800	2
0.58	0.800	2
1.00	0.800	2
1.33	0.387	1
1.67	0.327	1
1.83	0.309	1
2.00	0.311	1
2.17	0.331	1
2.42	0.363	1
2.67	0.449	1
2.92	0.622	1
3.33	0.692	1
4.08	0.692	2
4.50	0.700	2

algorithm converged to local minima of the fitness function for the signaling networks studied. Two-point crossovers are performed by randomly selecting two source parameter sets with replacement (i.e., the same set can be chosen twice). Two crossover points are chosen at random from the list of parameters. Two blocks of parameters are taken from the first source and the other block is taken from the second source, leading to a newly constructed set of parameters. Additionally, the probability of mutation for each parameter is chosen such that on average approximately three parameters within the crossover product are mutated (the probability is 3/35 for the CFF/NFB network and 3/26 for the PFB network). When a parameter is selected for mutation, the decision to increase or decrease the value is made with equal probability. The parameter is then multiplied or divided, respectively, by a uniformly distributed value between 1 and 2. When this decision would result in a parameter exceeding its upper bound, a uniform random value between the current parameter value and its upper limit is used instead. The source parameter sets are updated each generation by replacing 25 randomly selected source sets with parameter sets having the lowest fitness scores from the current population.

Table 3.3: Parameter ranges. Ranges explored for each kinetic parameter within the evolutionary algorithm. (*) The lower limit of 2 was used for the CFF/NFB network; the lower limit of 1 was used for the PFB network. (#) This range was used for all k-parameters except for k_{degr} associated with EIN2-regulated EBF degradation in the PFB network. For this one case, 719 of the evolved sets used the full range. We then restricted the remaining 528 cases to sample from the range of 50-100 h^{-1}

Parameter	Min	Max
$k_{\text{prod}}, k_{\text{degr}}, k_{\text{cat}}$ (h^{-1})	0	100 #
$K_{\text{prod}}, K_{\text{degr}}, K_{\text{cat}}$	0	1
N	1 or 2 *	6

We ran the EA for 400 generations and recorded the parameter set with the lowest fitness score for additional analysis. This evolutionary process was repeated independently 500-4000 times depending on the network topology. Each parameter set was screened for the targeted network behavior and the resulting data was used to characterize ethylene response kinetics and identify features of the evolved parameters.

3.2.3 Screening for targeted responses

After running the EA, we check whether the resulting growth responses exhibit wildtype two-phase growth inhibition and/or *ein3;eil1* mutant partial growth recovery. Specifically, network responses are checked to ensure that the wildtype response meets the following conditions:

1. With no ethylene, steady state growth is sufficiently high
2. After applying ethylene, the minimal growth rate is sufficiently low
3. A plateau-like region separates the first and second phases of growth inhibition
4. Following removal of ethylene, growth recovers to a sufficiently high level

Logic diagrams for discriminant functions are included inl Fig. 3.4 and Fig. 3.5. For mutant behavior, traits 1 and 2 were used to check for proper behavior. The discriminant functions were designed to make the inclusion of false positives unlikely.

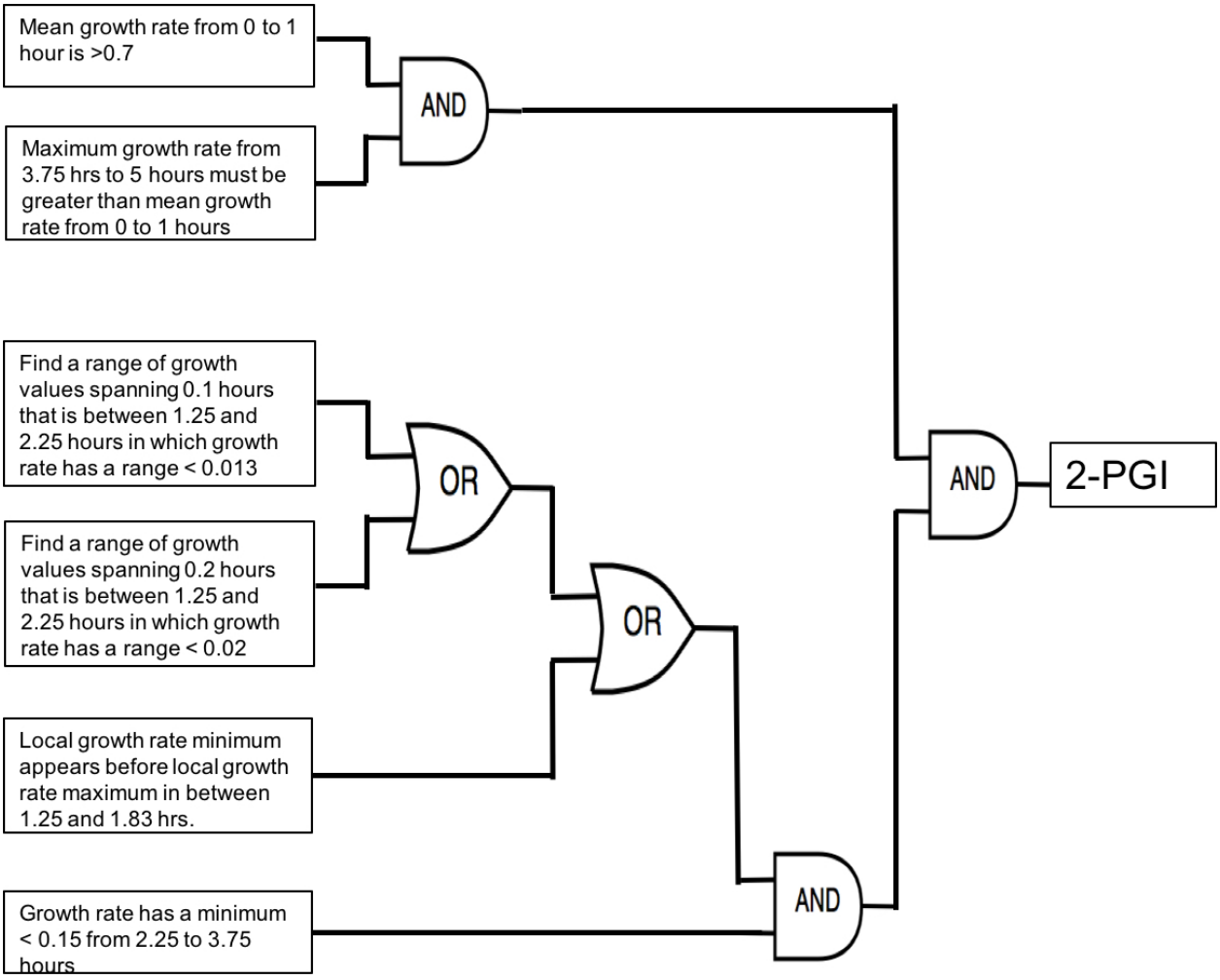


Figure 3.4: Logic diagram for 2-PGI screening. Growth responses under wildtype saturating ethylene conditions are screened using the outlined logic diagram. Growth responses reaching the final **2-PGI** terminal pass the screening procedure.

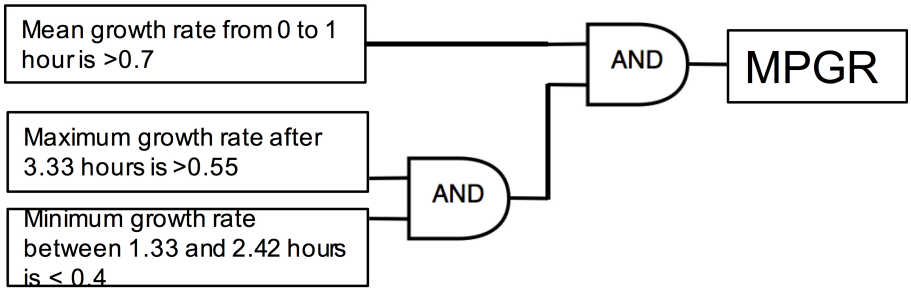


Figure 3.5: Logic diagram for MPGR screening. Growth responses under *ein3;eil1* mutant saturating ethylene conditions are screened using the outlined logic diagram. Growth responses reaching the final **MPGR** terminal pass the screening procedure.

3.2.4 Ethylene dose response

We additionally screen evolved parameters sets to identify whether they exhibit sub-saturating ethylene dose response kinetics similar to experimental observations (see Fig. 3.2C). The level of ethylene that leads to a sub-saturating response depends on the parameters of the network. Thus, we first identify the range of ethylene concentrations over which the network is responsive to concentration variations. For each evolved parameter set, we use a binary search method to identify the ethylene concentration range in which (i) the maximum dose produces long-time growth rate between 0.5 and 1.0% above minimum growth observed in the saturated response and (ii) the minimum dose produces a minimum growth rate between 0.5 and 1.0% below pre-ethylene steady state growth. We consider 20 evenly distributed ethylene concentrations between these bounds to test for partial growth recovery in the presence of sustained ethylene exposure. A parameter set is considered to exhibit sub-saturating growth recovery if there exists at least one ethylene concentration at which the growth maximum that occurs one hour or longer after the introduction of ethylene exceeds the minimum growth observed within the first hour of ethylene exposure by at least 0.1 (maximum possible growth is unity).

3.3 Results

3.3.1 Coherent feedforward/negative feedback network

The CFF/NFB network (Fig. 3.3A) was proposed by Kim *et al.* [82] based on growth kinetics in response to the addition and removal of ethylene. It can be broken down into three distinct regions: (i) the initial linear signaling cascade consisting of ethylene receptors, CTR1, and EIN2; (ii) a coherent feedforward loop with EIN2 as the initial node that inhibits growth both directly and indirectly (via EBF and EIN3); (iii) a negative feedback loop consisting of growth and GA. The coherent feedforward cascade interacts with the negative feedback loop as a result of the inhibitory effect of EIN3 on GA. As indicated in Fig. 3.3A, we treat EBF1 and EBF2 as well as EIN3 and EIL1 as single entities. We refer to these nodes as EBF and EIN3, respectively. This reduces the complexity of the model and the dimensionality of the parameter space while keeping key topological features of the network. Additionally, with existing experimental data, it is difficult to elucidate differences between these individual components, which could be included in a more detailed computational model.

We conducted multiple independent trials of the EA, obtaining 3774 sets of optimized parameters. Using the discriminant functions described previously, each evolved parameter set was screened for wildtype two-phase growth inhibition (2-PGI) and *ein3;eil1* mutant partial growth recovery (MPGR). Figure 3.6 shows examples of results that exhibit both 2-PGI and MPGR, as well as those that exhibit only one of the responses. Approximately 36% of the evolved parameter sets exhibit both 2-PGI and MPGR, 20% exhibit only 2-PGI, and 23% exhibit only MPGR. The large number of parameter sets yielding one or both of the targeted growth responses provides a large data set for analysis.

Dynamical response of network components

In Fig. 3.7, we plot the time dependence of each network component. This provides insight into how the feedforward and feedback loops shape growth response dynamics. For each component, we plot the mean response of the 1344 parameter sets exhibiting both 2-PGI and MPGR behavior. We also display the standard deviation about the mean (shaded

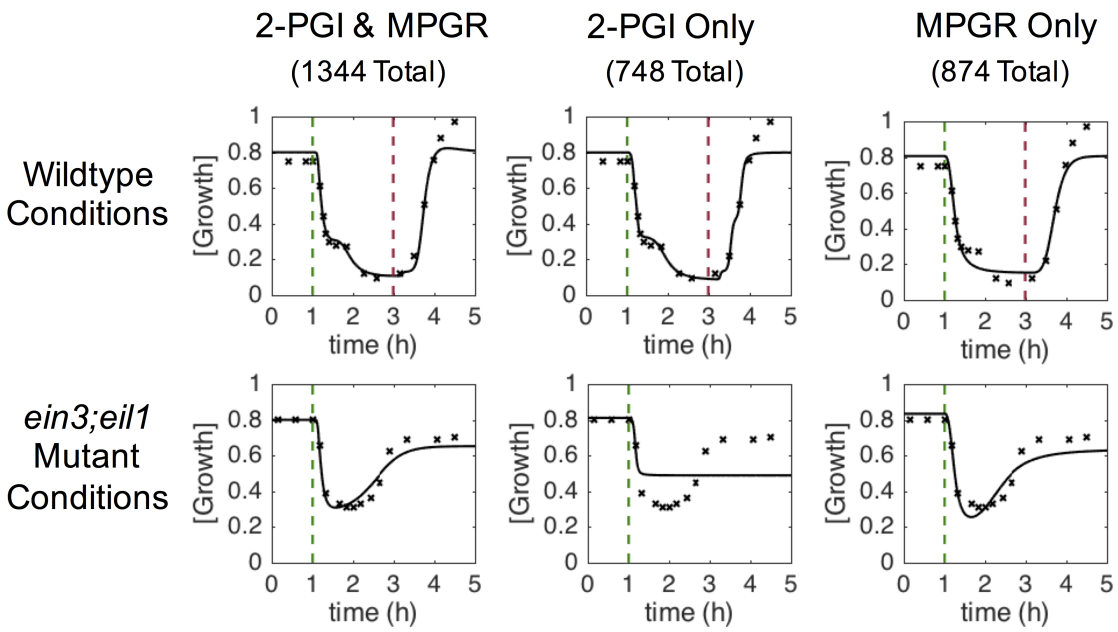


Figure 3.6: Characteristic growth responses at saturating ethylene doses. Columns show examples of time-dependent growth responses passing different combinations of screening criteria. Each column corresponds to a single set of evolved parameters. Rows show different simulated conditions (wildtype and *ein3;eil1* mutant). Targeted growth rates are denoted by x and the evolved response is shown by solid lines. Dashed vertical lines indicate time points at which ethylene was introduced (green) and removed (red).

regions) to characterize the heterogeneity of the response. Analogous results for parameter sets exhibiting only 2-PGI or MPGR behavior are provided in Fig. 3.8 and Fig. 3.9.

The response of the linear portion of the signaling cascade is identical for the wildtype (shown in Fig. 3.7A) and *ein3;eil1* mutant (not shown) topologies upon addition of ethylene. In response to the addition of ethylene at 1 h, the concentration of unbound receptors rapidly declines to levels near zero. This results in a decrease of CTR1 from a relatively high pre-ethylene concentration to a much lower concentration. Following this, EIN2 is no longer inhibited by active CTR1 and rapidly increases in concentration. There is a slight delay in the EIN2 response to ethylene due to the time required for the signal to propagate through the upstream components of the linear signaling cascade. Upon removal of ethylene at 3 h in the wildtype response, components return to their pre-ethylene levels.

The remaining network connections differentiate the wildtype response from the *ein3;eil1* mutant response. Components of the wildtype network are shown in Fig. 3.7B. Increasing

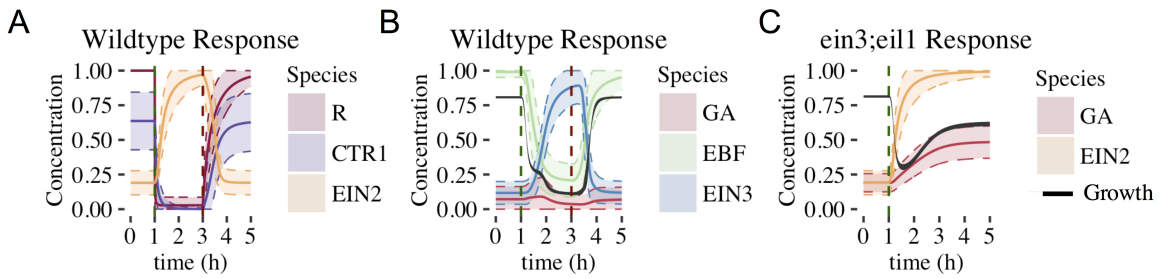


Figure 3.7: Time evolution of CFF/NFB network components (wildtype and *ein3/eil1* mutant conditions). Figures show the mean ±1 standard deviation for each component of the CFF/NFB network for cases exhibiting both wildtype and *ein3;eil1* mutant growth responses (1344 parameters sets). Components include: (A) early signaling components (wildtype conditions), (B) components downstream of EIN2 (wildtype conditions), and (C) negative feedback components affecting growth response (*ein3;eil1* mutant conditions). Black regions in (B) and (C) indicate the mean ±1 standard deviation of growth. Dashed vertical lines indicate time points at which ethylene was introduced (green) and removed (red).

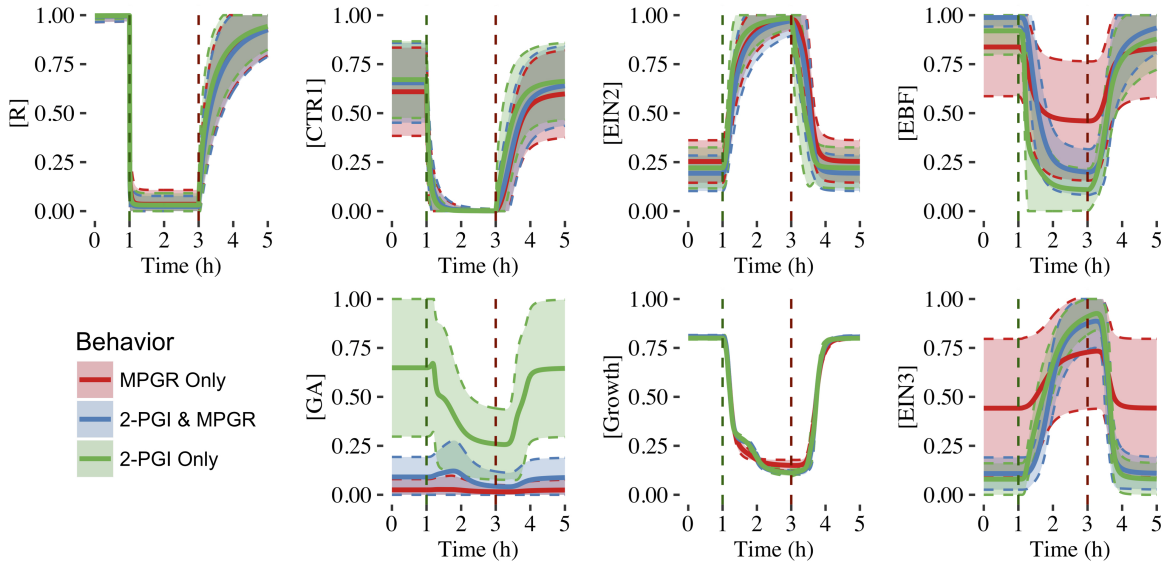


Figure 3.8: Time evolution of CFF/NFB network components (wildtype conditions). The mean response ± 1 standard deviation for all components from evolved CFF/NFB parameter sets passing 2-PGI and/or MPGR screening under wildtype saturating ethylene conditions.

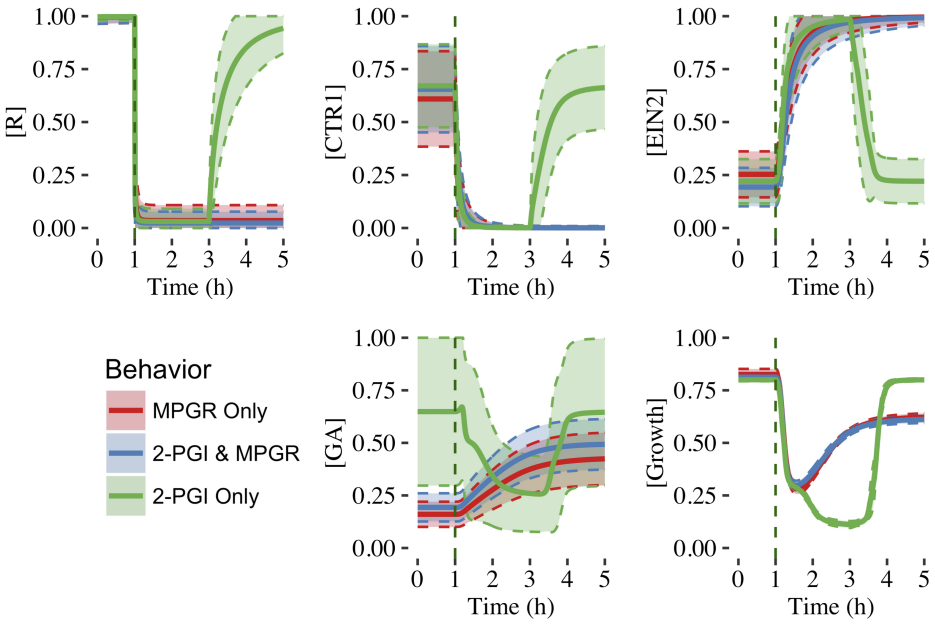


Figure 3.9: Time evolution of CFF/NFB network components (*ein3;eil1* mutant conditions). The mean response ± 1 standard deviation for all components from evolved CFF/NFB parameter sets passing 2-PGI and/or MPGR screening under *ein3;eil1* mutant saturating ethylene conditions.

EIN2 concentration acts to inhibit both growth and EBF, which is part of the indirect feedforward loop. The direct inhibitory effect of EIN2 on growth is responsible for the first phase of growth inhibition. In response to decreasing EBF concentration, EIN3 concentration increases from an initially low value approximately 30 minutes after ethylene is introduced. Once EIN3 reaches a sufficiently high concentration, a pronounced second phase of growth inhibition begins. Thus, the coherent feedforward loop leads to the desired 2-PGI.

The effect of the negative feedback loop can be understood by examining the dynamics of growth and GA. For the wildtype topology, GA is inhibited by both EIN3 and growth. A limited increase in GA levels accompanies the first phase of growth inhibition, and is driven by decreasing growth rates. During the second phase of growth inhibition, increasing EIN3 concentration inhibits GA, with EIN3 inhibition outcompeting the effect of decreasing growth rate. This drives GA to a low concentration, minimizing the effect of the negative feedback loop on growth. Thus, in the presence of increased EIN3, the effects of the negative feedback loop are suppressed. In the *ein3;eil1* mutant, however, the inhibitory action of EIN3 on the

negative feedback loop is lost. Thus, the negative feedback loop plays a more prominent role since GA is not inhibited by EIN3 (Fig. 3.7C). Additionally, the indirect path of growth inhibition is removed, eliminating the second phase of growth inhibition. Fig. 3.7C shows the response of key network components under these conditions. When increasing EIN2 levels cause a decrease in growth, GA levels increase in response, promoting growth and leading to partial growth recovery. This illustrates the importance of the negative feedback loop for partial growth recovery in the *ein3;eil1* mutant.

After the removal of ethylene at 3 h, it is interesting to note that the average growth rate does not exhibit a large overshoot compared with pre-ethylene levels (Fig. 3.7B). When analyzing individual parameter sets, none of the responses exhibit an overshoot that exceeds pre-ethylene levels by more than 10%, only 5 of 1344 exhibit >5% overshoot, and only 44 of 1344 exhibit >1% overshoot. This is in contrast with experimental results and suggests that modifications of the network or additional components might be needed to adequately capture the overshoot behavior. However, as discussed above, our results show that the core CFF/NFB topology generates key features of the 2-PGI and MPGR responses.

Analysis of parameter sets

The roles of the feedforward and feedback loops can be further understood by examining evolved parameters. In particular, it is instructive to characterize the distributions of evolved parameter values for the CFF/NFB network, as certain parameters are constrained to small ranges or excluded from certain parameter regimes. In Fig. 3.10, we compare the distributions of select parameters when the evolved parameter sets are categorized by their behavior (both 2-PGI and MPGR, only 2-PGI, or only MPGR). The distributions of all parameters are shown in Fig. 3.11. Parameter values are scaled by normalizing the maximum value to one, and the width of each distribution is scaled such that the maximum width is equal in each distribution. Comparing the distributions for specific parameters highlights key network features leading to each response.

The parameter distributions in Fig. 3.10 provide additional evidence that the feedforward loop plays a key role in generating 2-PGI. The activation coefficient for EIN2-regulated growth degradation of EBF ($K_{\text{degr}}^{\text{EIN2}}$) is excluded from low values in evolved parameter

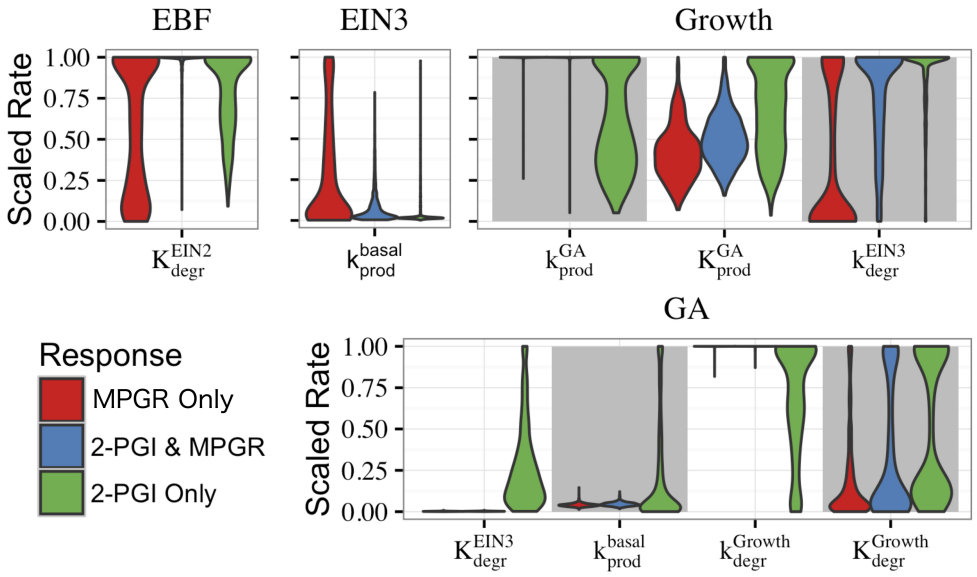


Figure 3.10: Distributions of parameters from evolved sets of CFF/NFB network parameters. Parameter sets exhibiting different combinations of responses are shown. Parameter labels K and k indicate activation coefficients and rate constants, respectively, that are associated with the ODEs governing the species labeled above each figure. Subscripts indicate if the parameter regulates degradation (degr) or production (prod) and superscripts indicate the network component regulating the reaction. Basal rates indicate that the parameter is not regulated by another network component. All parameters were unit normalized using range rescaling.

sets exhibiting 2-PGI. As a consequence, EIN2 concentration must reach high levels to significantly inhibit EBF, which contributes to a delay before the second phase of growth inhibition. Interestingly, this parameter is most significantly constrained in evolved parameter sets exhibiting both 2-PGI and MPGR. This is in contrast with the broader distributions seen for the cases exhibiting only one of the targeted responses. Additionally, in parameter sets exhibiting 2-PGI, the rate constant associated with basal production of EIN3 ($k^{\text{basal}}_{\text{prod}}$) occurs at low values. This also contributes to a time delay in the feedforward loop, which is needed for a second phase of growth inhibition.

