Unraveling the complex regulatory relationships between metabolism and signal transduction in breast cancer

By

Michelle L. Wynn

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Bioinformatics)
in the University of Michigan
2013

Doctoral Committee:

Professor Sofia D. Merajver, Co-Chair
Associate Professor Santiago Schnell, Co-Chair
Professor Philip C. Andrews
Professor Charles F. Burant
Professor Jennifer J. Linderman

Acknowledgements

The work presented in this thesis was funded in part by the Rackham Merit Fellowship, the Center for the Education of Women, the CCMB Pilot Grant Program, the National Science Foundation, the National Institutes of Health, and the Breast Cancer Research Foundation.

I am indebted to a number of people for their help and support throughout my graduate career at the University of Michigan. First, I would like to thank both my mentors, Santiago Schnell and Sofia Merajver. I will be forever grateful for their mentorship and friendship. I am also grateful for the support I received from members of their laboratories. I am especially indebted to Lauren Van Wassenhove, Zhi Fen Wu, and Li Wei Bao for helping me with many of the cell culture experiments included in this work. I also thank the other members of my thesis committee, Jennifer Linderman, Philip Andrews, and Charles Burant, for their continued advice and support over the last few years. In addition, I owe a debt of gratitude to Alejandra Ventura, Megan Egbert, Firas Midani and Nikita Consul, all of whom helped in invaluable ways with theoretical modeling projects. A special thank you to Charles Evans is also required for performing LC-MS analysis as well as providing a frequent sounding board for metabolic questions.

Finally, I am very grateful for my friends and family and am especially grateful for my husband, Eric, without whom this journey would have been immeasurably harder.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

2/3PG - 2-phosphoglycerate and 3-phosphoglycerate pool

AKG - alpha-ketoglutarate

ASP - aspartate

BAD - bcl-2-associated death promoter

DCIS - ductal carcinoma in situ

E4P - erythrose 4-phosphate

EGF - epidermal growth factor

ERK - extracellular signal-regulated kinases

F6P - fructose 6-phosphate

FAS - fatty acid synthesis

FBP - fructose 1,6-bisphosphate

G6P - glucose 6-phosphate

GAP - or glyceraldehyde 3-phosphate

GL-OH3P - Glycerol 3-phosphate

GLU - glutamate

GLUT - glucose transporter proteins

GLY - glycine

GPG - 6-phosphogluconate

GSK3ß - Glycogen synthase kinase 3 beta

IBC - inflammatory breast cancer

IDH - isocitrate dehydrogenase

IRS1 - insulin receptor substrate 1

KRAS - v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (isoform of RAS)

LY - LY294002, a PI3K inhibitor

MAPK - Mitogen-activated protein kinase

MEETHF - methylenetetrahydrofolate

MM - Michaelis-Menten

mTOR - mammalian target of rapamycin

NAA - N-acetylaspartate

NAAG - N-acetyl-aspartylglutamate

NTS - nucleotide synthesis

OAA - oxaloacetate

PEP - phosphoenolpyruvate

PFK1 - phosphofructokinase-1

PFK2 - phosphofructokinase-2

PI3K - phosphatidylinositol 3-kinase

PIK3CA - gene that encodes the catalytic subunit of PI3K

PIP3 - hosphatidylinositol (3,4,5)-triphosphate

PPP - pentose phosphate pathway

PTEN - phosphatase and tensin homolog

R5P - ribulose 5-phosphate

S7P - sedoheptulose 7-phosphate

TCA - tricarboxylic acid cycle

TGS - triglyceride synthesis

U0 - U0126, a MEK inhibitor

Abstract

Almost all cancer cells exhibit a metabolic phenotype characterized by high rates of glucose uptake known as the Warburg effect. Metabolic changes that are representative of distinct stages of breast cancer may suggest dependence by cancer cells on certain metabolic processes that are not relevant to normal cells. Importantly, these differences may help identify therapeutic targets that are non-lethal to normal cells. In this thesis, I present a set of models and methodologies developed using both experimental and theoretical approaches to investigate the complex intracellular signaling and metabolic networks associated with distinct stages of breast cancer. First, a detailed literature search was used to construct a logic network model of the PI3K signaling pathway, which is known to play an important regulatory role in glucose metabolism. Comparisons of experimental and simulated results suggest that the network model is well constructed but some regulatory crosstalk with MAPK requires additional refinement. Targeted therapeutic inhibitors frequently induce off-target effects. This thesis also explored the role of retroactivity in generating off-target effects in signaling networks using a computational model. The simulation results suggest that the kinetics governing covalently modified cycles in a cascade are more important for propagating an upstream off-target effect via retroactivity than the binding affinity of a drug to a targeted protein, which is a commonly optimized property in drug development. Finally, ^{13}C tracer experiments were used to infer relative glucose and glutamine derived flux in cell lines representing distinct stages of breast cancer. Steady state metabolic flux analysis was also used to computationally fit the absolute TCA cycle flux in these cell lines. Variations in acetyl-CoA, citrate, pyruvate, and alpha-ketoglutarate flux were identified. A particularly important finding was a large reductive carboxylation flux from alpha-ketoglutarate to citrate in SUM-149 cells. Together,

the models presented in this thesis provide a framework for identifying mechanistic drivers of the Warburg effect, which may represent important therapeutic targets for modulating cancer proliferation and progression.

Chapter 1
Background and Introduction

Portions of this chapter were previously published by the author in *Advances in Experimental Medicine and Biology* [1].

1.1 Background and significance

Despite the enormous genetic diversity found within a single tumor and across different cancers, the same signaling pathways are routinely altered in cancer cells [2]. A net result of this dysregulation is almost always the deregulation of apoptotic programs and the upregulation of survival and proliferation programs. As these interconnected programs are altered, a cancer cell's metabolic processes also undergo changes and adaptations to manage energy and nutrient needs. A critical component of a cancer cell's survival program is an altered metabolic phenotype characterized by high rates of glucose uptake and glycolysis, even under aerobic conditions. This altered metabolism, first described by Otto Warburg [3], is referred to as the Warburg effect or "aerobic glycolysis" [4-7] and is so pervasive among cancers that it is routinely leveraged in the clinic with fluorodeoxyglucose-positron emission tomography (FDG-PET). In general, high tumor glucose uptake observed by an FDG-PET scan correlates with the number of metabolically active tumor cells and their proliferation rates, both of which are robust prognostic indicators of adverse outcomes [5, 8].