It is also informative to consider the parameters governing the negative feedback loop (Fig. 3.10). Parameter sets exhibiting MPGR have highly restricted ranges associated with the rate constant for basal production of GA and the parameters for GA inhibition by growth. These restrictions lead to low pre-ethylene levels of GA and a reasonable response

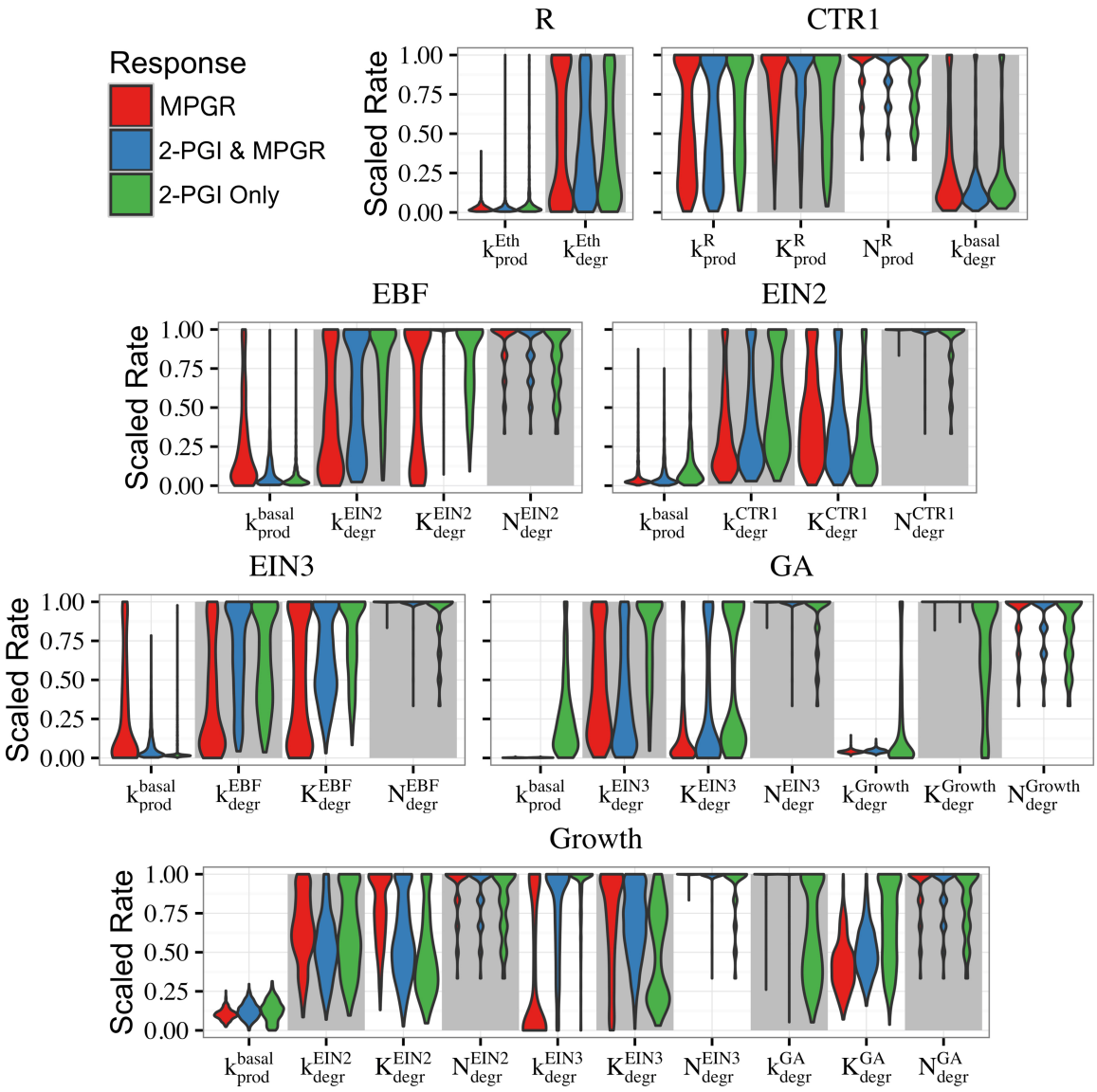


Figure 3.11: Complete parameter distributions for the CFF/NFB network. Parameter distributions of all CFF/NFB network parameters from evolved sets passing 2-PGI and/or MPGR screening. The width of each distribution is normalized such that the maximum width is equal across all parameters. All rates are unit normalized. Hill coefficients (N) are restricted to integer values.

of GA as growth declines following ethylene exposure. In cases exhibiting MPGR, the rate constant governing promotion of growth by GA is also restricted to high values and the activation coefficient for the promotion of growth by GA is excluded from the lowest values. Thus, a moderate increase in GA concentration will result in a significant increase in growth. However, excluding the activation coefficient from low values prevents increases from occurring with small changes in GA. Thus, tight regulation of parameters of the negative feedback loop is most readily apparent in parameter sets exhibiting MPGR. This further suggests the importance of the negative feedback loop for MPGR.

Ethylene dose response

Given that our network parameters were evolved to target only 2-PGI and MPGR behavior, we were interested in whether other experimentally observed behavior emerged as well. As such, we examined the sub-saturating ethylene dose-response behavior of evolved parameter sets exhibiting both 2-PGI and MPGR. Figure 3.12A shows an example of sub-saturating ethylene growth recovery (SSGR) that passes our screening criteria. Figure 3.12B shows an example of a typical growth response failing to exhibit SSGR behavior. Here, there is no growth recovery observed at any ethylene concentration. Of the evolved parameter sets exhibiting both 2-PGI and MPGR, 26% also exhibit SSGR. Partial growth recovery at large sub-saturating ethylene concentrations was observed experimentally (e.g., at 100 ppb in Fig. 3.2C) but was not observed in the CFF/NFB network model. However, the observed adaptive behavior occurring at lower ethylene concentrations is qualitatively consistent with experiments. This emergent property provides additional support for the proposed CFF/NFB network topology.

To gain insight into features that lead to SSGR, we compared the parameter distributions that passed SSGR screening to those that failed SSGR screening. Surprisingly, this revealed nearly identical parameter distributions except for the activation coefficient for GA inhibition by EIN3 and the rate constant associated with inhibition of growth by EIN3 (Fig. 3.13 and Fig. 3.14). The activation coefficient regulating GA inhibition by EIN3 occurs at higher values in sets producing SSGR. The rate constant for growth inhibition by EIN3 occurs at lower values more frequently in cases giving SSGR. We see the distributions of these

parameters across all evolved sets exhibiting 2-PGI and/or MPGR behavior (Fig. 3.10). Higher values of the activation coefficient for GA inhibition by EIN3 are found primarily in parameter sets exhibiting 2-PGI, while lower values of the rate constants for inhibition of growth by EIN3 occur primarily in parameter sets exhibiting MPGR. These restrictions apply to the regulation of EIN3 on components of the negative feedback loop. This again suggests that inhibition of the negative feedback loop by the coherent feedforward loop may play a key role in ethylene signaling.

Simplified CFF/NFB networks

We have shown that the CFF/NFB network can produce multiple experimentally-observed features of Arabidopsis growth responses to ethylene. Using this network topology as a guide, we probed simplified networks containing coherent feedforward and negative feedback motifs. The networks explored are shown in Fig. 3.15. Ethylene (E) acts as either the first node of the coherent feedforward loop (Fig. 3.15A-C) or as a direct input into the first node of the loop (Fig. 3.15D-E). Additionally, EBF and EIN3 are combined into a single node (Y) which acts to inhibit growth. In the simplest network (Fig. 3.15A), we remove the GA node and allow growth to directly inhibit its own production. For the remaining networks, the role of GA in the negative feedback loop is performed by node Z. To probe the inhibition of the negative feedback loop by the coherent feedforward loop, we tested network topologies with and without the inhibition of Z by Y. Approximately 500 independent optimization runs were performed for each simplified network topology. *Ein3;eil1* mutants were simulated by eliminating node Y. Evolved parameter sets were screened for 2-PGI and MPGR responses and a summary of results are shown in Table 3.4.

The simplest network (Fig. 3.15A) failed to produce any parameter sets passing 2-PGI or MPGR screening procedures. Examining the dynamical response of evolved parameter sets revealed two phases of growth inhibition that occurred too early and above the desired growth range. Additionally, no growth recovery was observed upon removal of Y. The addition of node Z to the negative feedback loop (Fig. 3.15B) produced two parameter sets exhibiting 2-PGI but none showing MPGR. When the inhibition of Z by Y is included (Fig. 3.15C), we begin to observe substantial numbers of parameter sets exhibiting either

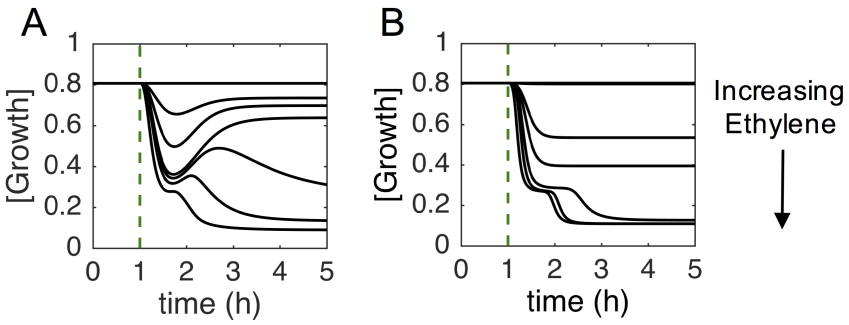


Figure 3.12: Characteristic growth responses at sub-saturating ethylene doses. Parameter sets exhibiting both 2-PGI and MPGR were screened for partial growth recovery to sustained sub-saturating ethylene doses (SSGR). Figures show typical growth responses for parameter sets: (A) passing SSGR screening and (B) failing SSGR screening.

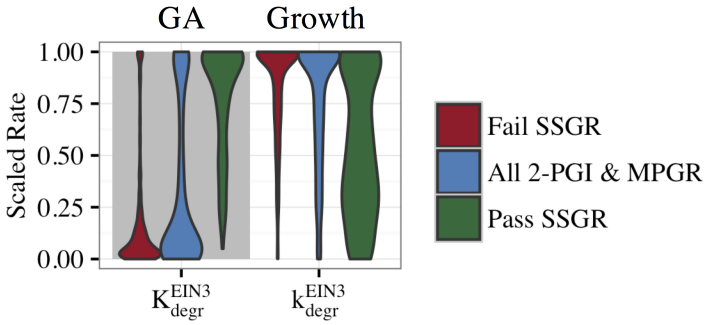


Figure 3.13: Distributions of parameters from the CFF/NFB network screened for SSGR behavior. A comparison of parameter distributions passing and failing SSGR screening (screened parameter sets exhibit both 2-PGI and MPGR).

Table 3.4: Screening results for ethylene growth responses of simplified CFF/NFB networks.

Network	Total	2-PGI & MPGR	2-PGI Only	MPGR Only
A	504	0	0	0
B	500	0	2	0
C	500	0	80	18
D	499	5	1	3
E	500	92	73	37

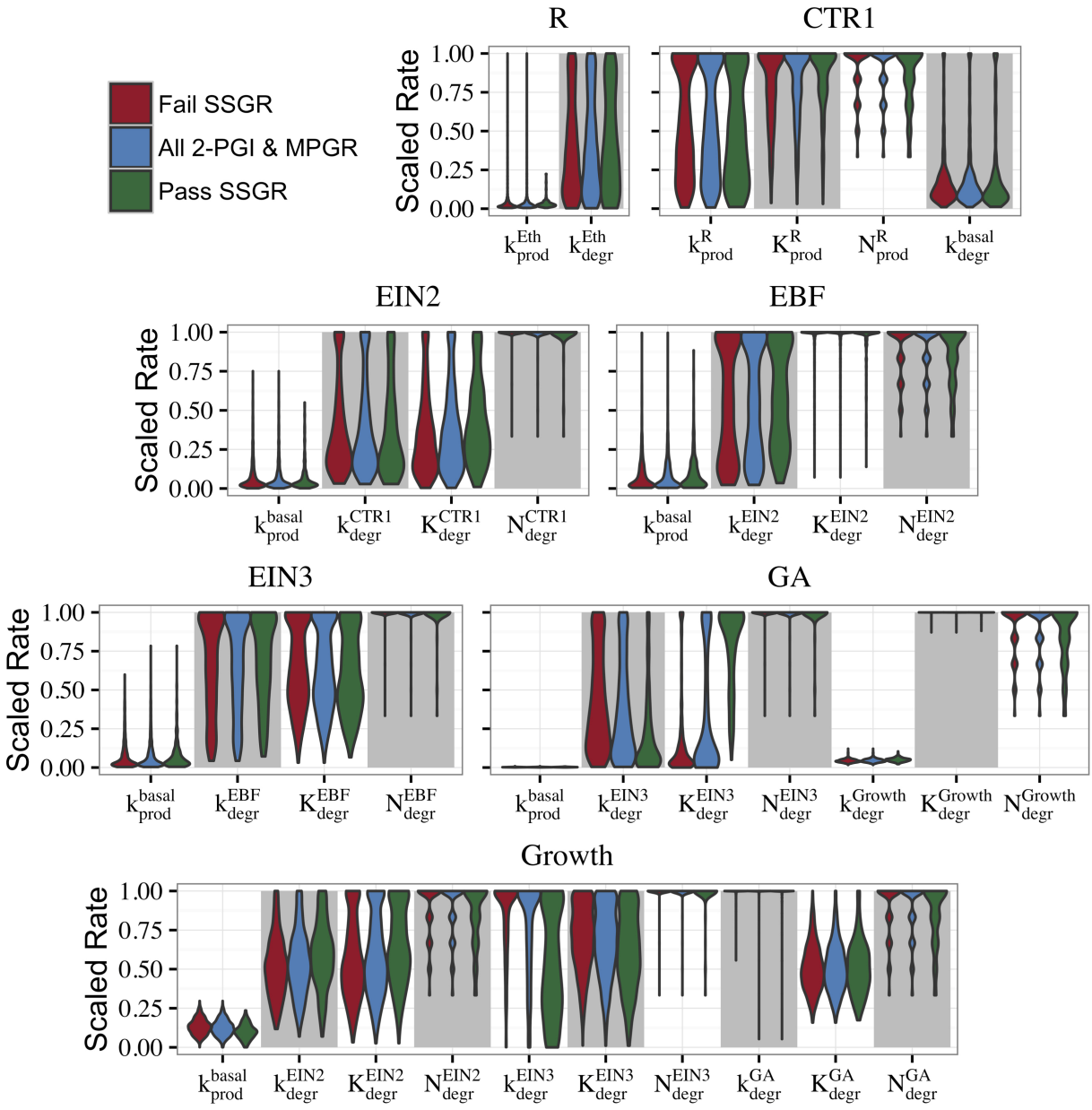


Figure 3.14: Parameter distributions for the CFF/NFB network based on SSGR screening. Comparison of the parameter distributions of all CFF/NFB network parameters for three cases: all sets passing both 2-PGI and MPGR screenings and the subset of these passing and failing SSGR screening. The width of each distribution is normalized such that the maximum width is equal across all parameters. All rates are unit normalized. Hill coefficients (N) are restricted to integer values.

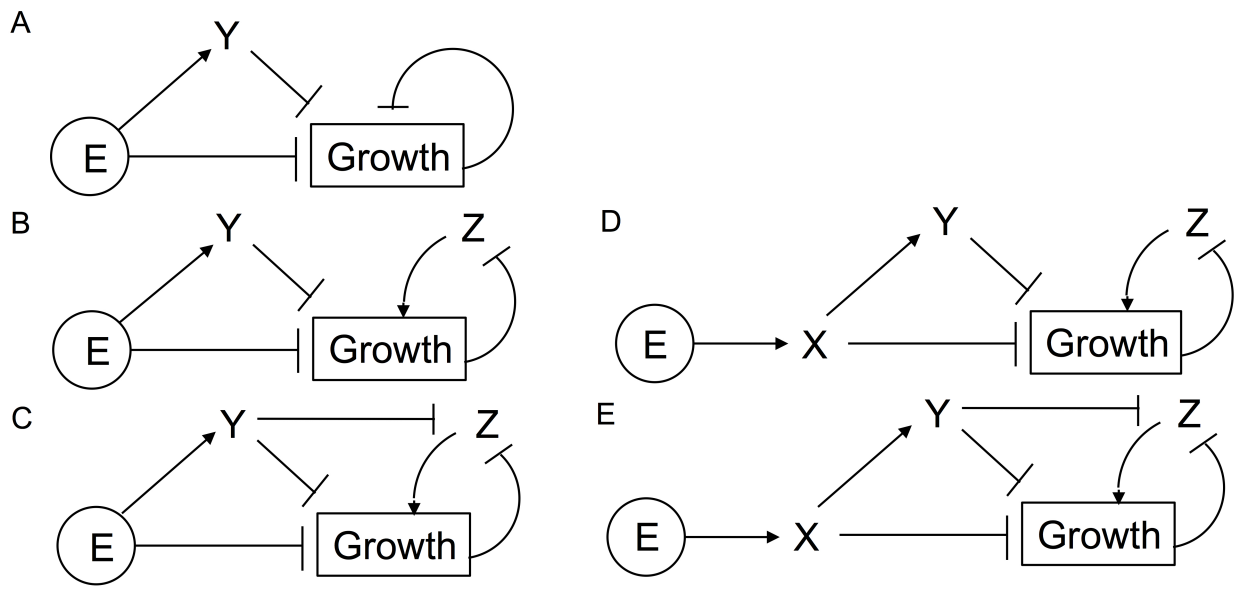


Figure 3.15: Simplified CFF/NFB networks. Simplified networks tested. Only networks D and E exhibit both 2-PGI and MPGR.

2-PGI or MPGR. However, no parameter sets simultaneously produced both responses. 2-PGI and MPGR were observed together only when ethylene promoted the first node of the coherent feedforward cascade, which more closely mimics the initial linear signaling cascade. In networks in which Z is not directly inhibited by Y (Fig. 3.15D), 1.0% of parameter sets exhibit both 2-PGI and MPGR responses. When Y regulates Z (Fig. 3.15E), 18.4% of parameter sets exhibit both targeted growth responses. These results suggest the importance of (i) the initial linear cascade in achieving proper timing of growth inhibition and (ii) the inhibition of negative feedback by the coherent feedforward loop in expanding the parameter space in which 2-PGI and MPGR are observed.

3.3.2 EIN2 cleavage with positive feedback network

In this section, we consider the second proposed network topology with EIN2 cleavage and a positive feedback loop (PFB network, Fig. 3.3B). The network was evolved with the same target responses as before. We obtained evolved parameter sets that produced both 2-PGI and MPGR behavior, but the number was substantially lower than in the CFF/NFB network. Out of 1247 independent runs of the EA, only 5 evolved parameter sets produce both 2-PGI

and MPGR. Parameter sets exhibiting only 2-PGI were also uncommon (3 sets). However, a significant proportion of parameter sets exhibited only MPGR (726 sets). Two of the 5 parameter sets exhibiting both 2-PGI and MPGR also display sub-saturating ethylene growth response (SSGR). The limited number of parameter sets exhibiting both targeted responses precludes analysis of parameter distributions. However, studying the dynamic response of the best-performing parameter set exhibiting 2-PGI, MPGR, and SSGR provides valuable insight (Fig. 3.16). Results are representative of the other parameter sets exhibiting 2-PGI and MPGR.

Dynamical response of network components

As in the CFF/NFB network, the introduction of ethylene decreases CTR1 levels (Fig. 3.16A). In the PFB network, the cleavage of EIN2 is no longer inhibited and EIN2-C is produced (Fig. 3.16A). As EIN2-C increases in concentration it inhibits both growth and EBF (Fig. 3.16A,B). EIN2 levels drop during this phase of network response as basal production of EIN2 cannot compensate for the rapid conversion of EIN2 to EIN2-C. EIN2 reaches low concentrations, limiting the resources available for production of EIN2-C. This leads to a transient decline in EIN2-C, which causes the plateau-like region of growth inhibition. Within the feedback loop, lower EBF levels decrease the inhibition of EIN3, which rises and promotes production of EIN2. The rapid rise of EIN2 provides more resources for EIN2-C production. This leads to the second phase of growth inhibition.

Within the *ein3;eil1* mutant, the positive feedback loop is absent. The addition of ethylene leads to EIN2 being converted to EIN2-C, resulting in a decline of EIN2. Without the positive feedback loop, there is no mechanism to further increase EIN2 production and its concentration monotonically decreases. EIN2-C initially increases but then declines as basal degradation eventually dominates the low rates of EIN2-C production associated with low levels of EIN2. As EIN2-C concentration decreases, partial growth recovery is observed (Fig. 3.16C).

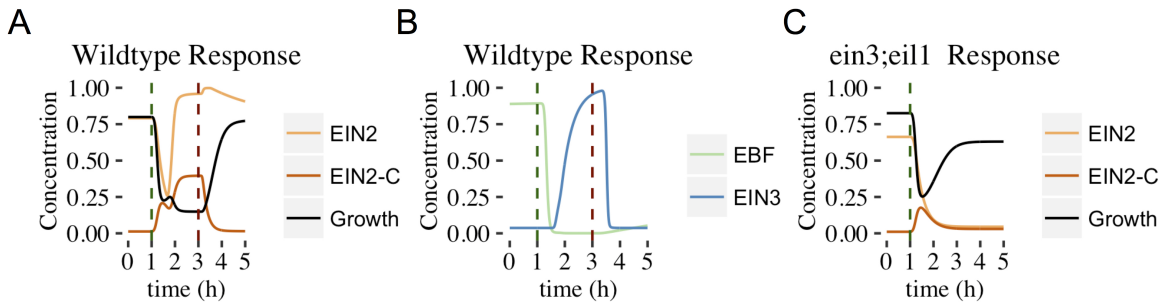
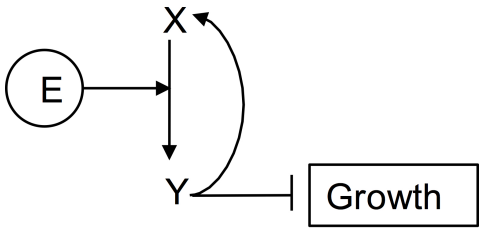


Figure 3.16: Time evolution of PFB network components. Figures show the behavior of the best-performing evolved parameter set that passed 2-PGI, MPGR, and SSGR screening. (A) Response of growth, EIN2, and EIN2-C (wildtype conditions). (B) Response of components in the positive feedback loop (wildtype conditions). (C) Response of growth, EIN2, and EIN2-C (*ein3;eil1* mutant conditions).

Simplified PFB network

We again explored a simplified network topology that keeps key features of the PFB network. We found that a four component network in which ethylene directly promotes the conversion of X (EIN2) to Y (EIN2-C) can produce both 2-PGI and MPGR (Fig. 3.17). In the network, Y directly inhibits growth and promotes the production of X. We performed 500 optimizations of this network and obtained 142 evolved parameter sets exhibiting both 2-PGI and MPGR. The marked increase in the fraction of parameter sets exhibiting both 2-PGI and MPGR suggests parameter evolution in the full PFB network is hindered by interactions in the positive feedback loop. Three parameters regulate the positive feedback from Y to X in the simplified network, while 9 parameters govern the positive feedback loop in the full network (associated with interactions between EIN2-C, EBF, EIN3, and EIN2). This apparently makes it difficult for our EA to evolve large numbers of parameter sets producing both 2-PGI and MPGR. The large fraction of simplified PFB networks exhibiting both 2-PGI and MPGR again provides support for EIN2 cleavage with positive feedback as a viable network topology.



Response	Count
Total	500
2-PGI and MPGR	142
2-PGI Only	0
MPGR Only	315

Figure 3.17: Simplified PFB network. Network diagram of the minimal PFB network. The table enumerates results of growth-response screening for evolved parameter sets.

3.4 Conclusions

We used computational methods, including supervised machine learning, to explore hypothesized network topologies underlying ethylene signaling responses in Arabidopsis. We focused on two core networks that are topologically distinct. Using an evolutionary algorithm to explore parameter space, we showed that both network topologies can produce dynamical responses consistent with experimental time-dependent growth data. The core topologies are (i) a coherent feedforward loop that inhibits growth and a negative feedback from growth onto itself (CFF/NFB), and (ii) a network in which ethylene promotes the cleavage of EIN2, with the product of the cleavage inhibiting growth and promoting the production of EIN2 through a positive feedback loop (PFB).