The Warburg effect is considered paradoxical because, despite the high energy demands needed for rapid proliferation, this form of non-oxidative metabolism generates only 2 molecules of ATP per molecule of glucose. In contrast, under normal oxidative metabolism, cells generate approximately 36 molecules of ATP. While a tumor's reliance on an inefficient form of energy

metabolism may appear paradoxical, evidence suggests that the use of non-oxidative metabolism to fuel proliferation is evolutionarily conserved across unicellular and multicellular organisms [6]. There is also evidence that cancer cells divert a significant amount of carbon from glucose to nucleotide and fatty acid synthesis [5], suggesting that the dependence on glycolysis exhibited by cancer cells likely sustains the high biosynthetic requirements of rapid proliferation. Moreover, aggressive cancer cells appear to be increasingly dependent on glutamine metabolism [5], and some cancer cells appear unable to proliferate *in vitro* without sufficient extracellular glutamine [4, 5, 7]. It should be noted that there is debate about whether increased glucose uptake translates into net glycolytic ATP gain in cancer [9]. It is possible that a significant amount of the glucose uptake in cancer cells is shunted to pathways other than glycolysis (e.g., to the pentose phosphate pathway). Nonetheless, metabolic transformation is increasingly recognized as an essential hallmark of cancer [5, 10].

While the Warburg effect appears to be intrinsic to most cancers, a precise etiology remains elusive. Both oncogenic signaling [6, 11] and interactions with the tumor microenvironment [12] are likely involved in the induction and maintenance of the malignant metabolic phenotype.

1.2 Targeted cancer therapies

Targeted molecular inhibition is now a leading anti-cancer strategy [13, 14]. A number of targeted therapies are in wide clinical use and many more are in clinical trials. The phosphatidylinositol 3-kinase (PI3K) pathway is one the most commonly altered signaling pathways in human cancers [15] and AKT, a key component of this pathway, is known to play a critical role in stimulating glycolysis as well as inhibiting apoptosis [16, 17]. As a result, the PI3K pathway is viewed as a promising anti-cancer therapeutic target and a number of drugs targeting this pathway are in clinical trials [18-20].

In practice, only a small percentage of targeted therapies are approved for clinical use, however, because of harmful off-target effects or lower than expected efficacy, despite promising pre-clinical data [14]. For targeted

therapies approved for clinical use, the emergence of drug resistant clones is common [13]. The large number of off-target effects associated with targeted cancer inhibitors was recently termed the “whack a mole problem” [13] because the inhibition of a target molecule often results in the unexpected activation of one or more non-targeted molecules. In the cancer research community, off-target effects are generally attributed to non-specific drug interactions or non-linear regulatory feedback within an intracellular network, sometimes without explicit evidence. In **Chapter 4**, it is demonstrated that off-target effects may also arise via a process termed retroactivity [21], where a downstream perturbation in a signaling network induces an upstream response without any direct regulatory feedback connections. This non-linear behavior may be an important but overlooked source of off-target effects.

It is increasingly likely that the inability of many targeted cancer therapies to successfully control the disease when administered as monotherapy is related to the complexity of non-linear intracellular networks as well as the complexity associated with a tumor as a whole. Tumors exhibit a variety of non-linear, emergent, and adaptive behaviors and consist of many layers of spatial complexity. Given the extreme complexity of intracellular signaling networks and the regulatory feedback connections between component pathways and modules, the use of experimental intuition alone will not be enough to make robust predictions about where off-target effects are likely to occur in a network. The development of computational or mathematical models that can simulate a non-linear response in a complex signaling network as a consequence of a molecular perturbation (such as treatment with one or more targeted inhibitors) is needed.

1.3 Cancer systems biology

Systems biology is the integration of theoretical and experimental methods to build predictive models of complex biological phenomena. The objective of cancer systems biology is to understand the interactions that occur across microscopic and macroscopic scales in a tumor and, critically, to exploit these interactions in a predictive way. Ideally, models built using cancer systems

biology methods will have translational significance and can be used to predict candidate therapeutic targets as well as their likelihood to induce an off-target effect or therapeutic resistance.

Several important studies have identified molecular differences between normal and malignant metabolism [17, 22-27]. Genomic and proteomic approaches are increasingly used to identify differences in expression levels between normal and malignant phenotypes [23, 28]. The current statistical methods associated with high-throughput approaches cannot always distinguish small fold changes from experimental noise [29], however. Because an expression change in a metabolic enzyme or metabolite of only 1 or 2 fold is often sufficient to substantially alter the rate of a metabolic reaction, high-throughput methods should ideally be complemented with methods that can detect smaller changes in metabolic reaction rates [29].

Modeling intracellular biochemical processes

It is never practical to create a model that is an exact replica of a complex system, and trade-offs must be made between the scope and level of detail included in a model [30]. Complex cellular processes are commonly modeled with systems of continuous ordinary (ODE) or partial (PDE) differential equations. ODE and PDE models are built from underlying biophysical principles and, as a consequence, are inherently predictive. The use of continuous ODE based approximations is justified when the system is assumed to be well mixed and the number of molecules of a given reactant ranges from 100 to 1000 [31]. ODE based systems, which are commonly employed to build models of protein signal transduction and metabolism, are generally based on mass action and Michaelis-Menten (MM) kinetics [30, 32-34]. MM kinetics depends on the quasi-steady-state approximation that assumes the formation of the complex occurs on a much faster timescale than the other reactants involved in the reaction. When building theoretical models of biochemical processes, it is important to recognize that these assumptions may be invalid under physiological conditions [35, 36].