For the CFF/NFB network, high-throughput use of the evolutionary algorithm led to a large number of parameter sets producing responses consistent with experimental growth kinetics under various conditions and genotypes. The results emphasize the importance of various network features for regulating dynamic responses. For example, the two branches of the coherent feedforward loop collectively produce two-phase growth inhibition (2-PGI), and the negative feedback loop is critical for mutant partial growth recovery (MPGR). Our study additionally suggests that 2-PGI and MPGR coexist in a broader parameter regime when the negative feedback loop is suppressed by an intermediate component of the

coherent feedforward cascade. The large number of parameter sets producing 2-PGI and MPGR behavior provide insight into important regimes of parameter space. Additionally, a large fraction of these parameter sets also exhibit sub-saturating ethylene growth response (SSGR), even though this was not a targeted response by the evolutionary algorithm. Taken together, these results provide support for the CFF/NFB network as a viable network topology underlying ethylene signaling.

For the PFB network, the evolutionary algorithm led to far fewer parameter sets producing both 2-PGI and MPGR behavior, yet the dynamics of their responses provided insight into the mechanisms underlying the network topology. A key feature of the network is that EIN2 is converted to EIN2-C and its transient depletion upon the addition of ethylene is responsible for the plateau phase of growth inhibition. Two of the evolved parameter sets also exhibited SSGR, indicating that this emergent behavior is also possible in the PFB network. Although we generated far more parameter sets producing 2-PGI and MPGR for the CFF/NFB network, this does not necessarily imply that it is biologically more likely. For example, the region of parameter space for the PFB network that gives the desired behavior may be smaller or more difficult to identify with our EA, but this does not exclude the PFB network as biologically feasible.

It is interesting to note that different plant species have qualitatively different ethylene response kinetics.[82] For example, some plant species (millet) have only a transient first phase response and some (rice) have only a prolonged second phase response. The paper by Kim et al. first proposed the CFF/NFB network studied here. For millet, Kim et al. proposed that the circuit controlled by EIN3/EIL1 was missing to give the transient response; for rice, it was proposed that the rapid, EIN3/EIL1-indepenendent output of EIN2 is missing. The first case was analyzed in this paper when we analyzed the MPGR response. An interesting feature of the CFF/NFB model is that there is a simple conceptual way to modify the network to generate responses consistent with other species. It is less clear how the PFB network could be modified in an analogous manner to generate growth response kinetics consistent with the rice and millet studies. Further exploration of the network topology across species is an interesting area for future exploration.

Even though both models exhibited SSGR behavior that was similar to what has been observed experimentally, the kinetics of the computational responses are subtly different from experimental observations. In particular, there was no long-time recovery at high sub-saturating ethylene concentrations and incomplete recovery at low concentrations. It has been suggested that responses to low levels of ethylene are in large part a result of receptor clustering, where ligand occupancy of one receptor affects the signaling state of surrounding receptors through direct interactions and results in signal amplification at low ethylene levels.[53, 18, 20] Computational models invoking receptor clustering indicate this element can affect both sensitivity and adaptation.[25] Our models did not incorporate this feature, which would likely affect features of the SSGR. Additionally, our models do not incorporate spatial information. For instance, it is now known that EIN2-C translocates to the nucleus to affect ethylene signaling.[76, 106, 132] Cleavage of EIN2 was not incorporated into the CFF/NFB network and translocation of EIN2-C was not explicitly incorporated into either model. This translocation also may have diverse functions since it has recently been found that EIN2-C in the cytosol also has a role in ethylene signaling.[87, 96] It is likely that spatial changes in important components such as EIN2-C have a role in adaptation.

Despite these differences, our calculations show that several simple networks can recapitulate the ethylene growth responses observed experimentally. The dynamic responses observed provide opportunities for experimental exploration. A comparison of the dynamical response of individual components for each network is shown in Fig. 3.18. For example, the PFB network shows that when ethylene is added there is a transient decrease in EIN2 levels followed by accumulation of EIN2. By contrast, the CFF/NFB model predicts qualitatively different accumulation kinetics for EIN2, with no transient decrease. Thus, one avenue of experimentation can be to obtain more detailed spatio-temporal information about the accumulation of EIN2 (and EIN2-C) to determine if the details predicted by the calculations in either model occur when saturating levels of ethylene are added. For example, a detailed time-course of EIN2-C accumulation or EIN2 full-length protein is lacking. Such information would help determine which, if either, model correctly predicts the accumulation pattern for EIN2. Additionally, removing EIN3 from the CFF/NFB model has minimal effect on the time-course of EIN2 accumulation when ethylene is added, but has

a profound effect on both EIN2 and EIN2-C levels in the PFB model. Thus, experiments examining EIN2 and EIN2-C levels in *ein3;eil1* double mutants would also be informative. Another example is the involvement of GA in the CFF/NFB network where it plays a larger role in the growth kinetics observed in the *ein3;eil1* mutants. Detailed information about changes in GA levels would provide a test of this model and whether the negative feedback loop needs to be incorporated into the PFB network. Such experimental details will help determine which network topology, if either, could serve as the ethylene signaling transduction network of Arabidopsis. It is also possible that a combination of the two models or different network topologies will yield emergent properties that are closer to experimental observations. Additional experimental details about the spatio-temporal changes that occur in each component of the pathway will allow us to refine the above models or develop additional network topologies.

In summary, these calculations show that a basic mechanistic understanding of ethylene growth response and recovery kinetics is possible without detailed knowledge of the molecular mechanisms or enzymatic kinetic parameters. Given that ethylene signal transduction has been highly studied for several decades, we anticipate that major advances in our understanding about this pathway will be to provide details about network interactions, reaction kinetics, and changes in the spatial distribution of proteins in the pathway. Our hope is that with more refined experimental input, we can refine the network models to provide insights into how plants respond to ethylene.

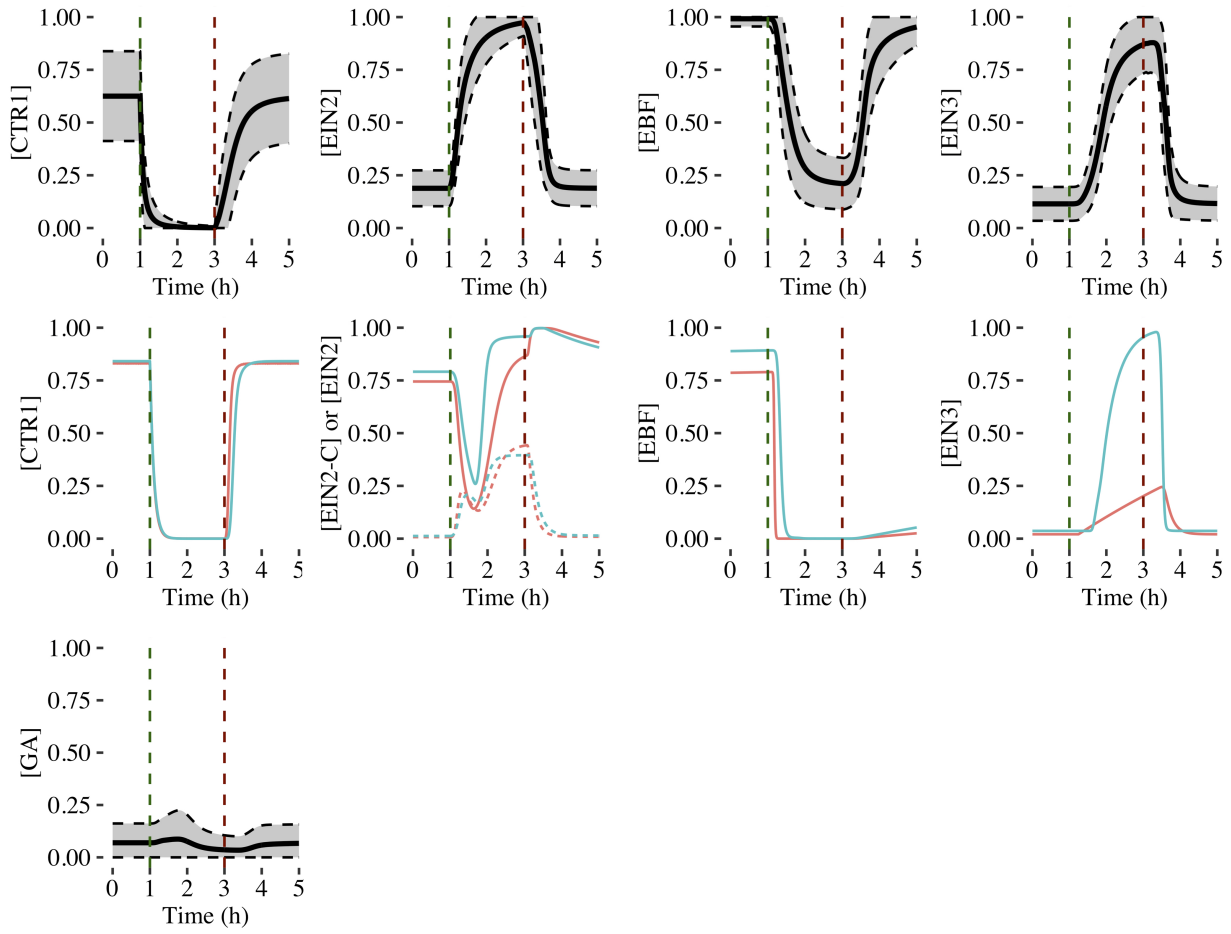


Figure 3.18: Time evolution of CFF/NFB and PFB network components (wildtype conditions). Comparison of the time evolution of components common to both the CFF/NFB (row 1) and PFB (row 2) networks. Results are from cases that exhibit 2-PGI, MPGR, and SSGR behavior. Row 1 shows the mean ± 1 standard deviation of responses from the CFF/NFB network. Row 2 shows the time evolution of components from the two evolved sets passing all screening from the PFB network. EIN2-C acts as the component inhibiting growth in the PFB network. This role is served by EIN2 in the CFF/NFB network. Comparing these components is illuminating for understanding the differences in the two networks. We plot together both EIN2-C (dashed line) and EIN2 (solid line) for the PFB network. Due to the critical nature of GA of the CFF/NFB network, we also plot its time evolution on the third row.

Chapter 4

Resdesigning Signaling Network Responses *In Silico*

This chapter was reproduced with some modifications from the following publication: Prescott, A. M. and Abel, S. M. Combining *in silico* and nonlinear dimensionality reduction to redesign responses of signaling networks. *Phys. Biol.*, 13(6):066015 (2016).

4.1 Introduction

Cells have evolved signal transduction networks capable of generating a wide range of responses. Recent efforts in synthetic biology have focused on signaling networks, with a goal of designing and building customized signaling circuits.[89, 99, 97] Synthetic networks can be created by engineering the connectivity of existing networks and/or by modulating parameters governing interactions in networks. Focusing on the modulation of parameters raises an interesting question: To what degree can a specific network topology give rise to varied responses by varying the underlying kinetic parameters? In this paper, we use machine learning approaches that combine *in silico* evolution and a nonlinear dimensionality reduction technique to investigate this question and to explore the plasticity of two signaling networks. The first network is relevant to the activation of T cells, which are regarded as prime candidates for synthetic biology applications.[134, 112] The second network contains common feedback motifs and further demonstrates the utility of the computational approach.[23]

Our approach uses *in silico* evolution for supervised machine learning in order to identify kinetic parameters that give rise to desired network behavior. We use an evolutionary algorithm (EA) in which a population of potential solutions (sets of parameters) are evaluated for comparison with the targeted behavior. The top performers are used to populate subsequent generations, using mutation and crossover to generate new potential solutions. EAs have been used successfully in a number of applications involving biochemical reaction networks.[14, 24, 101, 35, 100, 72, 119, 48, 85, 49, 12, 45, 104] For example, they have generated network topologies giving rise to behavior such as bistable switching and oscillations,[48] highly specific ligand discrimination,[85] and spatial segmentation patterns motivated by embryo development.[49] EAs have also been utilized as a tool in parameter estimation.[121, 80, 10] For example, time-course data has been used to determine parameters of small genetic networks[80] and parameters associated with signal transduction in neurons.[10]

In contrast with previous work, we consider fixed-topology networks and use an evolutionary algorithm to generate many sets of kinetic parameters giving the targeted network behavior. We couple this approach with a dimensionality reduction technique that allows us to cluster similar sets of kinetic parameters. Analyzing the distributions of parameters that characterize different groups then provides insight into constraints imposed by the targeted behavior, which can provide useful insight for synthetic biology applications. We use a nonlinear dimensionality reduction technique known as t-distributed stochastic neighbor embedding (t-SNE).[123] The central idea of the method is that a set of points in a high-dimensional space, which is characterized by the similarity between each pair of points, is embedded in lower dimensions in a way that attempts to preserve the pairwise similarities. It was used recently with mass cytometry data to identify differences between normal and leukemic bone marrow cells [7] and to identify populations of myeloid cells in mice.[16] Shekhar et al. further combined t-SNE with a kernel-based estimate of the density of points to identify phenotypic subpopulations of T cells in mice.[115] In contrast with linear dimensionality reduction techniques such as principal component analysis, t-SNE is not restricted to linear projections of the data from higher dimensions, which can be important in nonlinear systems such as signal transduction networks.

Fig. 4.1 provides a schematic overview of the paper. The signaling networks are described in the remainder of the introduction, and relevant computational details are described in the Methods section. The results are organized to (i) demonstrate the ability to evolve various input-output relations, (ii) utilize dimensionality reduction to group results of the evolutionary algorithm, and (iii) characterize the groups that emerge by visualizing distributions of parameters. We conclude by discussing the results in the context of designing and engineering signaling networks to generate desired cellular behavior.

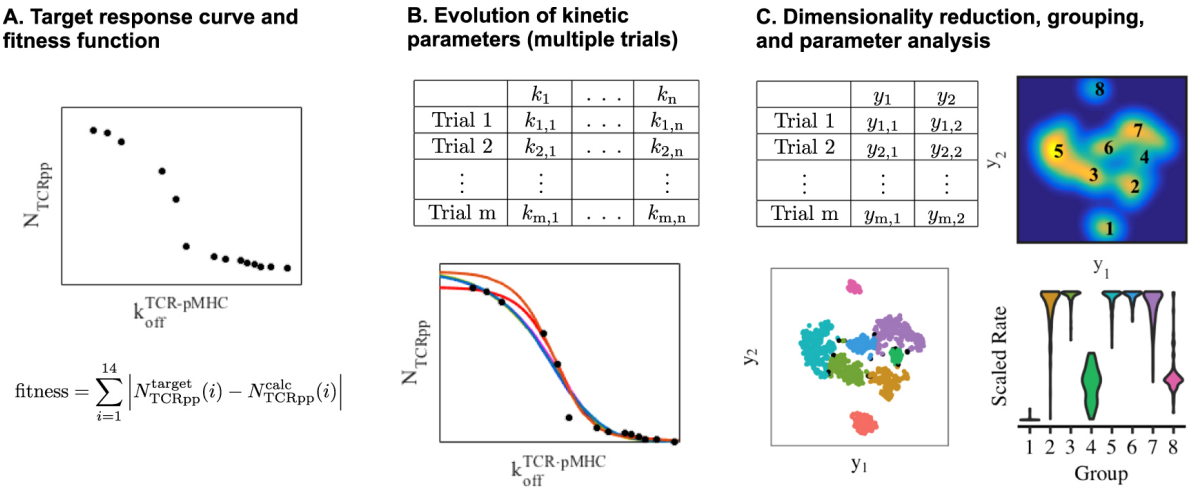


Figure 4.1: Schematic overview of the method. (A) A target response curve is defined along with an associated fitness function (shown here for the TCR network). Dynamics of the signaling networks are described by systems of ordinary differential equations. (B) Many trials are conducted to evolve the kinetic parameters through the use of an evolutionary algorithm (each trial seeks to minimize the fitness score). (C) A nonlinear dimensionality reduction technique (t-SNE) is used to embed the evolved parameters sets in two dimensions. Local peaks are identified using a kernel-based density estimation technique, and data points are assigned to a nearby peak. The distribution of parameters is analyzed as a function of group identity to give insight into constraints placed on the parameters by the targeted response.

4.1.1 Early TCR signaling network

T cells are a key element of the adaptive immune system and identify molecular signatures of pathogens presented on the surfaces of other cells. The binding of T cell receptors (TCRs) to antigenic ligands leads to a cellular response that is mediated by a well-studied signal transduction pathway.[44, 27, 124, 31] Because T cells are of central importance in detecting

disease, there is strong and growing interest in engineering T cells for therapeutic purposes such as cancer treatment and targeted immune suppression.[134] Thus, computational methods that aid in designing the changes needed to control T cell behavior will likely play a key role in engineering the behavior of T cells and other cell types.

T cell activation is controlled largely by the binding of T cell receptors (TCRs) to peptides presented by MHC proteins (pMHC) on the surfaces of other cells. Experimental evidence suggests that the lifetime of the TCR-pMHC complex is the major feature determining whether a T cell is activated.[111, 44] As such, we regard the off rate between TCR and pMHC ($k_{\text{off}}^{\text{TCR·pMHC}}$) as a measure of input strength in our models below. The binding of pMHCs to TCRs starts a signaling cascade in which a number of proteins are recruited to the T cell receptor complex. Among the recruited proteins are the coreceptors CD4 and CD8, which are found on the surface of helper T cells and cytotoxic T cells, respectively. The extracellular portions of the coreceptors bind to conserved regions of MHC proteins and the intracellular portions recruit the tyrosine kinase Lck, which phosphorylates residues within immunoreceptor tyrosine-based activation motifs (ITAMs) of the TCR complex. Lck can bind to both coreceptors and ITAM regions within the TCR complex, and hence can indirectly couple coreceptors to the TCR complex.

In this work, we consider a network model used by Artyomov et al. and Hoerter et al. to study the role of coreceptors in early T cell activation events.[11, 70] The model consists of three basic components: TCR, pMHC, and coreceptor (denoted by CD). Each component can bind with each of the other two, leading to the binary and ternary complexes shown in Fig. 4.2. Coreceptors are assumed to be bound to Lck, which allows binding between the coreceptor and the TCR complex in the model. The reactions describing the system are shown in Table 4.1. There is a binding (on) and unbinding (off) rate for each pair of binding partners. The off rate between pMHC and TCR ($k_{\text{off}}^{\text{TCR·pMHC}}$) is regarded as the model input, with small off rates (long binding times) leading to T cell activation in the native T cell response.[11, 70] The kinase activity of Lck leads to phosphorylation of the TCR when the TCR is directly bound to both pMHC and coreceptor. Phosphatase activity is modeled as a first order reaction that occurs only when a TCR is unbound from both coreceptor and pMHC. Although the TCR complex has multiple residues that can be

phosphorylated, in this model, TCRpp is regarded as representative of highly phosphorylated TCR. The phosphorylation state can be thought of as the state of a single ITAM domain. Using biologically relevant parameters, the network exhibits a sharp increase in TCR phosphorylation as $k_{\text{off}}^{\text{TCR·pMHC}}$ is decreased.[11, 70] High levels of TCR phosphorylation are associated with T cell activation.

4.1.2 Robust oscillator network

In addition to the early TCR signaling network, we study the "robust oscillator model" of Brandman and Meyer.[23] This is a toy model that contains both positive and negative feedback loops (Fig. 4.3) and that exhibits oscillations of network components in some parameter regimes. The network has a similar number of parameters as the TCR network described above, and we use the published parameter values that lead to oscillations as a starting point for *in silico* evolution. We regard this as a case study for having well-characterized network parameters that we wish to modulate to produce different output behavior. Here, we modify the period of oscillation of the output variable (A) in response to a step change in the input (I).

4.2 Methods

The computer code implementing the methods described below can be found at the Open Science Framework repository: https://osf.io/j6894/

4.2.1 Kinetics of the TCR network

Using the reaction network described above, we constructed a set of coupled ordinary differential equations (ODEs) obeying mass action kinetics that describe the deterministic, well-mixed time evolution of the concentration of each chemical species. As discussed in Fig. 4.2, the reactions result in 24 distinct species, and hence there are 24 coupled ODEs governed by 8 kinetic rates. The ranges of kinetic rates considered are tabulated in Table 4.1, and the complete set of ODEs is provided in Appendix A.

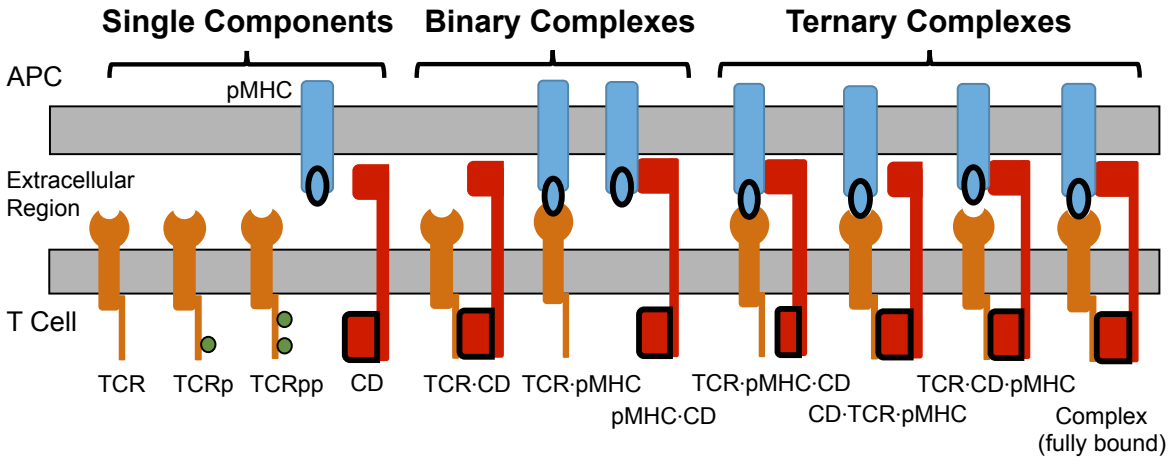


Figure 4.2: Schematic of TCR network species. The TCR and coreceptor (CD) are present at the T cell membrane; pMHC is present at the membrane of an antigen presenting cell (APC). Note that the TCR can exist in three states, representing different phosphorylation states, but we have displayed only the unphosphorylated state in the binary and ternary complexes. Accounting for multiple phosphorylation states, there are 24 chemical species in total (5 single components, 7 binary complexes, and 12 ternary complexes).

Table 4.1: Chemical reactions and ranges of kinetic rates for the TCR network. Phosphorylation of the TCR (reaction rate k_{kinase}) occurs only when both CD and pMHC are bound to the TCR.

Reaction	Reaction rate	Parameter range
TCR·pMHC $\rightarrow$ TCR + pMHC	$k_{\text{off}}^{\text{TCR·pMHC}}$	varied (input)
pMHC + CD $\rightarrow$ pMHC·CD	$k_{\text{on}}^{\text{pMHC·CD}}$	1 - 10^9 M^{-1} s^{-1}
pMHC·CD $\rightarrow$ pMHC + CD	$k_{\text{off}}^{\text{pMHC·CD}}$	0 - 50 s^{-1}
TCR + pMHC $\rightarrow$ TCR·pMHC	$k_{\text{on}}^{\text{TCR·pMHC}}$	1 - 10^9 M^{-1} s^{-1}
TCR + CD $\rightarrow$ TCR·CD	$k_{\text{on}}^{\text{TCR·CD}}$	1 - 10^9 M^{-1} s^{-1}
TCR·CD $\rightarrow$ TCR + CD	$k_{\text{off}}^{\text{TCR·CD}}$	0 - 50 s^{-1}
TCRpp $\rightarrow$ TCRp $\rightarrow$ TCR	k_{phosph}	0 - 2 s^{-1}
TCR $\rightarrow$ TCRp $\rightarrow$ TCRpp	k_{kinase}	0 - 0.5 s^{-1}
Amplification factor	α	10^5 or 1 - 10^9 M

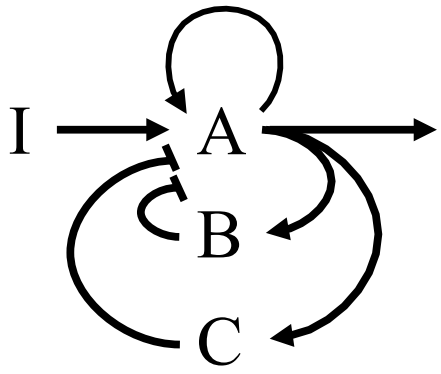


Figure 4.3: Robust oscillator network. Species I is the input, and the concentration of species A is regarded as the output variable. The network exhibits oscillations in some parameter regimes.

Our well-mixed deterministic model includes the parameter α, which we call the amplification factor, that accounts for proximity effects that occur within the ternary complex depicted in Fig. 4.2. For ternary complexes with two of the three possible bonds formed, the effective rate at which the third bond forms is enhanced because the reaction partners are in close proximity (i.e., the reaction volume is significantly reduced). To account for this effect in well-mixed simulations, the binding rate of the two unbound components is increased by a factor of α.

Initial TCR signaling events occur near the cell membrane. To model the membrane-proximal region, we used a system volume of 1 μm $\times$ 1 μm $\times$ 0.01 μm with a total of 300 TCR, 100 pMHC, and 100 coreceptors. These values are representative of typical surface concentrations.[11, 70] The initial conditions were 200 unbound TCR, 100 unbound coreceptors, and 100 TCR-pMHC complexes, with all other species having an initial population of zero. Rates were converted using the total volume so that ODEs describe the time-dependence of the number of each chemical species.

The ODEs were solved numerically in MATLAB using the solver ode23s, with solutions determined over a sufficiently long time to ascertain the steady state numbers of each species. For each case, we determined the total number of fully phosphorylated TCR (N_{TCRpp}) at steady state for use in the fitness function. This number includes both unbound TCRpp and those found in complexes.

4.2.2 Kinetics of the robust oscillator network

The robust oscillator network has one input variable (I) that affects three species (A, B, and C). The concentration of species A is regarded as the output variable. It autoregulates itself through a positive feedback loop. Species A also promotes species B and C, which in turn inhibit A, thus forming two separate negative feedback loops. Thus, as depicted in Fig. 4.3, the dynamics of A are controlled by an input variable, one positive feedback loop, and two negative feedback loops. We solve the following ODEs as formulated by Brandman and Meyer:[23]

$$\frac{d[A]}{dt} = [I](1 - k_2[C]) + \frac{[A]^{n_1}}{K_a^{n_1} + [A]^{n_1}} - [A] - k_1[A][B]^{n_2} \tag{4.1}$$

$$\frac{d[B]}{dt} = k_B([A] - [B]) \tag{4.2}$$

$$\frac{d[C]}{dt} = k_C([A] - [C]) \tag{4.3}$$

Square brackets denote concentrations, k denotes a reaction rate, K_a is an activation coefficient in a Hill equation, and n denotes a Hill-like parameter. Note that parameters, concentrations, and time are all unitless in this formulation. It is straightforward to introduce additional parameters to generalize the model and to introduce dimensional equations. The parameter ranges used in the evolutionary algorithm are provided in Table 4.2.

Table 4.2: Parameters and parameter ranges for the robust oscillator network (all parameters are unitless).

Parameter	Parameter range	Description
k_1	0 - 100	Negative regulation of A by B
k_2	0 - 10	Negative regulation of A by C
k_B	0 - 1	Production and degradation of B
k_C	0 - 1	Production and degradation of C
K_a	0 - 1	Activation coefficient, positive autoregulation of A
n_1	1, 2 , $\cdots$, 6	Hill coefficient, positive autoregulation of A
n_2	1, 2 , $\cdots$, 6	Hill coefficient, negative regulation of A by B

4.2.3 *In silico* evolution of the networks

We first implemented an evolutionary algorithm (EA) to optimize the T cell signaling network for sets of target values. For each set of target values, we performed multiple independent runs of the evolutionary algorithm. Each optimization run used 50 sets of parameters that evolved at least 100 generations and up to 750 generations (details are provided in Results). The evolved parameter sets consist of the kinetic rates listed in Table 4.1 (excluding the input rate), with the addition of the amplification factor in some cases. Each initial parameter set was chosen at random, as described below.

For each generation of the algorithm, the steady state number of fully phosphorylated TCR (N_{TCRpp}) was determined at 14 different values of $k_{\mathrm{off}}^{\mathrm{TCR{\cdot}pMHC}}$, which correspond to user-defined target values. These values were used to calculate the fitness for each set of parameters using

$$\text{fitness} = \sum_{i=1}^{14} \left| N_{\mathrm{TCRpp}}^{\mathrm{target}}(k_{\mathrm{off},i}^{\mathrm{TCR{\cdot}pMHC}}) - N_{\mathrm{TCRpp}}^{\mathrm{calc}}(k_{\mathrm{off},i}^{\mathrm{TCR{\cdot}pMHC}}) \right| \tag{4.4}$$

The ten top-performing parameter sets (lowest fitness scores) were selected to serve as an initial parent pool for the next generation. Fifty new parameter sets, which constitute the next generation, were obtained by subjecting the parent pool to mutation and crossover.

To generate a new parameter set, two parents were picked at random from the parent pool, with replacement (i.e., the same parent could be chosen twice). Given the ordering of kinetic rates in Table 4.1, a crossover point was chosen at random, with all rates before the crossover point coming from one parent and all rates after the crossover point coming from the second parent. Additionally, each parameter was mutated so that on average, each parameter set experienced one mutation per generation. When mutated, the new parameter value was chosen at random: For first order reactions, uniformly distributed random values were selected from the ranges shown in the table. For second order (binding) reactions, a random variable (ξ) uniformly distributed between 0 to 9 was generated and used to define the rate as $k_{\mathrm{on}} = 10^{\xi}$ (i.e., the logarithm of the binding rate was chosen to be uniformly distributed). When regarded as a variable, the amplification factor was chosen in an analogous manner to the second order rate constants.

For each generation after the first, the five parameter sets that produced the best fitness were chosen to replace five randomly selected members of the parent pool. This produces an updated parent pool, which was used to generate the next generation. For each set of target values, we performed many independent optimizations, each starting from a different initial pool of parameter sets. In total, 1500-2000 independent optimization runs were performed for each input-output relation. The top-performing parameter set was obtained from each optimization and screened for the desired response. The aggregated parameter sets were mined for emergent patterns and behaviors.

For the robust oscillator network, we implemented the EA to evolve different periods of oscillation in response to a step change in the input. Starting from an input of $[I] = 0.1$, we introduce a step change to $[I] = 0.4$ and target different periods of oscillation (T_{target}) of species A using the following fitness function:

$$\text{fitness} = \left(N_p - \left\lfloor \frac{t_{\text{tot}}}{T_{\text{target}}} \right\rfloor\right)^2 + \sum_{i=1}^{N_p} \left(\left(t_{\text{peak}}(i+1) - t_{\text{peak}}(i)\right) - T_{\text{target}}\right)^2 \tag{4.5}$$

Here, t_{tot} is the total observation time after the change in input and T_{target} is the targeted period. After solving the ODEs, we calculate N_p, the number of maxima of $[A]$ observed after the change in input, and $t_{\text{peak}}(i)$, the time of the i^{th} maximum. The fitness function penalizes trajectories that have the wrong number of peaks and timing between peaks that differs from the target value. The EA is implemented as in the TCR network, with the exception that it is initialized with a population parameterized by previously published values [23] and it is evolved for 50 generations.

4.2.4 Dimensionality reduction

For each netowork, we utilized t-distributed stochastic neighbor embedding (t-SNE) to embed our original data sets in two dimensions to facilitate analysis and visualization. This method is a nonlinear dimensionality reduction technique that seeks to maintain pairwise similarities between points when mapping from a high-dimensional space to a lower-dimensional space. The method was introduced by van der Maaten and Hinton (discussed in detail in Ref. [123]) and is briefly outlined here. The method begins by converting the Euclidean distance between

each pair of points in high dimensions to a similarity based on the conditional probability,

$$p_{i|j} = \frac{\exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\|\mathbf{x}_i - \mathbf{x}_k\|^2/2\sigma_i^2\right)}\,, \tag{4.6}$$

where σ_i^2 is the variance of a Gaussian centered on point i that is calculated through an iterative process. The conditional probability is then used to construct pairwise probabilities, $p_{ij} = (p_{i|j} + p_{j|i})/2M$, where M is the total number of data points. Higher probabilities are assigned to points that are closer together. To embed the data in a lower-dimensional space, points are initially placed stochastically. Pairwise probabilities (q_{ij}) are calculated in the lower dimensional space using

$$q_{ij} = \frac{\left(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2\right)^{-1}}{\sum_{k\neq i}\left(1 + \|\mathbf{y}_i - \mathbf{y}_k\|^2\right)^{-1}} \tag{4.7}$$

The algorithm seeks to minimize the difference in the joint probability distribution in the high-dimensional space ($\{p_{ij}\}$) and in the low-dimensional space ($\{q_{ij}\}$), as measured by the Kullback-Leibler (KL) divergence,

$$D_{\mathrm{KL}}\left(\{p_{ij}\}\|\{q_{ij}\}\right) = \sum_{i,j} p_{ij}\log\frac{p_{ij}}{q_{ij}} \tag{4.8}$$

A gradient descent method is used to minimize the KL divergence by rearranging the points in the low-dimensional space.[123]

Before applying t-SNE, all parameters were unit normalized as follows. For the TCR network, first order rates were divided by ranges shown in Table 4.1. For second order rates and α, the logarithm (base 10) of each was divided by 9. For the robust oscillator network, all rates were divided by the ranges shown in Table 4.2. We utilized the tsne function from the tsne Vers.0.1 package in R.[41] To calculate the quantity σ_i, a quantity known as the perplexity must be specified. This quantity can be viewed as a measure of the number of nearest neighbors of a point. The perplexity for all data sets was chosen to be 30. We also tested values of the perplexity of 10 and 50 and verified that the results presented in the paper were robust to these changes.

4.2.5 Identification of groups

To identify groups within the embedded data, we used kernel-based density estimation with the Gaussian kernel

$$K_\gamma(\mathbf{y}) = \frac{1}{2\pi\gamma^2} \sum_{\mathbf{y}' \in \mathcal{Y}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{y}'\|^2}{2\gamma^2}\right) , \tag{4.9}$$

where $\mathbf{y}$ is a point in two dimensions and the sum extends over all embedded data points.[115] The parameter γ controls the width of the density function around each point and was typically chosen in the range of 10 - 20. Given the kernel density estimate, we identified local maxima and mapped these maxima onto different "group" identities, with the data points near each peak being members of the group. To minimize the effects of local density fluctuations, we combined local peaks residing close to one another in the two-dimensional space (we used a cutoff distance of 5).

Once we identified local maxima, we used a hill-climbing algorithm to map each data point onto one of the local maxima. The two-dimensional space was discretized into a square lattice with unit-length lattice spacing. Starting from the location of each embedded data point, we first allowed the point to undergo 10 biased random walk steps on the lattice, with the probability of hopping in a given direction proportional to the density at that site (this was useful for assigning points near boundaries between different maxima). The point then underwent a strict hill-climbing process, choosing the direction of maximum density at each step. The data point was then grouped with all other data points that reached a given maximum. For each point, the random walk and hill-climbing procedure was performed 50 times, and only points that reached the same maximum at least 80% of the time were assigned to a group. Parameter sets failing this requirement were left unassigned to groups.

Scaled parameter distributions for all groups are visualized using violin plots produced using the ggplot2 package in R.[133] The width of each violin plot is proportional to the number of parameter sets having values in a given range. The width of each plot is scaled such that all distributions have equal maximum widths.

4.3 Results

4.3.1 Evolution and analysis of the TCR network

Evolving input-output relations.

Given the TCR signaling network described above, we investigate the ability of the network to produce varied input-output responses by modulating the kinetic parameters governing the underlying dynamics of the network. In particular, using *in silico* evolution, we show that:

- The TCR activation threshold can be shifted by orders of magnitude, so that the transition from an inactive to active state occurs at vastly different values of the TCR-pMHC off rate.
- The response can be altered so that it exhibits a graded input-output relationship.
- The input-output relation can be "inverted," so that TCRs are highly phosphorylated for short TCR-pMHC binding times but not for long binding times.

By conducting many independent optimization runs for each input-output relation, we obtained multiple sets of kinetic parameters for each network response studied. We focus on the top-performing results in this section and characterize the population of results in subsequent sections.

Fig. 4.4 demonstrates the ability of the TCR network to exhibit a sharp signaling response with the threshold from inactive to active shifted over six orders of magnitude. Four separate input-output relations are shown in the figure. The black squares represent the "native" T cell response that characterizes the activation profile of a typical T cell.[11] The other three response curves are obtained by shifting each of the points on the native response curve by a factor of 10^{-2}, 10^{2}, and 10^{4} on the horizontal axis ($k_{\text{off}}^{\text{TCR·pMHC}}$). Hereafter, we refer to the different cases as native and shifted (further specified by the factor by which the curve is shifted). For each case, we performed at least 1900 independent runs of the EA. Each run consisted of 300 generations. The best fit from all of the runs is shown for each case (less fit cases exhibit similar behavior and are discussed below). The evolved input-output

relations capture the key switch-like shape of the targeted points and fit the values reasonably well. The results consistently "miss" one of the points in the transition region, generating a transition that is less steep than the target curve. The points we chose were motivated by biological responses,[11] yet we note that the network we consider is only the first part of the TCR signaling network. We also do not account for spatial and stochastic effects that may be important. Although these neglected features may contribute to generating steeper response curves, we regard the input-output relations obtained as good fits.

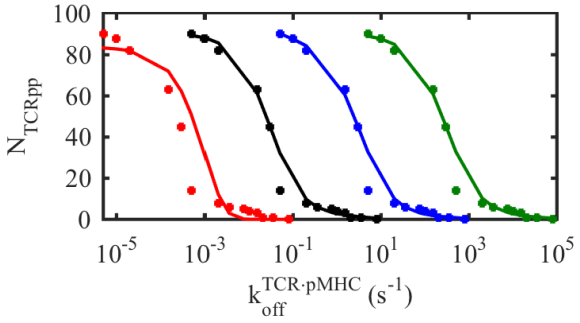


Figure 4.4: Shifting the T cell response threshold. Evolved kinetic parameters produce input-output responses with sharp activation patterns and varied activation thresholds. Optimized input-output responses (solid lines) are compared to target values (boxes) that correspond to the native response (black) and to horizontal shifts in the response curve by factors of 10^{-2} (red), 10^{2} (blue), and 10^{4} (green). For each case, we show a single evolved response curve corresponding to the best fit from many independent runs.

In Fig. 4.5, we show that the TCR network can exhibit input-output behavior that is qualitatively different than that in Fig. 4.4. In Fig. 4.5A, the TCR network exhibits a graded input-output response over four orders of magnitude of $k_{\text{off}}^{\text{TCR·pMHC}}$. Fig. 4.5B shows a more extreme change, in which the input-output relation is "inverted',' with small values of $k_{\text{off}}^{\text{TCR·pMHC}}$ (long binding times) leading to low levels of fully phosphorylated TCR and large values of $k_{\text{off}}^{\text{TCR·pMHC}}$ leading to high levels of phosphorylated TCR. The curve generated using *in silico* evolution clearly captures the key trend of the inverted response. This demonstrates a remarkable flexibility of the fixed-topology TCR network to exhibit qualitatively different input-output behavior. The inverted response was evolved for 750 generations.

For each set of target input-output relations, we considered input values ($k_{\text{off}}^{\text{TCR·pMHC}}$) that span over four orders of magnitude. As illustrated in Figs. 4.4 and 4.5, the EA generated

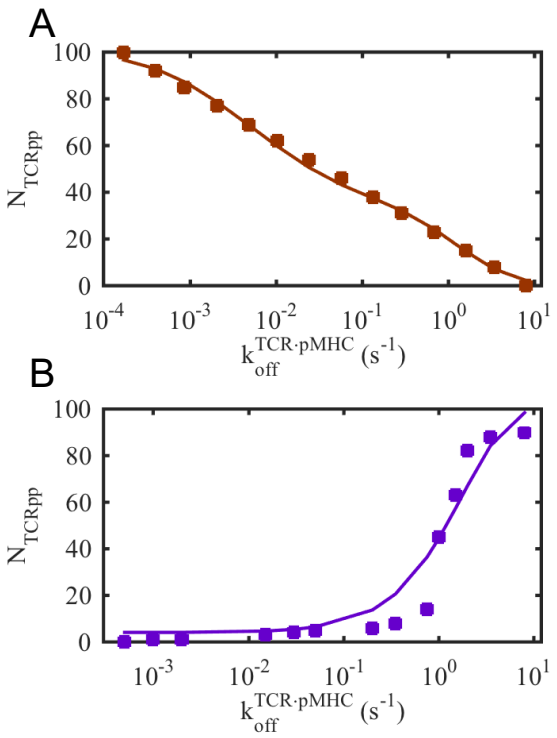


Figure 4.5: Generating graded and inverted signaling responses. Alternative input-output relations were produced using the TCR network: (A) a graded input-output response and (B) an inverted response in which increased TCR-pMHC unbinding rates lead to T cell activation. Solid lines correspond to the best fit from many optimization runs for each set of target values (squares).

parameters that give reasonably good fits of the target behavior. It is interesting to consider the behavior of the input-output relations outside of the range over which the fitness was evaluated. In Fig. 4.6A, examples corresponding to the native and shifted input-output responses are shown over an extended range. For each case, the results from six independent optimization runs are shown. Cases were selected so the the best fit and worst fit are shown, along with four intermediate cases selected at equal intervals along a rank-ordered list of fits. For values of $k_{\text{off}}^{\text{TCR·pMHC}}$ that exceed the range of target values, levels of TCR phosphorylation remain low. For values of $k_{\text{off}}^{\text{TCR·pMHC}}$ less than those considered in the fitted range, levels of TCR phosphorylation remain high. In the extended range of off rates, the response curves maintain their sigmoidal activation profile. Similar behavior is also seen for the other two cases in which the profiles are shifted in Fig. 4.4.

Interestingly, for the inverted response, the input-output behavior is qualitatively different when the range of $k_{\text{off}}^{\text{TCR·pMHC}}$ is extended. In Fig. 4.6B, it can be seen that as $k_{\text{off}}^{\text{TCR·pMHC}}$ increases beyond the target range, the levels of TCR phosphorylation peak and then eventually decrease to a negligible level. For off rates lower than the target range, TCR phosphorylation levels plateau over a range of values that are larger than those in the high $k_{\text{off}}^{\text{TCR·pMHC}}$ range (where they approach a negligible value). This suggests that the network topology predisposes the system towards activation at low values of $k_{\text{off}}^{\text{TCR·pMHC}}$ and an inactive state at high values. The inverted response obtained by evolving kinetic parameters corresponds to a nontrivial input-output shape with a local maximum when viewed over an input parameter range larger than that used in the fitness function. This behavior is consistent with a response that would detect ligands only within a certain range of binding times.

Dimensionality reduction and grouping.

In the results above, we used an evolutionary algorithm to show that the early TCR signaling network can generate a variety of input-output relations. We now investigate constraints imposed on the kinetic parameters by each target input-output relation. To do this, we conducted multiple independent runs of the evolutionary algorithm for each of the target input-output relations, generating between 600 and 2000 different optimized sets of parameters for each set of target values. To gain insight into the structure of the data, we use unsupervised machine learning algorithms on evolved parameter sets. Specifically, t-SNE was used to embed the original data in two dimensions.[123] We then identified groups of data points spatially proximal in the two-dimensional space using a data smoothing procedure to identify distinct high density regions.[115] The high density regions are indicative of parameter sets composed of similar kinetic rates. We also attempted traditional unsupervised machine learning algorithms including principal component analysis (PCA) and k-means clustering but found these methods less useful. For PCA, the first several principal components still left much of the variance in the data unexplained, and because it is a linear dimensionality reduction technique, it may miss important features of the data.

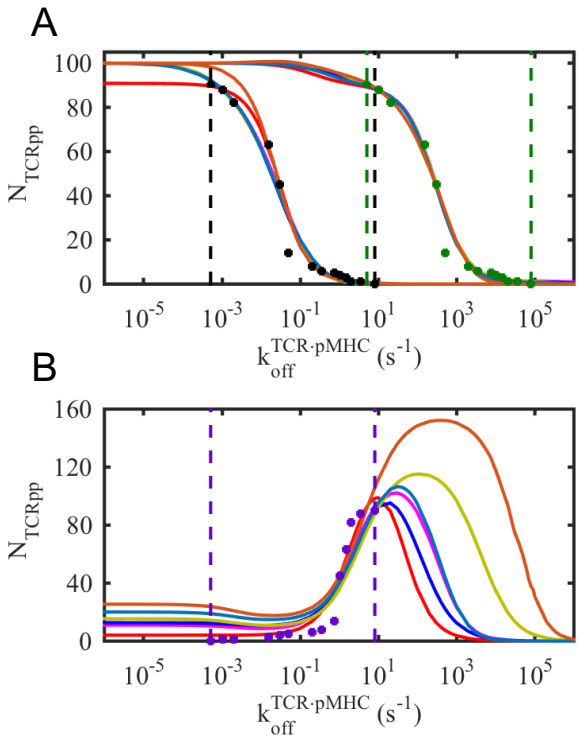


Figure 4.6: Response behavior outside of the target range. The EA optimizes kinetic parameters to fit specified behavior within the range of the target values (demarcated by dashed lines). Here we examine the response outside of the optimized range. (A) The results of multiple optimization runs (solid lines) are shown for the native response (black squares) and the case in which the response is shifted by a factor of 10^4 (green squares). For each case, the response curves maintain the sigmoidal activation profile when the TCR-pMHC off rate is extended beyond the target value range. (B) Analogous results are shown for the inverted activation response. Compared with the previous cases, the response shows qualitatively different behavior outside of the fitted range. As TCR-pMHC unbinding rates extend above the target value range, levels of TCR phosphorylation first peak and then approach zero.

Clustering analysis using k-means clustering suffers the disadvantage that the number of groups must be specified *a priori*.

We begin by considering the four target response curves shown in Fig. 4.4 (the native response curve and three curves with shifted thresholds). Taking the best-performing case from each run of the EA, we obtained 7861 sets of parameters in total. The network model contains seven optimized parameters, meaning that each set of values can be represented as a point in a 7-dimensional parameter space. Fig. 4.7A shows the embedding of each data point in two dimensions. By inspection, it is clear that points tend to cluster by their target value, although there are individual exceptions and some regions that have clustered points of one type in close proximity to clustered points of another type (e.g., points associated with the native target (black) and shift (10^2) target (blue) reside close to one another in the upper right portion of the figure). To categorize different classes of solutions, we associated with each point a Gaussian density kernel and display the resulting density map in Fig. 4.7B. Each peak was assigned a group number, and we performed a hill-climbing procedure from each data point to assign it to a group. The assignments are shown in Figs. 4.7C and D and are analyzed in more detail in the next section.

We now compare the inverted signaling response (Fig. 4.5B) to the native signaling response. In this case, the network model contains eight optimized parameters as we allowed the amplification factor (α) to vary for both the native and inverted response. Using the EA with a variable amplification factor yielded significantly better fits for the inverted target response than those using a fixed amplification factor. When analyzing the fitness, we discarded results from any runs of the EA that yielded a fitness score greater than 220. This cutoff yielded 604 parameter sets for the inverted response and allowed us to analyze only results which we regarded as reasonably good fits. We obtained 2000 parameter sets for the native response with mutable α and chose 604 at random for comparison with the inverted results (this gives a balanced comparison). The results of the t-SNE and subsequent grouping procedure are displayed in Fig 4.8. Interestingly, there are three well-separated groups that emerge in the embedded space, with points associated with the inverted response clustering in two regions. We identified three groups associated with the inverted response (Groups 1-3) and three groups associated with the native response (Groups 4-6). In the next section, we

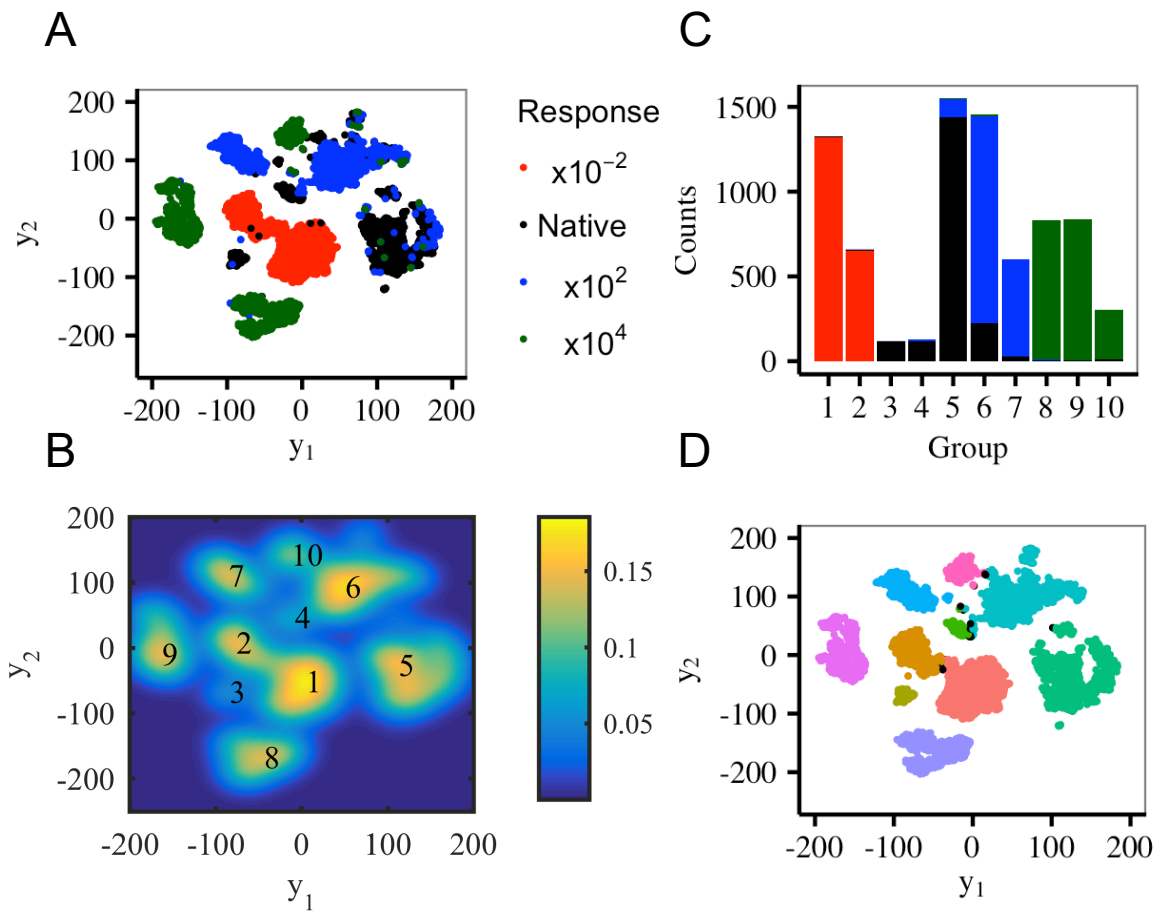


Figure 4.7: Dimensionality reduction of native and shifted T cell response results. (A) Results of the t-SNE method, with points colored according to the target input-output relations (native and shifted thresholds correspond to the target responses shown in Fig. 4.4). The two-dimensional embedding of 7861 data points is shown. Each point corresponds to an independent optimization run for a given target response curve. (B) Gaussian kernel density estimation associated with the embedded points (kernel parameter $\gamma = 19$). Local maxima in the density are labeled by number and used to assign individual points to groups. (C) Number of points assigned to each group, along with the breakdown between the target response curves. (D) Points are colored according to their assigned group. Black points were not assigned to a group.

analyze the distribution of parameters for each group and show that Group 1 is characterized by significantly different values of α compared with Groups 2 and 3.