An alternative to ODE based kinetic models are stoichiometric models where the known structure of a chemical pathway is used to understand the state

of the system under a set of specific conditions. The methods assume an optimization function (e.g., the goal of bacteria is continual production of biomass) and have demonstrated predictive power in prokaryotes. Because these methods do not include any regulatory or kinetic information in the model formulation [37], they lack predictive power for multifunctional mammalian cells [38], however. It would be extremely difficult to define an optimization function that adequately captures the complexity of a mammalian cell. Kinetic ODE models, therefore, tend to be more predictive than stoichiometric models because they can predict temporal dynamics. Kinetic ODE models require more knowledge *a priori* [37] than stoichiometric models, however, and this information is not always readily available.

Modeling signal transduction

A large number of ODE based models of signal transduction can be found in the literature [33, 34, 39-46]. In contrast to central carbon metabolism, however, significant information about the structure of signal transduction networks is often not known *a priori*, making ODE models of large scale signaling networks difficult to accurately construct. In addition, large scale ODE models are limited by extremely large and poorly defined parameter spaces. Alternative methods for modeling signal transduction networks include Bayesian network analysis, structural network analysis, Markov models, and discrete logic based models [47]. A number of logic-based network models of gene regulation and signal transduction have generated experimentally valid predictions [47-51]. Logic models are a parameter free method that can be used in the absence of detailed kinetic and mechanistic information. The theoretical underpinnings and limits of logic modeling are explored in **Chapter 2**.

Modeling cellular metabolism

Many ODE based models of metabolism exist in the literature [38, 52-54]. In general, metabolism is the set of enzyme catalyzed reactions that break down or synthesize molecules in a cell [32]. Enzymes regulate metabolism by catalyzing these reactions [55]. Specifically, an enzyme reacts selectively with a substrate and transforms it into a product. In experimental studies of

metabolism, enzyme concentrations are generally assumed to be constant during the catalyzed transformation of substrates into products [56, 57].

The most widely used theories of metabolic regulation are biochemical systems theory [58-60], metabolic control theory [61-63], and flux-oriented theory [64-66]. All three of these theories are applications of sensitivity analysis applied to biochemical reaction models. The models consist of coupled ODEs based on the law of mass action. Sensitivity analysis is used to investigate the effects of parameter value changes on model behavior [67]. It is not surprising, then, that the primary difference between these theories is the choice of which parameters to vary when evaluating model sensitivity [67, 68]. Another common attribute of these theories is the assumption of constant enzyme concentration [68, 69]. The assumption of constant enzyme concentration has been questioned for some time, however [70]. Enzymes are not indefinitely stable; they too are metabolites like their substrates and products [32]. The synthesis of enzymes is an essential part of metabolism and is catalyzed by other enzymes. A theory to investigate metabolic regulation in cancer cells that takes into account enzyme production and depletion is needed in system biology.

1.4 Modeling breast cancer complexity

In the United States, breast cancer is a leading cause of cancer deaths among women, second only to lung cancer [71]. Mammary ductal carcinoma *in situ* (DCIS) is an early form of breast cancer that occurs when luminal epithelial cells in a mammary duct begin to proliferate abnormally and disturb the ductal architecture. DCIS advances to invasive breast cancer when proliferating cells breach the basement membrane and begin the process of metastasis [72] (**Figure 1.1**), the most lethal stage of cancer. The 5-year survival rate after diagnosis with metastatic breast cancer is 23%, while the 5-year survival rate for other stages of breast cancer is greater than 80% [71]. A distinct breast cancer phenotype known as inflammatory breast cancer (IBC) represents the most deadly form of the disease [73, 74], with more than 90% of patients exhibiting metastases within 1 year of diagnosis [75]. New treatment strategies for IBC and other aggressive breast cancers are critically needed.

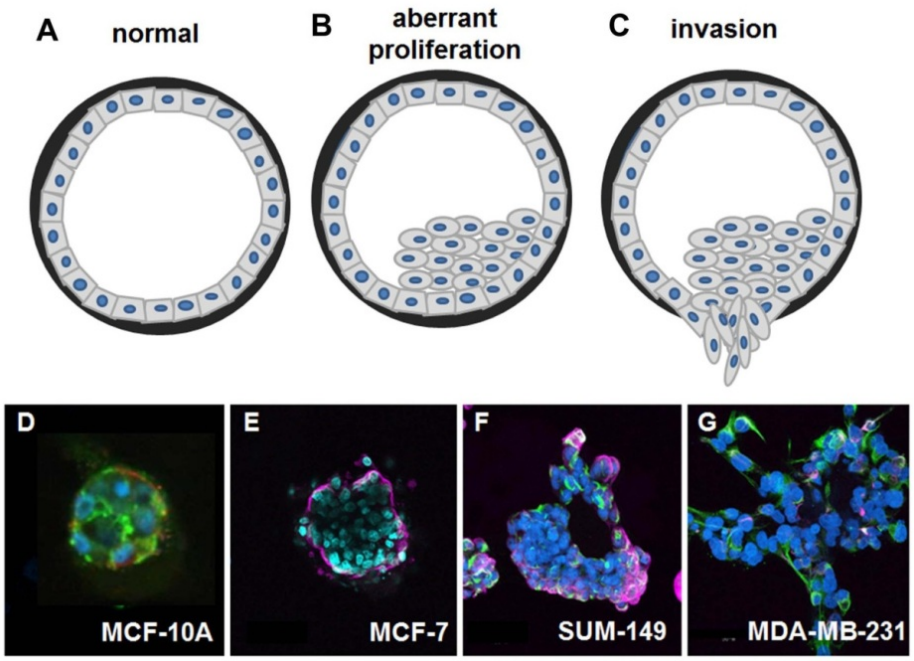


Figure 1.1. Breast cancer progression. (**A**) Normal adult mammary glands are hollow tube-like structures surrounded by polarized luminal epithelial cells (grey). (**B**) In ductal carcinoma development, aberrant luminal cell proliferation ensues, which may be followed by (**C**) invasion if the basement membrane surrounding the myopethelium is breached (black). (**D**-**G**) Representative images of 3-dimensional breast cell culture performed by the Merajver lab and imaged with fluorescent confocal microscopy. (**D**) The MCF-10A normal-like mammary epithelial cells form well-organized acini with empty central lumen resembling normal mammary ducts; compare to (**A**). In contrast, the tumorigenic (**E**) MCF-7 cancer cells form defined acini with filled lumen resembling ductal carcinoma *in situ*; compare to (**B**). (**F**) SUM -149 and (**G**) MDA-MB-231 are metastatic cell lines that exhibit forms of invasion and loss of organization; compare to (**C**).