Analysis of groups.

In the previous section, we identified distinct groups of data points by performing t-SNE and mapping points onto local maxima in the resulting density profile. We now compare characteristics of the various groups to gain insight into constraints placed on kinetic parameters by the various targeted input-output curves.

Fig. 4.9 shows the distributions of each kinetic rate for each group identified in Fig. 4.7 (the combined native and shifted response results). Each violin plot corresponds to a particular parameter and shows the distribution of the parameter for each group identified previously. Although the t-SNE procedure used parameter sets involving the kinetic parameters and not the fitness, it is instructive to compare the fitnesses associated with each group. The general trend is that thresholds occurring at higher values of $k_{\text{off}}^{\text{TCR·pMHC}}$ have lower scores, implying better fits. For example, Groups 1 and 2 have the highest fitness score and consist almost entirely of points associated with the signaling threshold shifted by a factor of 10^{-2}. Groups 7, 8, and 9 have low fitness scores and consist of points associated with the shift by a factor of 10^4. In other words, we obtain better fits of the sharp response threshold when the characteristic TCR-pMHC lifetime is short.

The violin plots in Fig. 4.9 are useful for (i) characterizing kinetic parameters associated with a single target response and (ii) comparing kinetic parameters associated with different responses. Groups 1 and 2, which are associated with a threshold shifted by a factor of 10^{-2}, are most clearly distinguished by the pMHC-coreceptor binding rate and the TCR-coreceptor binding rate (recall that this interaction is mediated by the protein kinase Lck). Group 1 is associated with a low pMHC-coreceptor binding rate but a relatively high TCR-coreceptor binding rate, whereas the relationships are reversed in group 2. There is also a strong constraint on the TCR-pMHC binding rate in these groups: It lies at the very bottom of the allowed range, which is consistent with having to distinguish ligands with a small off rate. In order to distinguish between long-lived TCR-pMHC complexes, the on rate must be sufficiently small so that the receptors are not saturated by persistently-bound pMHC. While

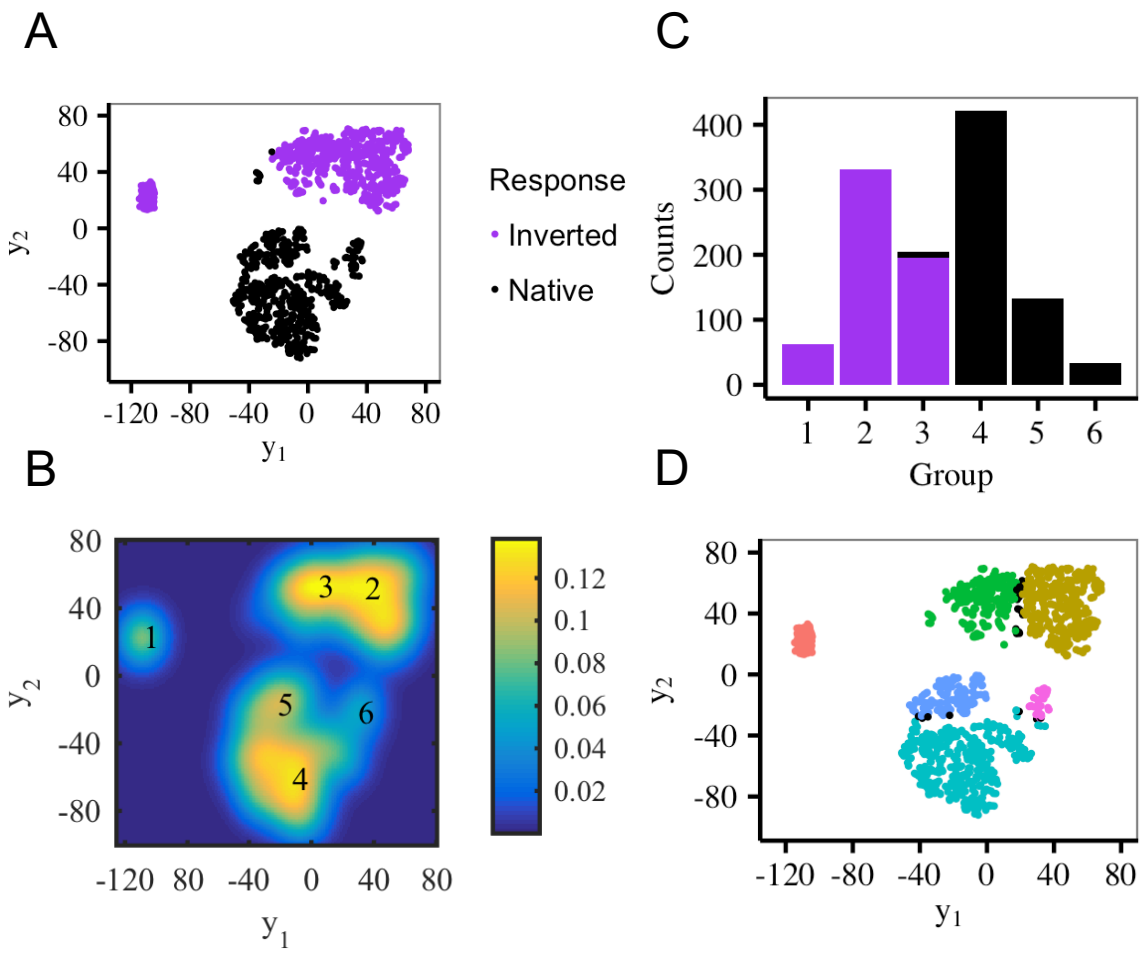


Figure 4.8: Dimensionality reduction of native and inverted T cell responses. (A) Results of the t-SNE method, with points colored according to the target input-output relations. In contrast with Fig. 4.7, the amplification factor (α) was regarded as a mutable parameter in these results. The two-dimensional embedding of 1208 data points is shown. (B) Density profile of embedded points ($\gamma = 10$). (C) Groups 1, 2, and 3 consist primarily of inverted response results, while groups 4, 5, and 6 consist exclusively of native response results. (D) Points are colored according to their group assignment.

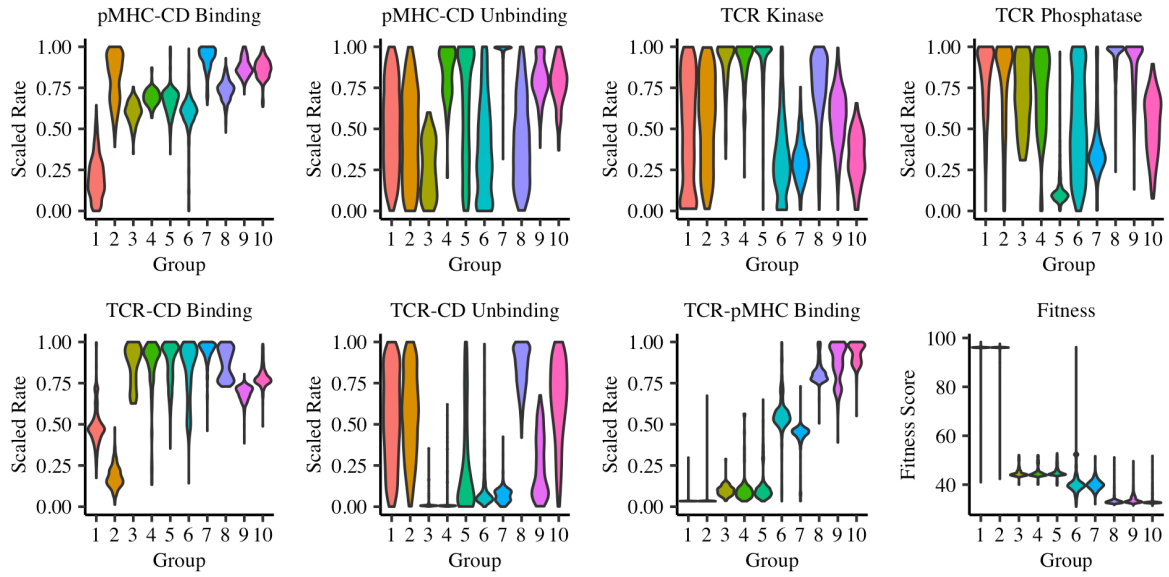


Figure 4.9: Distributions of parameters for groups identified in native and shifted T cell response results. Violin plots show the distribution of scaled parameters and fitness scores for the groups identified in Fig. 4.7, which includes the native and shifted threshold responses. The rate is labeled above each set of violin plots, and the parameter scaling is described in the Methods. Note that t-SNE was performed on the kinetic parameters and not the fitness; fitness was included for comparison between the groups.

the phosphatase rate tends to be high, the remaining parameters show broad distributions for Groups 1 and 2, meaning that there is not a favored regime for these parameters.

When comparing all groups in Fig. 4.9, a clear trend emerges with the TCR-pMHC binding rate increasing with increasing group number. The groups are ordered from lowest to highest value of $k_{\text{off}}^{\text{TCR·pMHC}}$ at the activation threshold. This is again consistent with the idea that the TCR-pMHC binding kinetics should balance the unbinding rate characteristic of the activation threshold so that the number of TCRs bound to pMHC is responsive to $k_{\text{off}}^{\text{TCR·pMHC}}$ near the threshold.

It is also interesting to observe the trend involving TCR-coreceptor kinetics in Groups 3 - 10. The binding rate of TCR and coreceptor is fast, with Groups 9 and 10 (associated with shift (10^4)) having moderately slower binding kinetics. The unbinding rates are slow for Groups 3 - 7 (primarily native and shift (10^2)) and highly variable for Groups 8 - 10 (primarily shift (10^4)). This suggests that coreceptor recruitment to the TCR by means of direct TCR-coreceptor interaction is important for native and shift (10^2) responses.

Fig. 4.10 shows the parameter ranges sampled by the six groups identified for the native and inverted response curves in Fig. 4.8. The parameter α was varied in the EA, and thus is included with the results. Groups 1 - 3 correspond primarily to the parameter sets that targeted the inverted response, and Groups 4 - 6 include only parameter sets that targeted the native response. The distribution of fitness scores highlights that the EA produced better fits (with substantially better scores) for the native case compared with the inverted case. For the inverted response, we analyzed only parameter sets with a fitness score less than 220, which we still regard as reasonably good fits.

The behavior of Groups 1 - 3 suggests that targeting the inverted response imposes strong constraints on a subset of the kinetic parameters. In particular, it can be seen that (i) the rate of phosphatase activity and the off rate of the pMHC-coreceptor bond reside at the low extreme of possible values, and (ii) the kinase rate is constrained to be highly active. Other kinetic parameters remain fairly constrained, but exhibit differences between groups. For example, compared to Groups 2 and 3, Group 1 has small α, high TCR-coreceptor binding and pMHC-coreceptor binding rates, and low TCR-coreceptor unbinding and TCR-pMHC binding rates. Groups 2 and 3 can be distinguished primarily by relatively small differences in distributions of TCR-coreceptor unbinding rates, TCR-pMHC binding rates, and α. Taken together, these results identify important kinetic parameters (i.e., the pMHC-coreceptor off rate, kinase rate, and phosphatase rate) that would need to be modulated to engineer this signaling response using the defined TCR network topology.

Groups 4 - 6 show less pronounced constraints on kinetic parameters. Compared with the inverted response, they exhibit higher TCR-coreceptor binding rates, higher pMHC-coreceptor unbinding (with broad distributions), lower kinase rates, and higher phosphatase rates. Groups 4 and 5 are distinguished by their phosphatase rates, and, by comparison to both, Group 6 has high TCR-pMHC binding rates and low kinase rates. Groups 4 - 6 can also be compared to the distribution of parameters for Groups 2 - 5 in Fig. 4.7, in which the native response was evolved with a fixed value of $\alpha = 10^5$. When allowed to vary, α tended to sample from the higher end of the allowed range, although the value used in Fig. 4.7 is within the range sampled. The primary differences are that when α is varied,

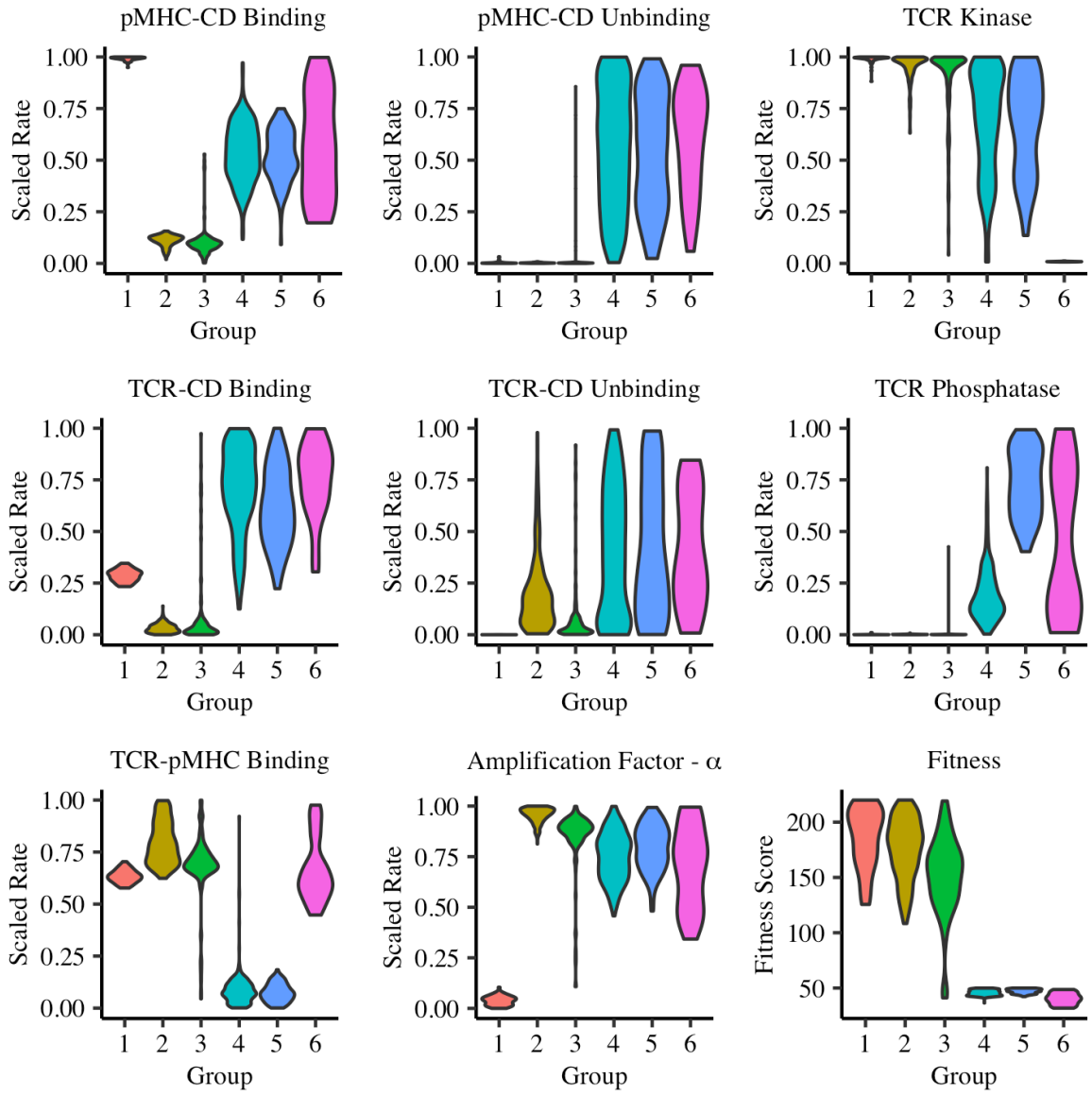


Figure 4.10: Distributions of parameters for groups identified in native and inverted T cell response results. Violin plots show the distribution of scaled parameters and fitness scores for the groups identified in Fig. 4.8, which includes the native and inverted threshold responses. Note that in these results the amplification factor was treated as a mutable parameter and the fitness was not included in the t-SNE procedure.

TCR-coreceptor unbinding rates sample a much larger range and a set of solutions emerges in which the kinase rate is low and the TCR-pMHC binding rate is high.

4.3.2 Evolution and analysis of the robust oscillator network

Given the robust oscillator network described above, we applied our computational methodology to evolve and explore different oscillatory patterns in response to a change in input. Fig. 4.11 demonstrates the ability of the network to generate three different periods of oscillations. Each trajectory is one example of a response curve generated by our EA for each targeted period. We obtained approximately 500 evolved parameter sets for each targeted period.

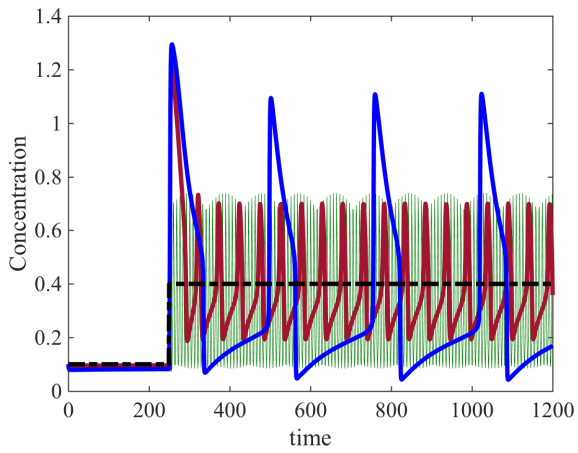


Figure 4.11: Evolving the robust oscillator network. Examples of network outputs that were optimized for the concentration of species A to oscillate with periods of 10 (green), 50 (red), and 250 (blue). The dashed black line corresponds to the input.

As before, to gain insight into the results obtained from the EA, we used t-SNE to embed the data in two dimensions. Fig. 4.12A shows the results of embedding 1525 data points from independent runs of the EA. By inspection, the points tend to cluster largely by their period, although there are regions in which the three populations are mixed. Fig. 4.12B shows the resulting density map, and Fig. 4.12C shows the content of each group identified by the grouping analysis. It is interesting to note that Group 3 has a mixed population with approximately equal numbers of points from each targeted period. Other groups are

dominated by a data points originating from a single targeted period. Fig. 4.12D shows the embedded points identified by group number.

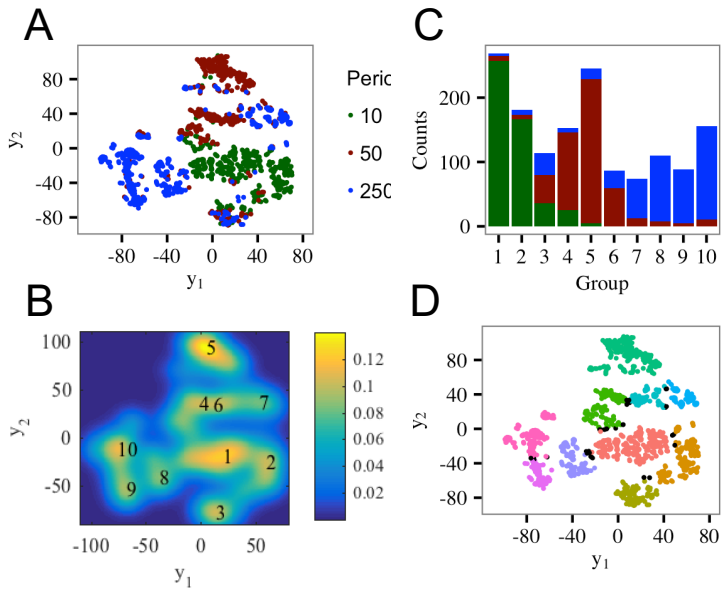


Figure 4.12: Dimensionality reduction of robust oscillator results. (A) Results of the t-SNE method, with points colored according to the targeted oscillation period. The two-dimensional embedding of 1525 data points is show. (B) Density profile of embedded points ($\gamma = 10$). (C) Number of points assigned to each group, along with the breakdown between the targeted oscillatory periods curves. (D) Points are colored according to their assigned group. Black points were not assigned to a group.

In Figure 4.13, we display the distributions of select parameters for each group (n_1 and n_2 are not included since they are restricted to integer values). Group 3 is interesting as it contains a mixture of points targeting the three different periods. It is distinguished by large k_2, which is associated with the negative regulation of A by C. The distribution of K_a (the activation coefficient associated with the positive autoregulation of A) is also low. This suggests a region of parameter space that, while distinct from other groups, one would want to avoid if engineering a network, as similar kinetic parameters lead to dramatically different outputs.

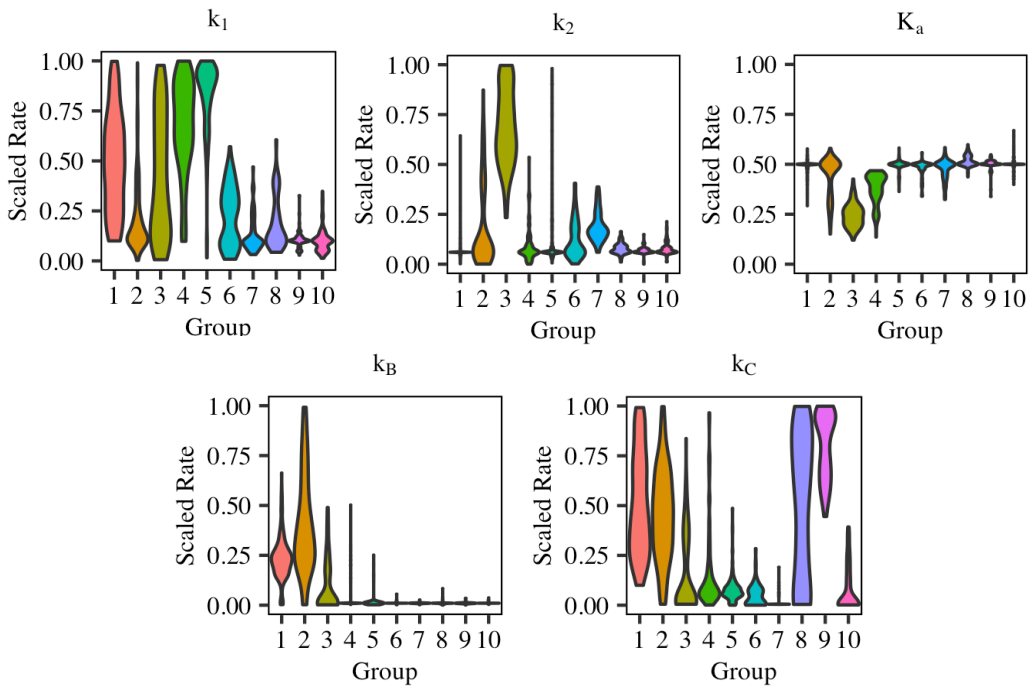


Figure 4.13: Distributions of parameters for groups identified in the robust oscillator results. Violin plots show the distribution of scaled parameters and fitness scores for the groups identified in Fig. 4.12, which includes three different periods of oscillation (10, 50, and 250). Note that the Hill coefficients (n_1 and n_2) were included in the t-SNE analysis but are not shown here.

4.4 Discussion

In this paper, we used *in silico* evolution as a supervised machine learning algorithm to explore the ability of fixed-topology signaling networks to generate a variety of input-output behavior. For each targeted input-output response, we conducted many independent runs of an evolutionary algorithm (EA) and used unsupervised machine learning in the form of nonlinear dimensionality reduction and density-based clustering to identify features of parameter sets producing the targeted behavior.

Our first focus was a model of the earliest portion of the TCR signaling network, which has been shown previously to exhibit sharp discrimination of pMHC based on the binding time of the TCR-pMHC complex.[11, 70] We sought to investigate the plasticity of the network in terms of generating varied output (TCR phosphorylation) in response to TCR-pMHC unbinding rates. We showed that with the fixed network topology, we could obtain different

responses by using an EA to identify sets of kinetic parameters giving rise to the targeted behavior. For example, we showed that the TCR network could generate (i) sigmoidal activation curves with the activation threshold shifted over 6 order of magnitude, (ii) a graded response curve, and (iii) a more extreme "inverted" response, in which TCR phosphorylation was high for short TCR-pMHC binding times. The inverted response was the most difficult to evolve: It required the most generations of evolution and yielded the highest average fitness values. Additionally, when viewed over a range of TCR-pMHC unbinding rates that extended beyond those specified in the fitness function, the inverted response had a shape that peaked at large values of $k_{\text{off}}^{\text{TCR·pMHC}}$ before decaying. Thus, while the evolutionary algorithm fit the inverted response over the specified range, the actual response is one in which an intermediate range of $k_{\text{off}}^{\text{TCR·pMHC}}$ would activate the T cell but high and low values would not. This is suggestive of a detector that would sense ligands only within a certain range of binding times. The results also suggest that the network is predisposed to generating responses in which short binding times yield low levels of TCR phosphorylation. This is consistent with the biological function of T cells: They must respond to antigens but not to self ligands, which have short binding times.

Conducting many independent runs of the EA for each input-output relation allowed us to explore parameter space and to identify many different sets of parameters that yielded good fits of the targeted behavior. We used t-SNE, a nonlinear dimensionality reduction method, to embed the data in two dimensions. This facilitated the identification of groups of results that were similar to one another but different (separated) from results in other groups. The distributions of parameters for each group can be compared to identify constraints placed on the parameters by the targeted response curves. It can also be used to identify correlations between parameters in a group. For example, in Fig. 4.10, Group 1 (inverted response) exhibits a high pMHC-coreceptor binding rate, which is associated with highly constrained values of several other parameters such as the TCR-coreceptor unbinding rate and α. Groups 2 and 3 have low values of pMHC-coreceptor binding rates, which are associated with parameters that are constrained to different ranges than those in Group 1.

We also considered a network topology containing both positive and negative feedback that can exhibit oscillations in certain parameter regimes.[23] We used our EA to evolve

responses in which the output variable oscillated with three different periods. The t-SNE and grouping analysis was again successful at identifying groups with different sets of characteristics. Interestingly, one of the groups identified contained approximately equal numbers of parameter sets from each targeted response. We did not observe this feature in the TCR network, and it suggests a region of parameter space in which experimental design would be difficult, since similar sets of parameters lead to significantly different behavior.

4.5 Conclusions

The approach described in this paper can be used to guide the design of kinetic rates for any signaling network of interest, including those with larger numbers of parameters. Potential applications include designing modified cells or synthetic systems *ex vivo*.[89, 99] A number of approaches exist for engineering kinetics of protein interactions.[9, 34, 43] In Hoerter et al., the binding interactions between MHC proteins and coreceptors were modulated over a range of values to explore the effects of coreceptor binding on T cell activation.[70] Having a computationally-guided set of kinetic regimes could help to streamline protein design and engineering. The approaches taken can also be used to constrain kinetic parameters for networks of known topology with data that is measured at steady state (e.g., response curves), and it can also be applied to dynamical data by using an alternative definition of the fitness function, as shown for the robust oscillator network.

Biologically, T cells coordinate the adaptive immune response by responding to molecular signatures of pathogens, and we chose to focus on early TCR signaling because T cells are prime candidates for synthetic biology applications. We have shown here that the earliest part of the TCR signaling network exhibits substantial plasticity in generating a variety of response curves. This can be useful from the perspective of designing and engineering T cell responses, and it also suggests the evolutionary potential of the network topology to perform functions other than sharp ligand discrimination at a particular threshold value of the binding time. Taken together, our results show that combining *in silico* evolution with dimensionality reduction and density-based clustering provides a useful way to guide the design of signaling networks with novel responses.

Chapter 5

Conclusions

Throughout this dissertation, we demonstrated the immense value to be distilled from data mining results from *in silico* models. Using systematic parameter exploration procedures, individual components within complex biological systems can be isolated and studied in a manner that is inaccessible to *in vivo*/*in vitro* models. For biological systems with poorly defined responses, a combinatoric approach can be applied to test various parameter values over each range of interest. In cases where the response profile is known, supervised machine learning can be applied to learn possible parameter values that produce observed responses. Regardless of the parameter exploration method employed, large quantities of simulation results can be produced. Data analytic techniques, including machine learning algorithms, can be applied to this data to discover critical mechanisms that govern system dynamics.

The process of data mining results produced from *in silico* parameter exploration is a versatile research technique. We applied it while investigating the role of aerobic glycolysis in tumor progression (Ch. 2), testing signal transduction network topologies associated with the growth response of *Arabidopsis* to ethylene exposure (Ch. 3), and testing fixed-topology signal transduction networks to kinetic redesign for use in synthetic biology (Ch. 4). In each instance, we obtained mechanistic insight into the experimental system which allowed us to refine *in silico* observations into testable hypotheses.

5.1 Exploring cancer growth and altered metabolism in heterogeneous environments (Chapter 2)

Within the heterogeneous tumor microenvironment, the role of some cancer hallmarks remain unclear . Specifically, there is still no consensus regarding the consequences of the switch to aerobic glycolysis has during tumor progression. To determine what benefits, if any, are conferred by this metabollic transition, we used a hybrid cellular automata (HCA) model to simulate tumor progression in a vascularized field of healthy cells. By varying the magnitude of several cell traits associated with aerobic glycolysis, we discovered four distinct growth regimes: isolated clustering, continuous growth, domain seeding, and tumor trapping. Each growth regime was associated with distinct combinations of local toxicity effect, malignant migration, and metabolite consumption rates.

Results from our simulations revealed the importance of including glucose consumption in simulations. Previous HCA models of vascularized tumors excluded glucose as a metabolite. Inclusion of glucose in our simulations led to the emergene of tumor trapping. Tumor trapping is similar to *in vivo* growth observed at many secondary tumor sites that form during metastasis and it has not observed in other simulations. It has been hypothesized that secondary tumor growth remains confined at a microscopic scale until initiation of the angiogenic switch, i.e. the formation of new vascular tissue. Our model results suggested an alternative path of secondary tumor growth that does not rely on angiogenesis. By moderately increasing the local toxicity effect and migration from the levels observed in tumor trapping, domain seeding produced tumor progression without the formation of new vascular sites. Domain seeding suggests an alternative explanation for the observed tissue bias exhibited by secondary tumors. It may be possible that, in the absence of the angiogenic switch, secondary tumors remain dormant unless implanted in tissues where conditions support tumor progression via domain seeding.

5.2 Analyzing signal transduction networks underlying ethylene growth response kinetics (Chapter 3)

Ethylene is a plant hormone that influences many plant behaviors. One of these behaviors is a reduction of growth rate observed in dark-grown *Arabidopsis* seedlings. Signal transduction network topologies have been proposed to explain response curves produced using dynamic image analysis. These network topologies provide qualitative descriptions consistent with observed growth responses. However, kinetic parameters governing network dynamics are unknown. This has prevented quantitative evaluation of proposed network.

Using evolutionary algorithms (EAs) for supervised machine learning, we discovered kinetic parameter sets that produce observed growth response profiles for two proposed network topologies. The first network contains coherent feedforward and negative feedback loops (CFF/NFB network) while the second utilizes a positive feedback loop (PFB network). Using evolved parameter values, we simulated the response produced by each network component. Network components that exhibited conserved responses served to indicate critical mechanisms governing growth response. Specifically, EIN2, a well-studied gene in *Arabidopsis*, showed a conserved response within the CFF/NFB network. This was also true for the PFB network. However, critical differences were observed in the dynamics of EIN2 between the two networks. Our results suggest that dynamic *in vivo* measurements of EIN2 activity could provide additional support for determining which, if any, of the proposed network topologies are correct.

5.3 Redesigning signal transduction network response dynamics *in silico* (Chapter 4)

The rational design of signal transduction networks is a central goal of synthetic biology. Network topology and kinetics are both critical components in the design process. The topologies for a large number of signal transduction networks are already known. However, the kinetics that drive network response and the tolerance for kinetic redesign of most

networks is not understood. As a first step in redesigning the kinetic response of signal transduction networks, we developed an *in silico* procedure to test the the ability of fixed-topology networks to generate a range of system responses. By using evolutionary algorithms (EAs) for supervised machine learning, we discovered parameters that produce desired network reponses. Using nonlinear unsupervised machine learning, clusters of similiar parameters were identified and compared to discover network mechanisms critical for each designed response.

These methods were applied to T cell receptor (TCR) activation and were used to generate (i) sigmoidal activation curves with the activation threshold shifted over 6 orders of magnitude, (ii) a graded response curve, and (iii) a more extreme response in which TCR activation was high for short TCR-antigen binding times. We also applied the methods to produce alternative periods of oscillation from the robust oscillator network. The analysis of parameters for each network indicated a subset of parameter values that were ideal for each designed response. Our method provides a tool by which one can focus their efforts on parameter regimes that are known to facilitate a desired network response during the *in vivo*/*in vitro* design process.

5.4 Closing

The feedback between computational and experimental biologists is a cyclical process of hypothesis refinement. As computational tools and resources grow, a greater proportion of this cycle can be placed in the hands of computational biologists. In their hands, high throughput simulations can generate vast quantities of data under precisely controlled conditions. By coupling this with increasingly powerful data mining approaches derived from the field of machine learning, the value produced from biological research has the potential for immense growth.

Bibliography

[1] Abeles, F. B., Morgan, P. W., and Saltveit Jr, M. E. (1992). Ethylene in plant biology. *San Diego: Academic*. 45

[2] Alarcon, T., Byrne, H., and Maini, P. (2004). A mathematical model of the effects of hypoxia on the cell-cycle of normal and cancer cells. *J. Theor. Biol.*, 229(3):395–411. 14

[3] Alarcón, T., Byrne, H. M., and Maini, P. K. (2003). A cellular automaton model for tumour growth in inhomogeneous environment. *J. Theor. Biol.*, 225(2):257–274. 15

[4] Alonso, J. M., Hirayama, T., Roman, G., Nourizadeh, S., and Ecker, J. R. (1999). EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. *Science*, 284(5423):2148–2152. 49

[5] Alpaydin, E. (2014). *Introduction to machine learning*. MIT press. 3

[6] Ambrosi, D. and Mollica, F. (2004). The role of stress in the growth of a multicell spheroid. *Journal of mathematical biology*, 48(5):477–499. 14

[7] Amir, E.-a. D., Davis, K. L., Tadmor, M. D., Simonds, E. F., Levine, J. H., Bendall, S. C., Shenfeld, D. K., Krishnaswamy, S., Nolan, G. P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.*, 31(6):545–52. 78

[8] An, F., Zhao, Q., Ji, Y., Li, W., Jiang, Z., Yu, X., Zhang, C., Han, Y., He, W., and Liu, Y. (2010). Ethylene-induced stabilization of ETHYLENE INSENSITIVE3 and EIN3-LIKE1 is mediated by proteasomal degradation of EIN3 binding F-box 1 and 2 that requires EIN2 in Arabidopsis. *Plant Cell*, 22(7):2384–2401. 46

[9] Ang, J., Harris, E., Hussey, B. J., Kil, R., and McMillen, D. R. (2013). Tuning response curves for synthetic biology. *ACS Synth. Biol.*, 2(10):547–567. PMID: 23905721. 105

[10] Arisi, I., Cattaneo, A., and Rosato, V. (2006). Parameter estimate of signal transduction pathways. *BMC Neurosci.*, 7(1):1. 50, 78

[11] Artyomov, M. N., Lis, M., Devadas, S., Davis, M. M., and Chakraborty, A. K. (2010). CD4 and CD8 binding to MHC molecules primarily acts to enhance Lck delivery. *Proc. Natl. Acad. Sci. USA*, 107(39):16916–16921. 80, 81, 83, 89, 90, 103

[12] Auliac, C., Frouin, V., Gidrol, X., and d'Alché Buc, F. (2008). Evolutionary approaches for the reverse-engineering of gene regulatory networks: A study on a biologically realistic dataset. *BMC Bioinform.*, 9(1):1–14. 50, 78

[13] Back, T. (1996). *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press. 6

[14] Bäck, T. and Schwefel, H.-P. (1993). An overview of evolutionary algorithms for parameter optimization. *Evol. Comp.*, 1(1):1–23. 50, 78

[15] Bakshi, A., Wilson, R. L., Lacey, R. F., Kim, H., Wupalapapati, S. K., and Binder, B. (2015). Identification of regions in the receiver domain of the ETHYLENE RESPONSE1 ethylene receptor of Arabidopsis important for functional divergence. *Plant Physiol.*, 169:219–232. 46

[16] Becher, B., Schlitzer, A., Chen, J., Mair, F., Sumatoh, H. R., Teng, K. W. W., Low, D., Ruedl, C., Riccardi-Castagnoli, P., Poidinger, M., et al. (2014). High-dimensional analysis of the murine myeloid cell system. *Nat. Immunol.*, 15(12):1181–1189. 78

[17] Bhaskar, H., Hoyle, D. C., and Singh, S. (2006). Machine learning in bioinformatics: A brief survey and recommendations for practitioners. *Computers in biology and medicine*, 36(10):1104–1125. 8

[18] Binder, B. M. and Bleecker, A. (2002). A model for ethylene receptor function and 1-methylcyclopropene action. In *Acta Hortic.*, pages 177–187. 74

[19] Binder, B. M., Mortimore, L. A., Stepanova, A. N., Ecker, J. R., and Bleecker, A. B. (2004a). Short-term growth responses to ethylene in Arabidopsis seedlings are EIN3/EIL1 independent. *Plant Physiol.*, 136(2):2921–2927. 46, 47, 48, 49

[20] Binder, B. M., O'Malley, R. C., Wang, W., Moore, J. M., Parks, B. M., Spalding, E. P., and Bleecker, A. B. (2004b). Arabidopsis seedling growth response and recovery to ethylene. A kinetic analysis. *Plant Physiol.*, 136(2):2913–2920. 46, 48, 74

[21] Binder, B. M., Walker, J. M., Gagne, J. M., Emborg, T. J., Hemmann, G., Bleecker, A. B., and Vierstra, R. D. (2007). The Arabidopsis EIN3 binding F-Box proteins EBF1 and EBF2 have distinct but overlapping roles in ethylene signaling. *Plant Cell*, 19(2):509–523. 46

[22] Bleecker, A. B., Estelle, M. A., Somerville, C., and Kende, H. (1988). Insensitivity to ethylene conferred by a dominant mutation in Arabidopsis thaliana. *Science*, 241(4869):1086–1089. 45

[23] Brandman, O. and Meyer, T. (2008). Feedback loops shape cellular signals in space and time. *Science*, 322(5900):390–395. 77, 81, 84, 86, 104

[24] Bray, D. and Lay, S. (1994). Computer simulated evolution of a network of cell-signaling molecules. *Biophys. J.*, 66(4):972–977. 50, 78

[25] Bray, D., Levin, M. D., and Morton-Firth, C. J. (1998). Receptor clustering as a cellular mechanism to control sensitivity. *Nature*, 393(6680):85–88. 74

[26] Bribiesca, E. (2008). An easy measure of compactness for 2D and 3D shapes. *Pattern Recogn.*, 41(2):543–554. 24

[27] Brownlie, R. J. and Zamoyska, R. (2013). T cell receptor signalling networks: branched, diversified and bounded. *Nat. Rev. Immunol.*, 13(4):257–269. 79

[28] Burg, S. P. (1973). Ethylene in plant growth. *Proc. Natl. Acad. Sci. USA*, 70(2):591–597. 46

[29] Butler, J., Mackay, F., Denniston, C., and Daley, M. (2016). Halting the hallmarks: a cellular automaton model of early cancer growth inhibition. *Nat. Comput.*, 15(1):15–30. 15

[30] Carreau, A., Hafny-Rahbi, B. E., Matejuk, A., Grillon, C., and Kieda, C. (2011). Why is the partial oxygen pressure of human tissues a crucial parameter? Small molecules and hypoxia. *J. Cell. Mol. Med.*, 15(6):1239–1253. 17

[31] Chakraborty, A. K. and Weiss, A. (2014). Insights into the initiation of TCR signaling. *Nat. Immunol.*, 15(9):798–807. 79

[32] Chambers, A. F., Groom, A. C., and MacDonald, I. C. (2002). Metastasis: dissemination and growth of cancer cells in metastatic sites. *Nat. Rev. Cancer*, 2(8):563. 44

[33] Chen, R., Binder, B. M., Garrett, W. M., Tucker, M. L., Chang, C., and Cooper, B. (2011). Proteomic responses in Arabidopsis thaliana seedlings treated with ethylene. *Mol. Biosyst.*, 7(9):2637–2650. 46

[34] Chen, T. S. and Keating, A. E. (2012). Designing specific protein–protein interactions using computation, experimental library screening, or integrated methods. *Protein Sci.*, 21(7):949–963. 105

[35] Chickarmane, V., Paladugu, S. R., Bergmann, F., and Sauro, H. M. (2005). Bifurcation discovery tool. *Bioinformatics*, 21(18):3688–3690. 78

[36] Christians, M. J., Gingerich, D. J., Hansen, M., Binder, B. M., Kieber, J. J., and Vierstra, R. D. (2009). The BTB ubiquitin ligases ETO1, EOL1 and EOL2 act collectively to regulate ethylene biosynthesis in Arabidopsis by controlling type2 ACC synthase levels. *Plant J.*, 57(2):332–345. 46

[37] DeBerardinis, R. J. and Chandel, N. S. (2016). Fundamentals of cancer metabolism. *Sci. Adv.*, 2(5):e1600200. 13, 14

[38] Díaz, J. and Álvarez-Buylla, E. R. (2006). A model of the ethylene signaling pathway and its gene response in Arabidopsis thaliana: pathway cross-talk and noise-filtering properties. *Chaos*, 16(2):023112. 49

[39] Díaz, J. and Alvarez-Buylla, E. R. (2009). Information flow during gene activation by signaling molecules: ethylene transduction in Arabidopsis cells as a study system. *BMC Syst. Biol.*, 3(1–13):1. 49

[40] Ding, D., Moskowitz, S. I., Li, R., Lee, S. B., Esteban, M., Tomaselli, K., Chan, J., and Bergold, P. J. (2000). Acidosis induces necrosis and apoptosis of cultured hippocampal neurons. *Exp. Neurol.*, 162(1):1–12. 14

[41] Donaldson, J. (2012). *tsne: T-distributed Stochastic Neighbor Embedding for R (t-SNE)*. R package version 0.1-2. 87

[42] Edelman, L. B., Eddy, J. A., and Price, N. D. (2010). In silico models of cancer. *WIRES Syst. Biol. Med.*, 2(4):438–459. 15

[43] Eriksen, D. T., Lian, J., and Zhao, H. (2014). Protein design for pathway engineering. *J. Struct. Biol.*, 185(2):234–242. 105

[44] Feinerman, O., Germain, R. N., and Altan-Bonnet, G. (2008). Quantitative challenges in understanding ligand discrimination by alphabeta T cells. *Mol. Immunol.*, 45(3):619–31. 79, 80

[45] Feng, S., Ollivier, J. F., Swain, P. S., and Soyer, O. S. (2015). BioJazz: in silico evolution of cellular networks with unbounded complexity using rule-based modeling. *Nucleic Acids Res.*, 43(19):e213. 50, 78

[46] Ferguson, A. (2017). Machine learning and data science in soft materials engineering. *Journal of Physics: Condensed Matter*. 8

[47] Folkman, J. (2002). Role of angiogenesis in tumor growth and metastasis. In *Semin. Oncol.*, volume 29, pages 15–18. Elsevier. 43

[48] François, P. and Hakim, V. (2004). Design of genetic networks with specified functions by evolution in silico. *Proc. Natl. Acad. Sci. USA*, 101(2):580–585. 50, 78

[49] François, P., Hakim, V., and Siggia, E. D. (2007). Deriving structure from evolution: metazoan segmentation. *Mol. Sys. Biol.*, 3(1). 78

[50] Freyer, J., Tustanoff, E., Franko, A., and Sutherland, R. (1984). In situ oxygen consumption rates of cells in V-79 multicellular spheroids during growth. *J. Cell. Physiol.*, 118(1):53–61. 17

[51] Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York. 3, 4, 5, 6

[52] Gagne, J. M., Smalle, J., Gingerich, D. J., Walker, J. M., Yoo, S.-D., Yanagisawa, S., and Vierstra, R. D. (2004). Arabidopsis EIN3-binding F-box 1 and 2 form ubiquitin-protein ligases that repress ethylene action and promote growth by directing EIN3 degradation. *Proc. Natl. Acad. Sci. USA*, 101(17):6803–6808. 46

[53] Gamble, R. L., Qu, X., and Schaller, G. E. (2002). Mutational analysis of the ethylene receptor ETR1. role of the histidine kinase domain in dominant ethylene insensitivity. *Plant Physiol.*, 128(4):1428–1438. 74

[54] Gao, Z., Wen, C.-K., Binder, B. M., Chen, Y.-F., Chang, J., Chiang, Y.-H., Kerris, R. J., Chang, C., and Schaller, G. E. (2008). Heteromeric interactions among ethylene receptors mediate signaling in Arabidopsis. *J. Biol. Chem.*, 283(35):23801–23810. 46, 47

[55] Gatenby, R., Smallbone, K., Maini, P., Rose, F., Averill, J., Nagle, R., Worrall, L., and Gillies, R. (2007). Cellular adaptations to hypoxia and acidosis during somatic evolution of breast cancer. *Brit. J. Cancer*, 97(5):646. 15, 21

[56] Gatenby, R. A., Gawlinski, E. T., Gmitro, A. F., Kaylor, B., and Gillies, R. J. (2006). Acid-mediated tumor invasion: a multidisciplinary study. *Cancer Res.*, 66(10):5216–5223. 14

[57] Gatenby, R. A. and Gillies, R. J. (2004). Why do cancers have high aerobic glycolysis? *Nat. Rev. Cancer*, 4(11):891–899. 14

[58] Gerlee, P. and Anderson, A. R. (2008). A hybrid cellular automaton model of clonal evolution in cancer: the emergence of the glycolytic phenotype. *J. Theor. Biol.*, 250(4):705–722. 17, 21, 22

[59] Gevertz, J. L. and Torquato, S. (2006). Modeling the effects of vasculature evolution on early brain tumor growth. *J. Theor. Biol.*, 243(4):517–531. 14

[60] Gillies, R. J., Raghunand, N., Garcia-Martin, M. L., and Gatenby, R. A. (5). pH imaging. *IEEE Eng. Med. Biol.*, 23(57–64):2004. 14

[61] Goeschl, J. D. and Kays, S. J. (1975). Concentration dependencies of some effects of ethylene on etiolated pea, peanut, bean, and cotton seedlings. *Plant Physiol.*, 55(4):670–677. 46

[62] Groebe, K., Erz, S., and Mueller-Klieser, W. (1994). Glucose diffusion coefficients determined from concentration profiles in EMT6 tumor spheroids incubated in radioactively labeled L-glucose. In *Oxygen Transport to Tissue XVI*, pages 619–625. Springer. 15, 17

[63] Gruetter, R., Ugurbil, K., and Seaquist, E. R. (1998). Steady-state cerebral glucose concentrations and transport in the human brain. *J. Neurochem.*, 70(1):397–408. 19

[64] Guo, H. and Ecker, J. R. (2003). Plant responses to ethylene gas are mediated by SCF EBF1/EBF2-dependent proteolysis of EIN3 transcription factor. *Cell*, 115(6):667–677. 46

[65] Guzman, P. and Ecker, J. R. (1990). Exploiting the triple response of Arabidopsis to identify ethylene-related mutants. *Plant Cell*, 2(6):513–523. 45

[66] Hall, A. E. and Bleecker, A. B. (2003). Analysis of combinatorial loss-of-function mutants in the Arabidopsis ethylene receptors reveals that the ers1 etr1 double mutant has severe developmental defects that are EIN2 dependent. *Plant Cell*, 15(9):2032–2041. 47

[67] Hanahan, D. and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, 100(1):57–70. 9, 12

[68] Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674. 9, 12, 13, 14, 17, 21, 43

[69] Hijmans, R. J., van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, O. P., Bevan, A., Racine, E. B., Shortridge, A., et al. (2016). Package 'raster'. *R package. https://cran. r-project. org/web/packages/raster/index. html (accessed 1 October 2016)*. 24

[70] Hoerter, J. A. H., Brzostek, J., Artyomov, M. N., Abel, S. M., Casas, J., Rybakin, V., Ampudia, J., Lotz, C., Connolly, J. M., Chakraborty, A. K., Gould, K. G., and Gascoigne, N. R. J. (2013). Coreceptor affinity for MHC defines peptide specificity requirements for TCR interaction with coagonist peptide-MHC. *J. Exp. Med.*, 210(9):1807–21. 80, 81, 83, 103, 105

[71] Hsu, P. P. and Sabatini, D. M. (2008). Cancer cell metabolism: Warburg and beyond. *Cell*, 134(5):703–707. 13, 21

[72] Ibrahim, B., Diekmann, S., Schmitt, E., and Dittrich, P. (2008). In-silico modeling of the mitotic spindle assembly checkpoint. *PLOS ONE*, 3(2):e1555. 78

[73] Jackson, M. (1983). Regulation of root growth and morphology by ethylene and other externally applied growth substances. *Br. Plant Gr.* 46

[74] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer. 3, 4

[75] Jiao, Y. and Torquato, S. (2011). Emergent behaviors from a cellular automaton model for invasive tumor growth in heterogeneous microenvironments. *PLOS Comput. Biol.*, 7(12):e1002314. 15

[76] Ju, C., Yoon, G. M., Shemansky, J. M., Lin, D. Y., Ying, Z. I., Chang, J., Garrett, W. M., Kessenbrock, M., Groth, G., and Tucker, M. L. (2012). CTR1 phosphorylates the central regulator EIN2 to control ethylene hormone signaling from the ER membrane to the nucleus in Arabidopsis. *Proc. Natl. Acad. Sci. USA*, 109(47):19486–19491. 46, 49, 74

[77] Kansal, A. R., Torquato, S., Harsh, G., Chiocca, E., and Deisboeck, T. (2000). Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. *J. Theor. Biol.*, 203(4):367–382. 15

[78] Karar, J. and Maity, A. (2011). PI3K/AKT/mTOR pathway in angiogenesis. *Front. Mol. Neurosci.*, 4:51. 13

[79] Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A., and Ecker, J. R. (1993). CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the raf family of protein kinases. *Cell*, 72(3):427–441. 45, 47

[80] Kikuchi, S., Tominaga, D., Arita, M., Takahashi, K., and Tomita, M. (2003). Dynamic modeling of genetic networks using genetic algorithm and s-system. *Bioinformatics*, 19(5):643–650. 50, 78

[81] Kim, H., Helmbrecht, E. E., Stalans, M. B., Schmitt, C., Patel, N., Wen, C.-K., Wang, W., and Binder, B. M. (2011). Ethylene receptor ETHYLENE RECEPTOR1 domain requirements for ethylene responses in Arabidopsis seedlings. *Plant Physiol.*, 156(1):417–429. 46

[82] Kim, J., Wilson, R. L., Case, J. B., and Binder, B. M. (2012). A comparative study of ethylene growth response kinetics in eudicots and monocots reveals a role for gibberellin in growth inhibition and recovery. *Plant Physiol.*, 160(3):1567–1580. 46, 47, 58, 73

[83] Kitano, H. (2002). Computational systems biology. *Nature*, 420(6912):206. 1

[84] Laan, P. (1934). *Der Einfluss von Aethylen auf die Wuchsstoffbildung bei Avena und Vicia*. PhD thesis, Utrecht: Van der Laan. 46