The Warburg effect in cancer is similar to proliferative metabolism in normal cells but molecular and kinetic differences have been discovered. *I hypothesize that there are key drivers of the Warburg effect that can be exploited to target breast cancer proliferation without substantial negative effects on normal tissue growth*. Because the intracellular networks involved in regulating cancer metabolism are highly complex and interconnected, identifying these drivers will require systems biology approaches.

The challenges associated with cancer's complexity are well recognized within the cancer systems biology community [76-84]. While a number of models of cancer complexity have been developed [51, 82, 85-88], few have included proliferative metabolism, or its regulation, in detail. [89, 90]. In this thesis, I present a set of models and methodologies developed using principles of cancer

systems biology to investigate the complex intracellular networks in distinct stages of breast cancer, including metabolic networks. The models in **Chapter 3** and **Chapter 5** rely on an *in vitro* model of breast cancer progression, where normal tissue is modeled using data from normal-like MCF-10A mammary epithelial cells lines and malignant tissue is modeled using data from three tumorigenic cell lines: MCF-7 (non-metastatic breast cancer), MDA-MB-231 (highly metastatic, triple negative breast cancer), and SUM-149 (highly aggressive metastatic IBC) (**Figure 1.1D-G**). The objective of all models presented is the elucidation of system level properties that would be difficult to discern by experimental methods alone. The models presented also provide a framework for identifying mechanistic drivers of the Warburg effect, which I hypothesize may be important potential therapeutic targets.

Chapter 2
Logic-based models: a predictive and parameter-free network analysis method

The content of this chapter was previously published by the author as a critical review in *Integrative Biology* [91].

2.1 Introduction

The emergence of molecular biology has produced a vast literature on the cellular function of individual genes and their protein products. It has also generated massive amounts of molecular interaction data derived from high-throughput methods as well as more classical low-throughput methods, such as immunoprecipitation, immunoblotting, and yeast two-hybrid systems. From this accumulation of interaction data, researchers can now attempt to reconstruct and analyze the highly complex molecular networks involved in cellular function.

The objective of systems biology is to integrate experimental data with theoretical methods to build predictive models of complex biological processes across a variety of spatial and temporal scales. Two very different paradigms of system biology are frequently used to construct and analyze network models of molecular interactions inferred from experimental data: structural network analysis methods and mathematical models based on differential equations. A third increasingly important network analysis paradigm in systems biology is the application of logic-based methods to generate predictive output [92, 93]. Although qualitative in nature, logic-based methods have the capacity to provide insights into the dynamics of highly complex gene regulatory and signal transduction networks without the burden of large parameter spaces.

Understanding the networks associated with cancer offers especially difficult challenges. Fundamental problems in understanding the transition from

the normal to near normal to dysplastic to neoplastic to metastatic states of cancer progression can theoretically be modeled by longitudinal comparisons of networks in which, as progression occurs, certain molecular interactions are rendered stronger (for instance through gene amplification) or lost (through mutation, deletion, down-regulation, or methylation). Logic models provide a framework in which these types of network comparisons are possible. Multi-state logic models can simulate signal amplification, and random order asynchronous logic models can simulate the heterogeneous response in a population of cells to diverse stimuli. Logic-models are also well suited for performing *in silico* molecular perturbations, which could be used to predict a population level response to a targeted therapy or a combination of therapies. In this review, we provide a tutorial on the use of logic-based methods as well as a discussion of their limitations, using biologically motivated examples.

Modeling intracellular networks

Typically, knowledge of molecular interactions is summarized in diagrams of varying complexity, commonly known as interaction networks [94]. In an interaction network diagram, each node represents a molecule and a line drawn between two nodes represents a molecular interaction, also referred to as an edge in graph theory. If the nature of an interaction between two nodes is known (e.g., which molecule is the regulator that activates or inhibits the other molecule), the edge is said to have directionality. If a correlation between the activities or expression levels of two nodes is known but the causal relationship underlying their interaction is not, the edge is said to lack directionality.

Structural network methods

Structural network analysis provides a picture of the correlations between molecules in very large networks. In structural network models, which are usually derived from high-throughput genomic or proteomic experimental methods, the directionality of interactions in the network is generally not known and it is the static correlations in expression patterns that are important. The primary objective of these methods is to infer functional patterns in large networks using statistical methods [95]. These methods are also used to

construct species specific interactomes [94]. A limitation of these methods, however, is that they generally provide only a static view of molecular interactions in a network at single point in a time. Additionally, the current experimental methods used to generate data for structural network models are extremely noisy [96], which further limits the predictive power of this method.

Differential equation methods

As discussed in **Chapter 1**, systems of ordinary differential equations (ODEs) are frequently used to model biochemical reactions involved in gene and protein regulation. In these models, information about the mechanistic nature of the interaction is essential and edges between species must be directional. ODE models are built from underlying biophysical principles, such as biochemical rate laws and the conservation of mass and energy. Consequently, ODE models have the capacity to be highly predictive [97-99]. This predictive power translates into the ability to generate a dynamic view of the concentration of each interacting species in the network over time as well as the ability to identify biologically realistic steady states [98]. The predictive power of ODE systems is dependent, however, on large numbers of kinetic parameters that are rarely known with any degree of certainty. These powerful methods are, therefore, limited both by the enormous parameter spaces involved in even a relatively simple network and by their need for detailed mechanistic knowledge *a priori*.

Logic-based methods

Logic-based network models were pioneered in the biomedical sciences by Kauffman [100, 101] and represent a compromise between structural analysis and ODE methods in terms of precision and complexity [97]. While logic models do not require mechanistic knowledge of interactions, they do require knowledge of edge directionality. In their simplest form, logic-based models permit each biochemical species (represented as nodes in a network) to be in one of two discrete states: ***ON*** or ***OFF***. The state of a logic network evolves in a dynamic fashion as nodes in the network are switched ***ON*** and ***OFF*** according to the state of other nodes in the network, until the network settles into an unchanging state, often referred to as an attractor [93]. Logic-based models with only two binary

states are generally known as Boolean models. While there is no explicit notion of time in a logic model, each round of updating can be considered an arbitrary time unit.

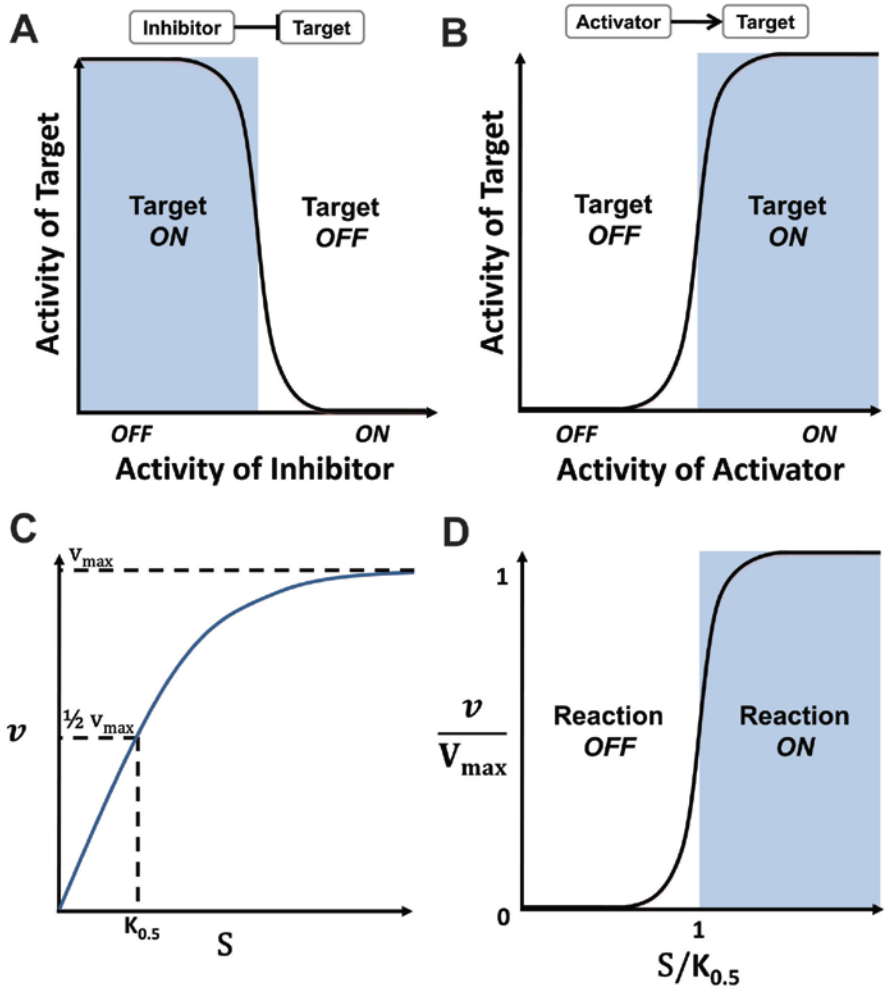


Figure 2.1. Logic models approximate molecular interactions. Logic models approximate molecular interactions. The motivation behind logic models comes from the sigmoidal relationship observed between regulatory molecules and their target molecules. This relationship approximates a Boolean switch and can be thought of as having two states: saturated (***ON***) and un-saturated (***OFF***). (**A**) An inhibitor regulates a target molecule in a logic model: the target remains ***ON*** until the inhibitor's activity surpasses a threshold of activation. (**B**) An activator regulates a target molecule in a logic model: the target remains ***OFF*** until the activator's activity surpasses a threshold of activation. In these models, ***ON*** means the molecule has reached a threshold of functional activation that is high enough to affect the state of other molecule(s) it directly regulates, while ***OFF*** means the molecule is not present at a high enough level to affect the molecules it directly regulates. Critically, ***OFF*** does not necessarily mean the molecule has a zero concentration. (**C**) A standard hyperbolic saturation curve measuring reaction velocity, *v*, as substrate concentration, **S**, increases. $\mathbf{K_{0.5}}$ is the substrate concentration needed to reach half the maximum velocity, $\mathbf{V_{max}}$. (**D**) A plot of $v/\mathbf{V_{max}}$ vs. $\mathbf{S/K_{0.5}}$ (the specific substrate concentration) takes the form of a sigmoidal saturation curve approximating a Boolean-like switch. When $\mathbf{S/K_{0.5}} << 1$, the reaction is effectively ***OFF***. When $\mathbf{S/K_{0.5}} >> 1$, the reaction is effectively ***ON***.

2.2 Logic-based models approximate biochemical regulation

The assumption that a molecule can have only two possible states is a simplification of biological complexity. It is a reasonable regulatory approximation, however, given the switch-like sigmoidal relationship often observed between an affector molecule and its target molecule (**Figure 2.1A-B**) [49, 93, 102, 103]. It is important to emphasize that when a molecule's node is ***OFF*** in a discrete logic model, it does not imply that the molecule has zero concentration in the system. Instead, it implies that the molecule is not present at a high enough level to induce a change in the molecules it directly regulates [103]. When a molecule's node is ***ON*** in a logic model, it means the molecule has reached a threshold of functional activation that is high enough to affect the state of the molecules it directly regulates. More specifically, a target molecule will remain ***OFF*** in a logic model until its activator reaches a specific threshold of activity (**Figure 2.1B**). Likewise, a target molecule will remain ***ON*** in a logic model until its inhibitor reaches a specific threshold of activity (**Figure 2.1A**) [103]. As a consequence, logic models can only provide qualitative approximations of molecular regulation. While this represents a limitation of the methodology, in reality, the majority of experimental data available on molecular regulations are also qualitative in nature [104].