[85] Lalanne, J.-B. and François, P. (2013). Principles of adaptive sorting revealed by *In Silico* evolution. *Phys. Rev. Lett.*, 110:218102. 78

[86] Larsen, P. B. and Chang, C. (2001). The Arabidopsis eer1 mutant has enhanced ethylene responses in the hypocotyl and stem. *Plant Physiol.*, 125(2):1061–1073. 47

[87] Li, W., Ma, M., Feng, Y., Li, H., Wang, Y., Ma, Y., Li, M., An, F., and Guo, H. (2015). EIN2-directed translational regulation of ethylene signaling in Arabidopsis. *Cell*, 163(3):670–683. 74

[88] Liberti, M. V. and Locasale, J. W. (2016). The Warburg effect: how does it benefit cancer cells? *Trends Biochem. Sci.*, 41(3):211–218. 9, 12, 13, 14, 19, 22

[89] Lim, W. A. (2010). Designing customized cell signalling circuits. *Nat. Rev. Mol. Cell Biol.*, 11(6):393–403. 77, 105

[90] Liu, D., Moberg, E., Kollind, M., Lins, P.-E., Adamson, U., and Macdonald, I. (1992). Arterial, arterialized venous, venous and capillary blood glucose measurements in normal man during hyperinsulinaemic euglycaemia and hypoglycaemia. *Diabetologia*, 35(3):287–290. 17

[91] Malaney, S. and Daly, R. J. (2001). The Ras signaling pathway in mammary tumorigenesis and metastasis. *J. Mammary Gland Biol.*, 6(1):101–113. 14

[92] Markowetz, F. (2017). All biology is computational biology. *PLOS Biol.*, 15(3):e2002050. 1

[93] Mattoo, A. K. and Suttle, J. C. (1991). *The plant hormone ethylene.* CRC press. 45

[94] McDaniel, B. K. and Binder, B. M. (2012). Ethylene receptor 1 (ETR1) is sufficient and has the predominant role in mediating inhibition of ethylene responses by silver in Arabidopsis thaliana. *J. Biochem.*, 287(31):26094–26103. 46

[95] McGoron, A. J., Nair, P., and Schubert, R. W. (1997). Michaelis-menten kinetics model of oxygen consumption by rat brain slices following hypoxia. *Ann. Biomed. Eng.*, 25(3):565–572. 17, 18

[96] Merchante, C., Brumos, J., Yun, J., Hu, Q., Spencer, K. R., Enríquez, P., Binder, B. M., Heber, S., Stepanova, A. N., and Alonso, J. M. (2015). Gene-specific translation regulation mediated by the hormone-signaling molecule EIN2. *Cell*, 163(3):684–697. 46, 74

[97] Morsut, L., Roybal, K. T., Xiong, X., Gordley, R. M., Coyle, S. M., Thomson, M., and Lim, W. A. (2016). Engineering customized cell sensing and response behaviors using synthetic notch receptors. *Cell*, 164(4):780 – 791. 77

[98] Nugent, L. J. and Jain, R. K. (1984). Extravascular diffusion in normal and neoplastic tissues. *Cancer Res.*, 44(1):238–244. 15, 17

[99] O'Shaughnessy, E. C. and Sarkar, C. A. (2012). Analyzing and engineering cell signaling modules with synthetic biology. *Curr. Opin. in Biotechnology*, 23(5):785–790. 77, 105

[100] Paladugu, S., Chickarmane, V., Deckard, A., Frumkin, J., McCormack, M., and Sauro, H. (2006). In silico evolution of functional modules in biochemical networks. *IEE Proceedings-Systems Biology*, 153(4):223–235. 78

[101] Patil, K. R., Rocha, I., Förster, J., and Nielsen, J. (2005). Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinform.*, 6(1):1. 50, 78

[102] Potuschak, T., Lechner, E., Parmentier, Y., Yanagisawa, S., Grava, S., Koncz, C., and Genschik, P. (2003). EIN3-dependent regulation of plant ethylene hormone signaling by two Arabidopsis F box proteins: EBF1 and EBF2. *Cell*, 115(6):679–689. 46

[103] Potuschak, T., Vansiri, A., Binder, B. M., Lechner, E., Vierstra, R. D., and Genschik, P. (2006). The exoribonuclease XRN4 is a component of the ethylene response pathway in Arabidopsis. *Plant Cell*, 18(11):3047–3057. 46

[104] Prescott, A. M., McCollough, F. W., Eldreth, B. L., Binder, B. M., and Abel, S. M. (2016). Analysis of network topologies underlying ethylene growth response kinetics. *Front. Plant Sci.*, 7. 78

[105] Qiao, H., Chang, K. N., Yazaki, J., and Ecker, J. R. (2009). Interplay between ethylene, ETP1/ETP2 F-box proteins, and degradation of EIN2 triggers ethylene responses in Arabidopsis. *Gene Dev.*, 23(4):512–521. 46, 49

[106] Qiao, H., Shen, Z., Huang, S.-s. C., Schmitz, R. J., Urich, M. A., Briggs, S. P., and Ecker, J. R. (2012). Processing and subcellular trafficking of ER-tethered EIN2 control response to ethylene gas. *Science*, 338(6105):390–393. 46, 49, 74

[107] Qiu, L., Xie, F., Yu, J., and Wen, C.-K. (2012). Arabidopsis RTE1 is essential to ethylene receptor ETR1 amino-terminal signaling independent of CTR1. *Plant Physiol.*, 159(3):1263–1276. 47

[108] Rai, M. I., Wang, X., Thibault, D. M., Kim, H. J., Bombyk, M. M., Binder, B. M., Shakeel, S. N., and Schaller, G. E. (2015). The ARGOS gene family functions in a negative feedback loop to desensitize plants to ethylene. *BMC Plant Biol.*, 15(1):1–14. 46, 47

[109] Rauser, W. E. and Horton, R. F. (1975). Rapid effects of indoleacetic acid and ethylene on the growth of intact pea roots. *Plant Physiol.*, 55(3):443–447. 46

[110] Roman, G., Lubarsky, B., Kieber, J. J., Rothenberg, M., and Ecker, J. R. (1995). Genetic analysis of ethylene signal transduction in Arabidopsis thaliana: five novel mutant loci integrated into a stress response pathway. *Genetics*, 139(3):1393–1409. 47

[111] Rosette, C., Werlen, G., Daniels, M. A., Holman, P. O., Alam, S. M., Travers, P. J., Gascoigne, N. R., Palmer, E., and Jameson, S. C. (2001). The impact of duration versus extent of TCR occupancy on T cell activation: a revision of the kinetic proofreading model. *Immunity*, 15(1):59–70. 80

[112] Roybal, K. T., Rupp, L. J., Morsut, L., Walker, W. J., McNally, K. A., Park, J. S., and Lim, W. A. (2016). Precision tumor recognition by T cells with combinatorial antigen-sensing circuits. *Cell*, 164(4):770 – 779. 77

[113] Santos, J. and Monteagudo, Á. (2012). Study of cancer hallmarks relevance using a cellular automaton tumor growth model. In *International Conference on Parallel Problem Solving from Nature*, pages 489–499. Springer. 15

[114] Scott, J. G., Fletcher, A. G., Anderson, A. R., and Maini, P. K. (2016). Spatial metrics of tumour vascular organisation predict radiation efficacy in a computational model. *PLoS Comput Biol*, 12(1):e1004712. 15, 17, 22

[115] Shekhar, K., Brodin, P., Davis, M. M., and Chakraborty, A. K. (2014). Automatic classification of cellular expression by nonlinear stochastic embedding (ACCENSE). *Proc Natl Acad Sci USA*, 111(1):202–207. 78, 88, 92

[116] Silva, A. S. and Gatenby, R. A. (2010). A theoretical quantitative model for evolution of cancer chemotherapy resistance. *Biol. Direct*, 5(1):25. 15, 21

[117] Silva, A. S., Yunes, J. A., Gillies, R. J., and Gatenby, R. A. (2009). The potential role of systemic buffers in reducing intratumoral extracellular pH and acid-mediated invasion. *Cancer Res.*, 69(6):2677–2684. 15

[118] Smallbone, K., Gatenby, R. A., Gillies, R. J., Maini, P. K., and Gavaghan, D. J. (2007). Metabolic changes during carcinogenesis: potential impact on invasiveness. *J. Theor. Biol.*, 244(4):703–713. 15, 21

[119] Spirov, A. and Holloway, D. (2013). Using evolutionary computations to understand the design and evolution of gene and cell regulatory networks. *Methods*, 62(1):39–55. 50, 78

[120] Strauss, W. A. (1992). *Partial differential equations*, volume 92. Wiley New York. 16

[121] Sun, J., Garibaldi, J. M., and Hodgman, C. (2012). Parameter estimation using metaheuristics in systems biology: A comprehensive review. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 9(1):185–202. 50, 78

[122] Tlsty, T. D. and Coussens, L. M. (2006). Tumor stroma and regulation of cancer development. *Annu. Rev. Pathol. Mech. Dis.*, 1:119–150. 9, 14

[123] van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.*, 9:2579–2605. 78, 86, 87, 92

[124] van der Merwe, P. A. and Dushek, O. (2011). Mechanisms for T cell receptor triggering. *Nat. Rev. Immunol.*, 11(1):47–55. 79

[125] Van Zanten, M., Basten Snoek, L., EckStouten, V., Proveniers, M. C., Torii, K. U., Voesenek, L. A., Peeters, A. J., and Millenaar, F. F. (2010). Ethylene-induced hyponastic growth in Arabidopsis thaliana is controlled by ERECTA. *Plant J.*, 61(1):83–95. 46

[126] Vandenbussche, F., Petrášek, J., Žádníková, P., Hoyerová, K., Pešek, B., Raz, V., Swarup, R., Bennett, M., Zažímalová, E., and Benková, E. (2010). The auxin influx carriers AUX1 and LAX3 are involved in auxin-ethylene interactions during apical hook development in Arabidopsis thaliana seedlings. *Development*, 137(4):597–606. 46

[127] Vander Heiden, M. G., Cantley, L. C., and Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*, 324(5930):1029–1033. 9, 12

[128] Vara, J. Á. F., Casado, E., de Castro, J., Cejas, P., Belda-Iniesta, C., and González-Barón, M. (2004). PI3K/Akt signalling pathway and cancer. *Cancer Treat. Rev.*, 30(2):193–204. 13

[129] Vaupel, P., Kallinowski, F., and Okunieff, P. (1989). Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review. *Cancer Res.*, 49(23):6449–6465. 19

[130] Ward, P. S. and Thompson, C. B. (2012). Metabolic reprogramming: a cancer hallmark even Warburg did not anticipate. *Cancer Cell*, 21(3):297–308. 13, 14

[131] Warner, H. and Leopold, A. (1971). Timing of growth regulator responses in peas. *Biochem. Biophys. Res. Commun.*, 44(4):989–994. 46

[132] Wen, X., Zhang, C., Ji, Y., Zhao, Q., He, W., An, F., Jiang, L., and Guo, H. (2012). Activation of ethylene signaling is mediated by nuclear translocation of the cleaved EIN2 carboxyl terminus. *Cell Res.*, 22(11):1613–1616. 46, 49, 74

[133] Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York. 88

[134] Wu, C.-Y., Rupp, L. J., Roybal, K. T., and Lim, W. A. (2015). Synthetic biology approaches to engineer T cells. *Curr. Opin. Immunol.*, 35:123–130. 77, 80

[135] Yanagisawa, S., Yoo, S.-D., and Sheen, J. (2003). Differential regulation of EIN3 stability by glucose and ethylene signalling in plants. *Nature*, 425(6957):521–525. 46

[136] Žádníková, P., Petrášek, J., Marhavý, P., Raz, V., Vandenbussche, F., Ding, Z., Schwarzerová, K., Morita, M. T., Tasaka, M., and Hejátko, J. (2010). Role of PIN-mediated auxin efflux in apical hook development of Arabidopsis thaliana. *Development*, 137(4):607–617. 46

Appendices

A Summary of Signal Transduction Ordinary Differential Equations

Listed below are complete descriptions of all coupled ordinary differential equation systems used to model all signal transduction network responses.

A.1 Ethylene Signaling in Arabidopsis

Shown below are the ordinary differential equations (ODEs) describing the time evolution of each network considered in chapter 3. Square brackets denote the concentration of a network component. We use the following notation for network parameters: k indicates a reaction rate, K indicates an activation coefficient appearing in a Hill equation, and N indicates a Hill coefficient. The subscripts "prod' and "degr" indicate whether a rate constant is associated with production or degradation; an associated superscript refers to the relevant network component. Remaining terms appear within large parentheses like the following:

$$-\left(k_{\text{degr}}\frac{[\text{CTR1}]^N}{K_{\text{degr}}^N+[\text{CTR1}]^N}\right)_{\text{CTR1}}[\text{EIN2}]$$

This term is associated with the degradation of EIN2 in the CFF/NFB network. Each of the three parameters within the parentheses has two implied associations: Each represents the effect of CTR1 (subscript of parentheses) on the degradation of EIN2 (term being modified by the factor in the parentheses). In other words, the subscript outside of the parentheses indicates the network component that regulates the interaction.

Complete Coherent Feedforward/Negative Feedback (CFF/NFB) Network (See Fig. 3.3)

$$
\begin{aligned}
\frac{d[\mathrm{R}]}{dt} &= k^{\mathrm{R}}_{\mathrm{prod}}(1-[\mathrm{R}]) - k^{\mathrm{E}}_{\mathrm{degr}}[\mathrm{E}][\mathrm{R}] \\
\frac{d[\mathrm{CTR1}]}{dt} &= \left(k_{\mathrm{prod}}\frac{[\mathrm{R}]^N}{K^N_{\mathrm{prod}}+[\mathrm{R}]^N}\right)_{\mathrm{R}}(1-[\mathrm{CTR1}]) - k^{\mathrm{CTR1}}_{\mathrm{degr}}[\mathrm{CTR1}] \\
\frac{d[\mathrm{EIN2}]}{dt} &= k^{\mathrm{EIN2}}_{\mathrm{prod}}(1-[\mathrm{EIN2}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{CTR1}]^N}{K^N_{\mathrm{degr}}+[\mathrm{CTR1}]^N}\right)_{\mathrm{CTR1}}[\mathrm{EIN2}] \\
\frac{d[\mathrm{EBF}]}{dt} &= k^{\mathrm{EBF}}_{\mathrm{prod}}(1-[\mathrm{EBF}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{EIN2}]^N}{K^N_{\mathrm{degr}}+[\mathrm{EIN2}]^N}\right)_{\mathrm{EIN2}}[\mathrm{EBF}] \\
\frac{d[\mathrm{EIN3}]}{dt} &= k^{\mathrm{EIN3}}_{\mathrm{prod}}(1-[\mathrm{EIN3}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{EBF}]^N}{K^N_{\mathrm{degr}}+[\mathrm{EBF}]^N}\right)_{\mathrm{EBF}}[\mathrm{EIN3}] \\
\frac{d[\mathrm{GA}]}{dt} &= k^{\mathrm{GA}}_{\mathrm{prod}}(1-[\mathrm{GA}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{EIN3}]^N}{K^N_{\mathrm{degr}}+[\mathrm{EIN3}]^N}\right)_{\mathrm{EIN3}}[\mathrm{GA}] \\
&\quad - \left(k_{\mathrm{degr}}\frac{[\mathrm{Growth}]^N}{K^N_{\mathrm{degr}}+[\mathrm{Growth}]^N}\right)_{\mathrm{Growth}}[\mathrm{GA}] \\
\frac{d[\mathrm{Growth}]}{dt} &= k^{\mathrm{Growth}}_{\mathrm{prod}}(1-[\mathrm{Growth}]) + \left(k_{\mathrm{prod}}\frac{[\mathrm{GA}]^N}{K^N_{\mathrm{prod}}+[\mathrm{GA}]^N}\right)_{\mathrm{GA}}(1-[\mathrm{Growth}]) \\
&\quad - \left(k_{\mathrm{degr}}\frac{[\mathrm{EIN2}]^N}{K^N_{\mathrm{degr}}+[\mathrm{EIN2}]^N}\right)_{\mathrm{EIN2}}[\mathrm{Growth}] \\
&\quad - \left(k_{\mathrm{degr}}\frac{[\mathrm{EIN3}]^N}{K^N_{\mathrm{degr}}+[\mathrm{EIN3}]^N}\right)_{\mathrm{EIN3}}[\mathrm{Growth}]
\end{aligned}
$$

Simplified Coherent Feedforward/Negative Feedback (CFF/NFB) Network A (See Fig. 3.15A)

$$\begin{aligned}
\frac{d[\mathrm{Y}]}{dt} &= \left(k_{\mathrm{prod}}\frac{[\mathrm{E}]^N}{K_{\mathrm{prod}}^N+[\mathrm{E}]^N}\right)_{\mathrm{E}}(1-[\mathrm{Y}]) - k_{\mathrm{degr}}^{\mathrm{Y}}[\mathrm{Y}] \\
\frac{d[\mathrm{Growth}]}{dt} &= k_{\mathrm{prod}}^{\mathrm{Growth}}(1-[\mathrm{Growth}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{E}]^N}{K_{\mathrm{degr}}^N+[\mathrm{E}]^N}\right)_{\mathrm{E}}[\mathrm{Growth}] \\
&\quad - \left(k_{\mathrm{degr}}\frac{[\mathrm{Y}]^N}{K_{\mathrm{degr}}^N+[\mathrm{Y}]^n}\right)_{\mathrm{Y}}[\mathrm{Growth}] \\
&\quad - \left(k_{\mathrm{degr}}\frac{[\mathrm{Growth}]^N}{K_{\mathrm{degr}}^N+[\mathrm{Growth}]^N}\right)_{\mathrm{Growth}}[\mathrm{Growth}]
\end{aligned}$$

Simplified Coherent Feedforward/Negative Feedback (CFF/NFB) Network B (See Fig. 3.15B)

$$\begin{aligned}
\frac{d[\mathrm{Y}]}{dt} &= \left(k_{\mathrm{prod}}\frac{[\mathrm{E}]^N}{K_{\mathrm{prod}}^N+[\mathrm{E}]^N}\right)_{\mathrm{E}}(1-[\mathrm{Y}]) - k_{\mathrm{degr}}^{\mathrm{Y}}[\mathrm{Y}] \\
\frac{d[\mathrm{Z}]}{dt} &= k_{\mathrm{prod}}^{\mathrm{Z}}(1-[\mathrm{Z}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{Growth}]^N}{K_{\mathrm{degr}}^N+[\mathrm{Growth}]^N}\right)_{\mathrm{Growth}}[\mathrm{Z}] \\
\frac{d[\mathrm{Growth}]}{dt} &= k_{\mathrm{prod}}^{\mathrm{Growth}}(1-[\mathrm{Growth}]) + \left(k_{\mathrm{prod}}\frac{[\mathrm{Z}]^N}{K_{\mathrm{prod}}^N+[\mathrm{Z}]^N}\right)_{\mathrm{Z}}(1-[\mathrm{Growth}]) \\
&\quad - \left(k_{\mathrm{degr}}\frac{[\mathrm{E}]^N}{K_{\mathrm{degr}}^N+[\mathrm{E}]^N}\right)_{\mathrm{E}}[\mathrm{Growth}] - \left(k_{\mathrm{degr}}\frac{[\mathrm{Y}]^N}{K_{\mathrm{degr}}^N+[\mathrm{Y}]^n}\right)_{\mathrm{Y}}[\mathrm{Growth}]
\end{aligned}$$

Simplified Coherent Feedforward/Negative Feedback (CFF/NFB) Network C (See Fig. 3.15C)

$$\begin{aligned}
\frac{d[\mathrm{Y}]}{dt} &= \left(k_{\mathrm{prod}}\frac{[\mathrm{E}]^N}{K_{\mathrm{prod}}^N+[\mathrm{E}]^N}\right)_{\mathrm{E}}(1-[\mathrm{Y}]) - k_{\mathrm{degr}}^{\mathrm{Y}}[\mathrm{Y}] \\
\frac{d[\mathrm{Z}]}{dt} &= k_{\mathrm{prod}}^{\mathrm{Z}}(1-[\mathrm{Z}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{Y}]^N}{K_{\mathrm{degr}}^N+[\mathrm{Y}]^N}\right)_{\mathrm{Y}}[\mathrm{Z}] \\
&\quad - \left(k_{\mathrm{degr}}\frac{[\mathrm{Growth}]^N}{K_{\mathrm{degr}}^N+[\mathrm{Growth}]^N}\right)_{\mathrm{Growth}}[\mathrm{Z}] \\
\frac{d[\mathrm{Growth}]}{dt} &= k_{\mathrm{prod}}^{\mathrm{Growth}}(1-[\mathrm{Growth}]) + \left(k_{\mathrm{prod}}\frac{[\mathrm{Z}]^N}{K_{\mathrm{prod}}^N+[\mathrm{Z}]^N}\right)_{\mathrm{Z}}(1-[\mathrm{Growth}]) \\
&\quad - \left(k_{\mathrm{degr}}\frac{[\mathrm{E}]^N}{K_{\mathrm{degr}}^N+[\mathrm{E}]^N}\right)_{\mathrm{E}}[\mathrm{Growth}] - \left(k_{\mathrm{degr}}\frac{[\mathrm{Y}]^N}{K_{\mathrm{degr}}^N+[\mathrm{Y}]^N}\right)_{\mathrm{Y}}[\mathrm{Growth}]
\end{aligned}$$

Simplified Coherent Feedforward/Negative Feedback (CFF/NFB) Network D (See Fig. 3.15D)

$$\begin{aligned}
\frac{d[\mathrm{X}]}{dt} &= \left(k_{\mathrm{prod}}\frac{[\mathrm{E}]^N}{K_{\mathrm{prod}}^N+[\mathrm{E}]^N}\right)_{\mathrm{E}}(1-[\mathrm{X}]) - k_{\mathrm{degr}}^{\mathrm{X}}[\mathrm{X}] \\
\frac{d[\mathrm{Y}]}{dt} &= \left(k_{\mathrm{prod}}\frac{[\mathrm{X}]^N}{K_{\mathrm{prod}}^N+[\mathrm{X}]^N}\right)_{\mathrm{X}}(1-[\mathrm{Y}]) - k_{\mathrm{degr}}^{\mathrm{Y}}[\mathrm{Y}] \\
\frac{d[\mathrm{Z}]}{dt} &= k_{\mathrm{prod}}^{\mathrm{Z}}(1-[\mathrm{Z}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{Growth}]^N}{K_{\mathrm{degr}}^N+[\mathrm{Growth}]^N}\right)_{\mathrm{Growth}}[\mathrm{Z}] \\
\frac{d[\mathrm{Growth}]}{dt} &= k_{\mathrm{prod}}^{\mathrm{Growth}}(1-[\mathrm{Growth}]) + \left(k_{\mathrm{prod}}\frac{[\mathrm{Z}]^N}{K_{\mathrm{prod}}^N+[\mathrm{Z}]^N}\right)_{\mathrm{Z}}(1-[\mathrm{Growth}]) \\
&\quad - \left(k_{\mathrm{degr}}\frac{[\mathrm{X}]^N}{K_{\mathrm{degr}}^N+[\mathrm{X}]^N}\right)_{\mathrm{X}}[\mathrm{Growth}] - \left(sk_{\mathrm{degr}}\frac{[\mathrm{Y}]^N}{K_{\mathrm{degr}}^N+[\mathrm{Y}]^N}\right)_{\mathrm{Y}}[\mathrm{Growth}]
\end{aligned}$$

Simplified Coherent Feedforward/Negative Feedback (CFF/NFB) Network E (See Fig. 3.15E)

$$\frac{d[\mathrm{X}]}{dt} = \left(k_{\mathrm{prod}} \frac{[\mathrm{E}]^N}{K_{\mathrm{prod}}^N + [\mathrm{E}]^N} \right)_{\mathrm{E}} (1 - [\mathrm{X}]) - k_{\mathrm{degr}}^{\mathrm{X}} [\mathrm{X}]$$

$$\frac{d[\mathrm{Y}]}{dt} = \left(k_{\mathrm{prod}} \frac{[\mathrm{X}]^N}{K_{\mathrm{prod}}^N + [\mathrm{X}]^N} \right)_{\mathrm{X}} (1 - [\mathrm{Y}]) - k_{\mathrm{degr}}^{\mathrm{Y}} [\mathrm{Y}]$$

$$\begin{aligned} \frac{d[\mathrm{Z}]}{dt} = {} & k_{\mathrm{prod}}^{\mathrm{Z}} (1 - [\mathrm{Z}]) - \left(k_{\mathrm{degr}} \frac{[\mathrm{Y}]^N}{K_{\mathrm{degr}}^N + [\mathrm{Y}]^N} \right)_{\mathrm{Y}} [\mathrm{Z}] \\ & - \left(k_{\mathrm{degr}} \frac{[\mathrm{Growth}]^N}{K_{\mathrm{degr}}^N + [\mathrm{Growth}]^N} \right)_{\mathrm{Growth}} [\mathrm{Z}] \end{aligned}$$

$$\begin{aligned} \frac{d[\mathrm{Growth}]}{dt} = {} & k_{\mathrm{prod}}^{\mathrm{Growth}} (1 - [\mathrm{Growth}]) + \left(k_{\mathrm{prod}} \frac{[\mathrm{Z}]^N}{K_{\mathrm{prod}}^N + [\mathrm{Z}]^N} \right)_{\mathrm{Z}} (1 - [\mathrm{Growth}]) \\ & - \left(k_{\mathrm{degr}} \frac{[\mathrm{X}]^N}{K_{\mathrm{degr}}^N + [\mathrm{X}]^N} \right)_{\mathrm{X}} [\mathrm{Growth}] - \left(k_{\mathrm{degr}} \frac{[\mathrm{Y}]^N}{K_{\mathrm{degr}}^N + [\mathrm{Y}]^N} \right)_{\mathrm{Y}} [\mathrm{Growth}] \end{aligned}$$