More complex logic-based methods have been developed, such as multi-state and fuzzy logic methods, which permit nodes to be in more than two discrete states. In addition, logic models that allow node states to vary continuously between states (e.g., from 0 to 1) have also been developed [47, 49, 51, 97]. Although theoretically able to more precisely simulate biochemical regulation [47], these more complicated approaches require parameter value estimates that are rarely known and, in some cases, are difficult to correlate with biophysical chemistry theory. Thus, discrete two-state logic models (Boolean models) are an intuitive and predictive method for describing biochemical interactions without requiring prior knowledge of complex mechanistic details of reaction kinetics (needed for ODE systems) or degrees of membership (needed for multi-state fuzzy logic systems).

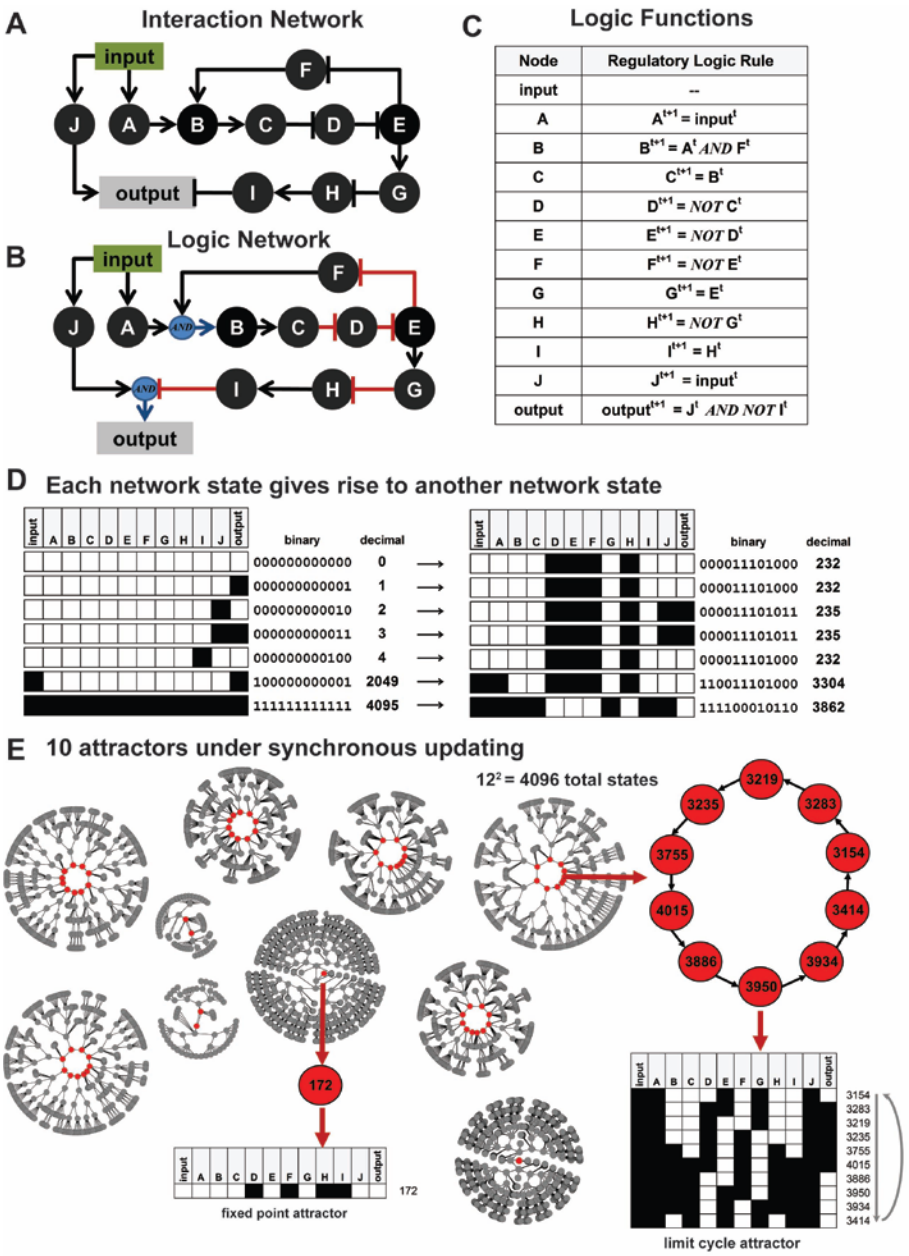


Figure 2.2. A hypothetical network example with 12 nodes. (**A**) An interaction network diagram summarizing a hypothetical network with 12 nodes where **input** is the only input node, **output** is the only output node, and **A**–**J** are internally regulated nodes. (**B**) A logic network diagram explicitly identifying how multiple signals (edges in the diagram) will be integrated to produce a response in nodes **B** and **output**, the only nodes with more than one regulator. (**C**) A table explicitly listing the logic functions used in the hypothetical logic model. The information in this table is equivalent to the information available in (**B**), and both (**B**) and (**C**) contain more information than (**A**). (**D**) Each unique network state in a logic network can be identified in a number of ways: a graphical table where white = ***OFF*** and black = ***ON***, a binary representation where 0 = ***OFF*** and 1 = ***ON***, or a decimal value that is equivalent to the binary representation. Each of these states will give rise to another state based on the logic functions (**B**–**C**) that control signal transfer in the network. (**E**) Eventually, all states will settle into an attractor, a state from which it is not possible to escape without an external perturbation to the system. In the graphs, attractor states are colored red and non-attractor states are colored grey. An example of a limit cycle and a point attractor are highlighted. In this model, there are a total of 2^{12} (4096) possible states that settle into 1 of 10 attractors. Blue circles indicate ***AND*** interactions.

Moreover, Boolean models can produce the same qualitative output as more quantitatively precise ODE models when well-constructed. For example, Albert and Othmer [48] used a Boolean model of the genetic regulation of segmentation patterns in *Drosophila* to produce results that were in close agreement with an earlier ODE model of the same system [105]. Fauré et al. [104] analyzed a simple Boolean model of cell cycle regulation and found qualitative steady state agreement with a complex ODE model [42]. More recently, Akman et al.[106] demonstrated that a series of Boolean models produce the same qualitative output as a series of ODE models of circadian clock regulation. In addition, Boolean models are well suited for the testing of hypothesized regulatory mechanisms [48, 107] and for helping to direct future experiments [50, 108-110]. They are also useful for performing a preliminary network analysis [110] prior to developing more detailed experimental or theoretical models. For all these reasons, the development and analysis of two-state Boolean models will be the primary focus of this review.

2.2.1 A basis for the Boolean approximation is also found in chemical reaction kinetics

In addition to the switch-like regulatory dependence described above, another switch-like behavior in biochemical systems is related to the substrate concentration (S) needed to reach half the maximum velocity (V_{max}) of a reaction, commonly represented by $K_{0.5}$. Enzymologists routinely use saturation curves depicting how reaction velocity (*v*) varies with S (**Figure 2.1C**) to estimate the $K_{0.5}$ of a reaction. In reality, *v* is not dependent on S but is, instead, dependent on the specific substrate concentration, defined as the ratio of S to $K_{0.5}$ ($S/K_{0.5}$). The well-known constant K_m is the substrate concentration needed to reach $K_{0.5}$ under Michaelis-Menten (MM) kinetics [32]. While $K_{0.5}$ is sometimes called the "apparent K_m", $K_{0.5}$ is not restricted to kinetic mechanisms that follow the MM approximation.

The standard MM expression for the velocity (v_{MM}) of a non-reversible enzyme catalyzed reaction is presented in equation (1). This expression is similar to the generalized expression for the velocity (*v*) of any enzyme catalyzed

reaction presented in equation (2), where S is converted to a product. Here, *n* refers to the Hill coefficient describing the degree of cooperativity in the reaction (i.e., positive, negative, or no cooperativity). Dividing both sides of equation (2) by V_{max} and factoring $K_{0.5}$ out of the right-hand side, gives equation (3), which is now in the form commonly used to normalize reaction velocities (*v'*). Equation (3) also resembles a standard Hill equation.

$$v_{\mathrm{MM}} = \frac{\mathrm{V}_{max}\mathrm{S}}{\mathrm{K_m} + \mathrm{S}} \quad (1) \qquad v = \frac{\mathrm{V}_{max}\,\mathrm{S}^n}{\mathrm{K}_{0.5}^n + \mathrm{S}^n} \quad (2)$$

$$v' = \frac{v}{\mathrm{V}_{max}} = \frac{\left(\frac{\mathrm{S}}{\mathrm{K}_{0.5}}\right)^n}{1 + \left(\frac{\mathrm{S}}{\mathrm{K}_{0.5}}\right)^n} \quad (3)$$

In equation (3), when $S/K_{0.5} = 1$, then $v' = 1/2$. Consequently, there are two distinct regions where the normalized reaction velocity responds in a characteristic way to $S/K_{0.5}$ (**Figure 2.1D**). When $S/K_{0.5} << 1$, then $S < (0.01 \times K_{0.5})$, meaning the enzyme is not saturated and the reaction rate is linear with respect to the substrate concentration. In contrast, when $S/K_{0.5} >> 1$, then $S > (100 \times K_{0.5})$, meaning the enzyme is saturated, the reaction rate is independent of the substrate concentration, and the reaction has reached (or is very near to) its maximum velocity. Thus, the specific substrate concentration serves as the on-off switch of a reaction. In logical terms, when the node representing S is ***OFF*** (or 0), then the specific substrate concentration will be less than 1 and the reaction cannot not proceed. Likewise, when the node representing S is ***ON*** (or 1), then the specific substrate concentration will be greater than 1 and the reaction will proceed at a rate near V_{max} (**Figure 2.1D**).

The use of logic to describe chemical change is not limited to the rate of biochemical reactions. Molecular substrates have been described as computational devices that process physical and chemical inputs into outputs according to Boolean logic [111]. Molecular logical operations are achieved by leveraging observable chemical changes to create a computational device.

These devices, for example, can be used to solve arithmetic or logical operations by exploiting changes in the conformations of chemical components. In the laboratory, molecular logic functions have been developed that rely on charge transfer, which affects the fluorescent state of a molecule [112, 113]. Logical functions have also been created to exploit charge transfers in cascades of coupled enzymes [114, 115]. In fact, molecular logic operations serve as the basis for many nanosensors currently used in the basic sciences, industry, and medicine [111].

2.2.2 Predictive logic-based models in the literature

Although the theoretical underpinnings of logic models provide qualitative approximations of molecular and biochemical regulation, in reality these models can only generate predictive output when the logic model is well constructed. Several examples of well-constructed logic models providing good agreement with experimental data exist in the literature [47, 48, 50, 51, 116-119]**.**

For example, Li et al. [117] constructed a Boolean model of the genetic network controlling the cell cycle in *Saccharomyces cerevisiae.* The authors found that 86% of the 2048 possible states in the network settled into a steady state (also known as a stable attractor) that corresponded to the G1 stationary phase of the cell cycle. Their analysis suggested that the regulatory network controlling the yeast cell cycle is resistant to stochastic perturbations. The authors interpreted these findings to mean that robustness in the underlying network is advantageous for the organism because, under normal conditions there is a high probability that the regulatory network dynamics will settle into the G1 state regardless of the current state of the network. Once in the G1 state, the network will remain in that state until a significant external signal perturbs the network and initiates another round of cellular division. Subsequent work by Davidich and Bornholdt [120] used a similar Boolean approach to study the cell cycle regulation in *Schizosaccharomyces pombe*. These authors found that the majority of *S. pombe* network states settle into a steady state corresponding to the G1 stationary phase, which is in agreement with the results of the *S. cerevisiae* model. However, they also found significant differences in the

regulatory network of *S. pombe* compared to *S. cerevisiae*, which yielded very different network dynamics.