Complete Positive Feedback (PFB) Network (See Fig. 3.3)

The parameters describing the CTR1-regulated conversion of EIN2 to EIN2-C appears as part of a production term in the EIN2 ODE and as part of a degradation term in EIN2-C. To clearly demarcate these parameters from the remaining parameters that are strictly classified as production or degradation, these parameters been given the subscript "cat".

$$
\begin{aligned}
\frac{d[\mathrm{R}]}{dt} &= k^{\mathrm{R}}_{\mathrm{prod}}(1-[\mathrm{R}]) - k^{\mathrm{E}}_{\mathrm{degr}}[\mathrm{E}][\mathrm{R}] \\
\frac{d[\mathrm{CTR1}]}{dt} &= \left(k_{\mathrm{prod}}\frac{[\mathrm{R}]^N}{K^N_{\mathrm{prod}}+[\mathrm{R}]^N}\right)_{\mathrm{R}}(1-[\mathrm{CTR1}]) - k^{\mathrm{CTR1}}_{\mathrm{degr}}[\mathrm{CTR1}] \\
\frac{d[\mathrm{EIN2}]}{dt} &= k^{\mathrm{EIN2}}_{\mathrm{prod}}(1-[\mathrm{EIN2}]) + \left(k_{\mathrm{prod}}\frac{[\mathrm{EIN3}]^N}{K^N_{\mathrm{prod}}+[\mathrm{EIN3}]^N}\right)_{\mathrm{EIN3}}(1-[\mathrm{EIN2}]) \\
&\quad - \left(k_{\mathrm{cat}}\frac{K^N_{\mathrm{cat}}}{K^N_{\mathrm{cat}}+[\mathrm{CTR1}]^N}\right)_{\mathrm{CTR1}}[\mathrm{EIN2}](1-[\mathrm{EIN2\text{-}C}]) \\
\frac{d[\mathrm{EIN2\text{-}C}]}{dt} &= \left(k_{\mathrm{cat}}\frac{K^N_{\mathrm{cat}}}{K^N_{\mathrm{cat}}+[\mathrm{CTR1}]^N}\right)_{\mathrm{CTR1}}[\mathrm{EIN2}](1-[\mathrm{EIN2\text{-}C}]) - k^{\mathrm{EIN2\text{-}C}}_{\mathrm{degr}}[\mathrm{EIN2\text{-}C}] \\
\frac{d[\mathrm{EBF}]}{dt} &= k^{\mathrm{EBF}}_{\mathrm{prod}}(1-[\mathrm{EBF}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{EIN2\text{-}C}]^N}{K^N_{\mathrm{degr}}+[\mathrm{EIN2\text{-}C}]^N}\right)_{\mathrm{EIN2\text{-}C}}[\mathrm{EBF}] \\
\frac{d[\mathrm{EIN3}]}{dt} &= k^{\mathrm{EIN3}}_{\mathrm{prod}}(1-[\mathrm{EIN3}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{EBF}]^N}{K^N_{\mathrm{degr}}+[\mathrm{EBF}]^N}\right)_{\mathrm{EBF}}[\mathrm{EIN3}] \\
\frac{d[\mathrm{Growth}]}{dt} &= k^{\mathrm{Growth}}_{\mathrm{prod}}(1-[\mathrm{Growth}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{EIN2\text{-}C}]^N}{K^N_{\mathrm{degr}}+[\mathrm{EIN2\text{-}C}]^N}\right)_{\mathrm{EIN2C}}(1-[\mathrm{Growth}])
\end{aligned}
$$

Simplified Positive Feedback (PFB) Network (See Fig. 3.18)

The parameters describing the E-regulated conversion of X to Y appears as part of a production term in the ODE for X and as part of a degradation term in the ODE for Y. To clearly demarcate these parameters from the remaining parameters that are strictly classified as production or degradation, these parameters been given the subscript "cat".

$$\begin{aligned}
\frac{d[\mathrm{X}]}{dt} &= k^{\mathrm{X}}_{\mathrm{prod}}(1-[\mathrm{X}]) + \left(k_{\mathrm{prod}}\frac{[\mathrm{Y}]^N}{K^N_{\mathrm{prod}}+[\mathrm{Y}]^N}\right)_{\mathrm{Y}}\cdot \\
&\qquad (1-[\mathrm{X}]) - \left(k_{\mathrm{cat}}\frac{[\mathrm{E}]^N}{K^N_{\mathrm{cat}}+[\mathrm{E}]^N}\right)_{\mathrm{E}}[\mathrm{X}](1-[\mathrm{Y}]) \\
\frac{d[\mathrm{Y}]}{dt} &= \left(k_{\mathrm{cat}}\frac{[\mathrm{E}]^N}{K^N_{\mathrm{cat}}+[\mathrm{E}]^N}\right)_{\mathrm{E}}[\mathrm{X}](1-[\mathrm{Y}]) - k^{\mathrm{Y}}_{\mathrm{degr}}[\mathrm{Y}] \\
\frac{d[\mathrm{Growth}]}{dt} &= k^{\mathrm{Growth}}_{\mathrm{prod}}(1-[\mathrm{Growth}]) - \left(k_{\mathrm{degr}}\frac{[\mathrm{Y}]^N}{K^N_{\mathrm{degr}}+[\mathrm{Y}]^N}\right)_{\mathrm{Y}}[\mathrm{Growth}]
\end{aligned}$$

A.2 Early T Cell Receptor Network

The following ordinary differential equations (ODEs) fully describe the dynamics of the early T cell response signaling network considered chapter 4 (see Fig. 4.2). As discussed in Methods, a complete description of the network requires 24 coupled ODEs. For ternary complexes, the bonds present are indicated by the order in which the compounds are written. For example, TCR·pMHC·CD indicates that the TCR and CD are directly bound to pMHC but not to each other. "Complex" refers to the fully bound ternary complex in which all binding partners are bound. The state of the TCR in the fully bound complex is indicated by Complex, Complex·p, and Complex·pp (corresponding to TCR, TCRp, and TCRpp, respectively).

$$
\begin{aligned}
\frac{d[\mathrm{CD}]}{dt} &= k_{\mathrm{off}}^{\mathrm{pMHC\cdot CD}}\Big([\mathrm{pMHC\cdot CD}] + [\mathrm{CD\cdot pMHC\cdot TCR}] + [\mathrm{CD\cdot pMHC\cdot TCRp}] \\
&\qquad + [\mathrm{CD\cdot pMHC\cdot TCRpp}]\Big) \\
&\quad + k_{\mathrm{off}}^{\mathrm{TCR\cdot CD}}\Big([\mathrm{TCR\cdot CD}] + [\mathrm{TCRp\cdot CD}] + [\mathrm{TCRpp\cdot CD}] \\
&\qquad + [\mathrm{CD\cdot TCR\cdot pMHC}] + [\mathrm{CD\cdot TCRp\cdot pMHC}] + [\mathrm{CD\cdot TCRpp\cdot pMHC}]\Big) \\
&\quad - [\mathrm{CD}]\Bigg(k_{\mathrm{on}}^{\mathrm{TCR\cdot CD}}\Big([\mathrm{TCR}] + [\mathrm{TCRp}] + [\mathrm{TCRpp}] \\
&\qquad\quad + [\mathrm{TCR\cdot pMHC}] + [\mathrm{TCRp\cdot pMHC}] + [\mathrm{TCRpp\cdot pMHC}]\Big) \\
&\qquad + k_{\mathrm{on}}^{\mathrm{pMHC\cdot CD}}\Big([\mathrm{pMHC}] + [\mathrm{TCR\cdot pMHC}] + [\mathrm{TCRp\cdot pMHC}] + [\mathrm{TCRpp\cdot pMHC}]\Big)\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
\frac{d[\text{pMHC}]}{dt} &= k_{\text{off}}^{\text{pMHC·CD}}\Big([\text{pMHC·CD}] + [\text{TCR·CD·pMHC}] + [\text{TCRp·CD·pMHC}] \\
&\quad + [\text{TCRpp·CD·pMHC}]\Big) \\
&+ k_{\text{off}}^{\text{TCR·pMHC}}\Big([\text{TCR·pMHC}] + [\text{TCRp·pMHC}] + [\text{TCRpp·pMHC}] \\
&\quad + [\text{CD·TCR·pMHC}] + [\text{CD·TCRp·pMHC}] + [\text{CD·TCRpp·pMHC}]\Big) \\
&- [\text{pMHC}]\Big(k_{\text{on}}^{\text{pMHC·CD}}\Big([\text{CD}] + [\text{TCR·CD}] + [\text{TCRp·CD}] + [\text{TCRpp·CD}]\Big) \\
&\quad + k_{\text{on}}^{\text{TCR·pMHC}}\Big([\text{TCR}] + [\text{TCRp}] + [\text{TCRpp}] + [\text{TCR·CD}] + [\text{TCRp·CD}] \\
&\quad + [\text{TCRpp·CD}]\Big)\Big) \\
\frac{d[\text{pMHC·CD}]}{dt} &= k_{\text{on}}^{\text{pMHC·CD}}[\text{CD}][\text{pMHC}] \\
&+ k_{\text{off}}^{\text{TCR·pMHC}}\Big([\text{CD·pMHC·TCR}] + [\text{CD·pMHC·TCRp}] + [\text{CD·pMHC·TCRpp}]\Big) \\
&+ k_{\text{off}}^{\text{TCR·CD}}\Big([\text{TCR·CD·pMHC}] + [\text{TCRp·CD·pMHC}] + [\text{TCRpp·CD·pMHC}]\Big) \\
&- [\text{pMHC·CD}]\cdot \\
&\Big(k_{\text{off}}^{\text{pMHC·CD}} + \Big(k_{\text{on}}^{\text{TCR·pMHC}} + k_{\text{on}}^{\text{TCR·CD}}\Big)\Big([\text{TCR}] + [\text{TCRp}] + [\text{TCRpp}]\Big)\Big) \\
\frac{d[\text{TCR}]}{dt} &= k_{\text{phosph}}[\text{TCRp}] + k_{\text{off}}^{\text{TCR·pMHC}}\Big([\text{TCR·pMHC}] + [\text{CD·pMHC·TCR}]\Big) \\
&+ k_{\text{off}}^{\text{TCR·CD}}\Big([\text{TCR·CD}] + \text{TCR·CD·pMHC}]\Big) \\
&- [\text{TCR}]\cdot \\
&\Big(k_{\text{on}}^{\text{TCR·pMHC}}\Big([\text{pMHC}] + [\text{pMHC·CD}]\Big) + k_{\text{on}}^{\text{TCR·CD}}\Big([\text{CD}] + [\text{pMHC·CD}]\Big)\Big) \\
\frac{d[\text{TCRp}]}{dt} &= 2k_{\text{phosph}}[\text{TCRpp}] + k_{\text{off}}^{\text{TCR·pMHC}}\Big([\text{TCRp·pMHC}] + [\text{CD·pMHC·TCRp}]\Big) \\
&+ k_{\text{off}}^{\text{TCR·CD}}\Big([\text{TCRp·CD}] + [\text{TCRp·CD·pMHC}]\Big) \\
&- [\text{TCRp}]\Big(k_{\text{phosph}} + k_{\text{on}}^{\text{TCR·pMHC}}\Big([\text{pMHC}] + [\text{pMHC·CD}]\Big) \\
&\quad + k_{\text{on}}^{\text{TCR·CD}}\Big([\text{CD}] + [\text{pMHC·CD}]\Big)\Big)
\end{aligned}
$$

$$
\begin{aligned}
\frac{d[\mathrm{TCRpp}]}{dt} &= k_{\mathrm{off}}^{\mathrm{TCR\cdot pMHC}}\Big([\mathrm{TCRpp\cdot pMHC}] + [\mathrm{CD\cdot pMHC\cdot TCRpp}]\Big) \\
&+ k_{\mathrm{off}}^{\mathrm{TCRpp\cdot CD}}\Big([\mathrm{TCRpp\cdot CD}] + [\mathrm{TCRpp\cdot CD\cdot pMHC}]\Big) \\
&- [\mathrm{TCRpp}]\Bigg(2k_{\mathrm{phosph}} + k_{\mathrm{on}}^{\mathrm{TCR\cdot pMHC}}\Big([\mathrm{pMHC}] + [\mathrm{pMHC\cdot CD}]\Big) \\
&\quad + k_{\mathrm{on}}^{\mathrm{TCR\cdot CD}}\Big([\mathrm{CD}] + [\mathrm{pMHC\cdot CD}]\Big)\Bigg) \\
\frac{d[\mathrm{TCR\cdot pMHC}]}{dt} &= k_{\mathrm{on}}^{\mathrm{TCR\cdot pMHC}}[\mathrm{pMHC}][\mathrm{TCR}] + k_{\mathrm{off}}^{\mathrm{pMHC\cdot CD}}[\mathrm{CD\cdot pMHC\cdot TCR}] \\
&+ k_{\mathrm{off}}^{\mathrm{TCR\cdot CD}}[\mathrm{CD\cdot TCR\cdot pMHC}] \\
&- [\mathrm{TCR\cdot pMHC}]\Big(k_{\mathrm{off}}^{\mathrm{TCR\cdot pMHC}} + \Big(k_{\mathrm{on}}^{\mathrm{pMHC\cdot CD}} + k_{\mathrm{on}}^{\mathrm{TCR\cdot CD}}\Big)[\mathrm{CD}]\Big) \\
\frac{d[\mathrm{TCRp\cdot pMHC}]}{dt} &= k_{\mathrm{on}}^{\mathrm{TCR\cdot pMHC}}[\mathrm{pMHC}][\mathrm{TCRp}] + k_{\mathrm{off}}^{\mathrm{pMHC\cdot CD}}[\mathrm{CD\cdot pMHC\cdot TCRp}] \\
&+ k_{\mathrm{off}}^{\mathrm{TCR\cdot CD}}[\mathrm{CD\cdot TCRp\cdot pMHC}] \\
&- [\mathrm{TCRp\cdot pMHC}]\Big(k_{\mathrm{off}}^{\mathrm{TCR\cdot pMHC}} + \Big(k_{\mathrm{on}}^{\mathrm{pMHC\cdot CD}} + k_{\mathrm{on}}^{\mathrm{TCR\cdot CD}}\Big)[\mathrm{CD}]\Big) \\
\frac{d[\mathrm{TCRpp\cdot pMHC}]}{dt} &= k_{\mathrm{on}}^{\mathrm{TCR\cdot pMHC}}[\mathrm{pMHC}][\mathrm{TCRpp}] + k_{\mathrm{off}}^{\mathrm{pMHC\cdot CD}}[\mathrm{CD\cdot pMHC\cdot TCRpp}] \\
&+ k_{\mathrm{off}}^{\mathrm{TCR\cdot CD}}[\mathrm{CD\cdot TCRpp\cdot pMHC}] \\
&- [\mathrm{TCRpp\cdot pMHC}]\Big(k_{\mathrm{off}}^{\mathrm{TCR\cdot pMHC}} + \Big(k_{\mathrm{on}}^{\mathrm{pMHC\cdot CD}} + k_{\mathrm{on}}^{\mathrm{TCR\cdot CD}}\Big)[\mathrm{CD}]\Big) \\
\frac{d[\mathrm{TCR\cdot CD}]}{dt} &= k_{\mathrm{on}}^{\mathrm{TCR\cdot CD}}[\mathrm{CD}][\mathrm{TCR}] + k_{\mathrm{off}}^{\mathrm{pMHC\cdot CD}}[\mathrm{TCR\cdot CD\cdot pMHC}] \\
&+ k_{\mathrm{off}}^{\mathrm{TCR\cdot pMHC}}[\mathrm{CD\cdot TCR\cdot pMHC}] \\
&- [\mathrm{TCR\cdot CD}]\Big(k_{\mathrm{off}}^{\mathrm{TCR\cdot CD}} + \Big(k_{\mathrm{on}}^{\mathrm{TCR\cdot pMHC}} + k_{\mathrm{on}}^{\mathrm{pMHC\cdot CD}}\Big)[\mathrm{pMHC}]\Big) \\
\frac{d[\mathrm{TCRp\cdot CD}]}{dt} &= k_{\mathrm{on}}^{\mathrm{TCR\cdot CD}}[\mathrm{CD}][\mathrm{TCRp}] + k_{\mathrm{off}}^{\mathrm{pMHC\cdot CD}}[\mathrm{TCRp\cdot CD\cdot pMHC}] \\
&+ k_{\mathrm{off}}^{\mathrm{TCR\cdot pMHC}}[\mathrm{CD\cdot TCRp\cdot pMHC}] \\
&- [\mathrm{TCRp\cdot CD}]\Big(k_{\mathrm{off}}^{\mathrm{TCR\cdot CD}} + \Big(k_{\mathrm{on}}^{\mathrm{TCR\cdot pMHC}} + k_{\mathrm{on}}^{\mathrm{pMHC\cdot CD}}\Big)[\mathrm{pMHC}]\Big)
\end{aligned}
$$

$$
\begin{aligned}
\frac{d[\text{TCRpp·CD}]}{dt} &= k_{\text{on}}^{\text{TCR·CD}}[\text{CD}][\text{TCRpp}] + k_{\text{off}}^{\text{pMHC·CD}}[\text{TCRpp·CD·pMHC}] \\
&\quad + k_{\text{off}}^{\text{TCR·pMHC}}[\text{CD·TCRpp·pMHC}] \\
&\quad - [\text{TCRpp·CD}]\left(k_{\text{off}}^{\text{TCR·CD}} + \left(k_{\text{on}}^{\text{TCR·pMHC}} + k_{\text{on}}^{\text{pMHC·CD}}\right)[\text{pMHC}]\right) \\
\frac{d[\text{CD·pMHC·TCR}]}{dt} &= k_{\text{off}}^{\text{TCR·CD}}[\text{Complex}] + k_{\text{on}}^{\text{pMHC·CD}}[\text{CD}][\text{TCR·pMHC}] \\
&\quad + k_{\text{on}}^{\text{TCR·pMHC}}[\text{TCR}][\text{pMHC·CD}] \\
&\quad - [\text{CD·pMHC·TCR}]\left(\alpha\, k_{\text{on}}^{\text{TCR·CD}} + k_{\text{off}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·pMHC}}\right) \\
\frac{d[\text{CD·pMHC·TCRp}]}{dt} &= k_{\text{off}}^{\text{TCR·CD}}[\text{Complex·p}] + k_{\text{on}}^{\text{pMHC·CD}}[\text{CD}][\text{TCRp·pMHC}] \\
&\quad + k_{\text{on}}^{\text{TCR·pMHC}}[\text{TCRp}][\text{pMHC·CD}] \\
&\quad - [\text{CD·pMHC·TCRp}]\left(\alpha\, k_{\text{on}}^{\text{TCR·CD}} + k_{\text{off}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·pMHC}}\right) \\
\frac{d[\text{CD·pMHC·TCRpp}]}{dt} &= k_{\text{off}}^{\text{TCR·CD}}[\text{Complex·pp}] + k_{\text{on}}^{\text{pMHC·CD}}[\text{CD}][\text{TCRpp·pMHC}] \\
&\quad + k_{\text{on}}^{\text{TCR·pMHC}}[\text{TCRpp}][\text{pMHC·CD}] \\
&\quad - [\text{CD·pMHC·TCRpp}]\left(\alpha\, k_{\text{on}}^{\text{TCR·CD}} + k_{\text{off}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·pMHC}}\right) \\
\frac{d[\text{TCR·CD·pMHC}]}{dt} &= k_{\text{off}}^{\text{TCR·pMHC}}[\text{Complex}] + k_{\text{on}}^{\text{pMHC·CD}}[\text{pMHC}][\text{TCR·CD}] \\
&\quad + k_{\text{on}}^{\text{TCR·CD}}[\text{TCR}][\text{pMHC·CD}] \\
&\quad - [\text{TCR·CD·pMHC}]\left(\alpha\, k_{\text{on}}^{\text{TCR·pMHC}} + k_{\text{off}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·CD}}\right) \\
\frac{d[\text{TCRp·CD·pMHC}]}{dt} &= k_{\text{off}}^{\text{TCR·pMHC}}[\text{Complex·p}] + k_{\text{on}}^{\text{pMHC·CD}}[\text{pMHC}][\text{TCRp·CD}] \\
&\quad + k_{\text{on}}^{\text{TCR·CD}}[\text{TCRp}][\text{pMHC·CD}] \\
&\quad - [\text{TCRp·CD·pMHC}]\left(\alpha\, k_{\text{on}}^{\text{TCR·pMHC}} + k_{\text{off}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·CD}}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{d[\text{TCRpp·CD·pMHC}]}{dt} &= k_{\text{off}}^{\text{TCR·pMHC}}[\text{Complex·pp}] + k_{\text{on}}^{\text{pMHC·CD}}[\text{pMHC}][\text{TCRpp·CD}] \\
&\quad + k_{\text{on}}^{\text{TCR·CD}}[\text{TCRpp}][\text{pMHC·CD}] \\
&\quad - [\text{TCRpp·CD·pMHC}]\left(\alpha\, k_{\text{on}}^{\text{TCR·pMHC}} + k_{\text{off}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·CD}}\right) \\
\frac{d[\text{CD·TCR·pMHC}]}{dt} &= k_{\text{off}}^{\text{pMHC·CD}}[\text{Complex}] + k_{\text{on}}^{\text{TCR·pMHC}}[\text{pMHC}][\text{TCR·CD}] \\
&\quad + k_{\text{on}}^{\text{TCR·CD}}[\text{CD}][\text{TCR·pMHC}] \\
&\quad - [\text{CD·TCR·pMHC}]\left(k_{\text{kinase}} + \alpha\, k_{\text{on}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·pMHC}} + k_{\text{off}}^{\text{TCR·CD}}\right) \\
\frac{d[\text{CD·TCRp·pMHC}]}{dt} &= k_{\text{kinase}}[\text{CD·TCR·pMHC}] + k_{\text{on}}^{\text{TCR·pMHC}}[\text{pMHC}][\text{TCRp·CD}] \\
&\quad + k_{\text{on}}^{\text{TCR·CD}}[\text{CD}][\text{TCRp·pMHC}] + k_{\text{off}}^{\text{pMHC·CD}}[\text{Complex·p}] \\
&\quad - [\text{CD·TCRp·pMHC}]\left(k_{\text{kinase}} + \alpha\, k_{\text{on}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·pMHC}} + k_{\text{off}}^{\text{TCR·CD}}\right) \\
\frac{d[\text{CD·TCRpp·pMHC}]}{dt} &= k_{\text{kinase}}[\text{CD·TCRp·pMHC}] + k_{\text{on}}^{\text{TCR·pMHC}}[\text{pMHC}][\text{TCRpp·CD}] \\
&\quad + k_{\text{on}}^{\text{TCR·CD}}[\text{CD}][\text{TCRpp·pMHC}] + k_{\text{off}}^{\text{pMHC·CD}}[\text{Complex·pp}] \\
&\quad - [\text{CD·TCRpp·pMHC}]\left(\alpha\, k_{\text{on}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·pMHC}} + k_{\text{off}}^{\text{TCR·CD}}\right) \\
\frac{d[\text{Complex}]}{dt} &= \alpha\, k_{\text{on}}^{\text{pMHC·CD}}[\text{CD·TCR·pMHC}] + \alpha\, k_{\text{on}}^{\text{TCR·pMHC}}[\text{TCR·CD·pMHC}] \\
&\quad + \alpha\, k_{\text{on}}^{\text{TCR·CD}}[\text{CD·pMHC·TCR}] \\
&\quad - [\text{Complex}]\left(k_{\text{kinase}} + k_{\text{off}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·pMHC}} + k_{\text{off}}^{\text{TCR·CD}}\right) \\
\frac{d[\text{Complex·p}]}{dt} &= k_{\text{kinase}}[\text{Complex}] + \alpha\, k_{\text{on}}^{\text{pMHC·CD}}[\text{CD·TCRp·pMHC}] \\
&\quad + \alpha\, k_{\text{on}}^{\text{TCR·pMHC}}[\text{TCRp·CD·pMHC}] + \alpha\, k_{\text{on}}^{\text{TCR·CD}}[\text{CD·pMHC·TCRp}] \\
&\quad - [\text{Complex·p}]\left(k_{\text{kinase}} + k_{\text{off}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·pMHC}} + k_{\text{off}}^{\text{TCR·CD}}\right) \\
\frac{d[\text{Complex·pp}]}{dt} &= k_{\text{kinase}}[\text{Complex·p}] + \alpha\, k_{\text{on}}^{\text{pMHC·CD}}[\text{CD·TCRpp·pMHC}] \\
&\quad + \alpha\, k_{\text{on}}^{\text{TCR·pMHC}}[\text{TCRpp·CD·pMHC}] + \alpha\, k_{\text{on}}^{\text{TCR·CD}}[\text{CD·pMHC·TCRpp}] \\
&\quad - [\text{Complex·pp}]\left(k_{\text{off}}^{\text{pMHC·CD}} + k_{\text{off}}^{\text{TCR·pMHC}} + k_{\text{off}}^{\text{TCR·CD}}\right)
\end{aligned}
$$

Vita

Aaron Prescott completed his BS in Biochemistry at Washington State University in 2010. Prior to this, he worked as a laboratory technician in the US Navy charged with monitoring and maintaining radiological and chemical conditions for the nuclear propulsion plants aboard aircraft carriers. After acquiring his BS, he spent two years working for the US Department of Agriculture as a lab technician. While at this position, he utilized molecular and microbiology methods to study soil-borne pathogens affecting cereal grains in no-till cropping systems. He came to the University of Tennessee in 2013 and has been conducting research as a PhD student under the guidance of Steven Abel, Assistant Professor of Chemical and Biomolecular Engineering. His research uses computational methods to understand emergent biological phenomena.