Using a somewhat different logic-based approach, Bolouri and colleagues [110, 121] constructed an *a priori* gene regulatory network of endomesoderm specification control in sea urchin embryos. The network was logic-based and generated a series of testable predictions. Using computational methodologies and large-scale perturbation analyses, the authors iteratively tested and revised their model by comparing model output to biological readouts. Their use of a regulatory network to logically map inputs and outputs for cis-regulatory elements identified system level properties that would not otherwise have been observable. From this information, the authors were able to draw important conclusions about the developmental features of endomesoderm specification.

In addition to gene regulatory networks, logic models can also be used to model signal transduction networks. Li, Assmann, and Albert [50] developed a Boolean model of the signal transduction network controlling abscisic acid regulation of stomatal closure in plants. The authors employed a network construction approach that inferred indirect molecular relationships from data to build the sparsest logic network possible that was compatible with available experimental data. A random order asynchronous Boolean approach was then used to simulate the heterogeneity in a population of cells. The model results were in good agreement with previous experimental findings and generated novel predictions about the conditions likely to have the strongest effect on stomatal closure. In a subsequent manuscript, which also serves as an excellent tutorial, Albert et al. [49] contrasted the asynchronous approach used in the Li, Assmann, and Albert [50] model with a continuous piecewise Boolean model of the same system that allowed node states to vary between 0 and 1. They reported that the asynchronous discrete Boolean model produced the same qualitative results as the continuous piecewise Boolean model.

For additional examples of predictive logic-based models in the literature, we direct the reader to comprehensive reviews by Morris et al. [97] and Albert et al. [49]. Both reviews provide detailed case studies of logic model

implementations and demonstrate the variety of ways logic-based models can be applied to answer biological questions.

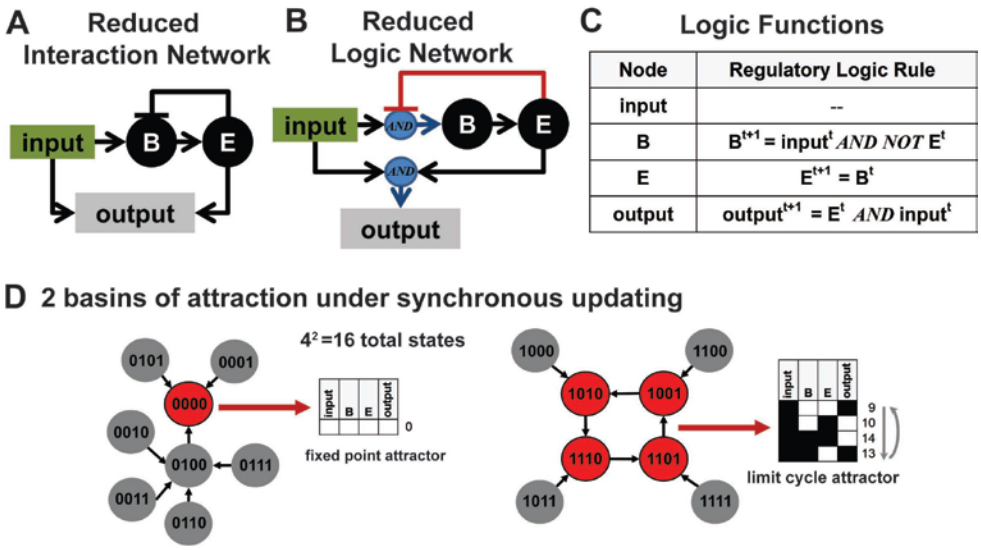


Figure 2.3. The hypothetical network example reduced to 4 nodes. (A) An interaction network diagram summarizing an alternate form of the network in **Figure 2.2A** where all linear regulations have been collapsed into single regulations. (**B**) A logic network identifying how multiple signals (edges in the diagram) will be integrated to produce a response in **B** and **output**, the only nodes with more than one regulator. (**C**) A table listing logic functions used in the reduced model. The information in this table is equivalent to the information available in (**B**), and both (**B**) and (**C**) contain more information than (**A**). (**D**) Eventually, all states will settle into an attractor. In the graphs, attractor states are colored red and non-attractor states are colored grey. In this model, there are 2^4 (16) possible states that settle into one of two attractors, which, because of their small size, can be easily visualized. The binary numbers on the nodes represent the state of each node in the network in the following order **input**, **B**, **E**, and **output**. Blue circles indicate ***AND*** interactions.

2.3 Time in a Logic Model

Each time a network is updated in a standard two-state Boolean model, signals are transferred according to logic functions in a synchronous and deterministic manner. In these models, all nodes are updated instantaneously in the same time step so that the state of the network is always fully determined by the state of all nodes in the previous time step. Thus, the underlying assumption is that all molecular interactions in the network take the same arbitrary amount of time to complete. In reality, the time it takes for molecular interactions to complete varies widely.

An alternate updating method involves asynchronous updating where one node is selected at random and instantly updated according to the current state of the network. In these models, the next state of the network is non-

deterministic. The random and instantaneous updating of a node is repeated many times with each random update representing a time step in the model. This non-deterministic updating scheme is thought to more closely resemble biological variation by eliminating temporal uniformity in the model [49, 92, 103, 122, 123]. Typically, implementation of this type of scheme involves running a large number of model simulations to calculate a probability that any given node will end up ***ON*** or ***OFF*** for any set of initial conditions [49, 50]. Synchronous and asynchronous updating methods are discussed in more detail in later sections of this chapter.

2.4 Building predictive logic-based models

Creating a logic model is relatively straight-forward. Building a logic model that can generate predictive output that can be leveraged by experimentalists, however, requires considerably more effort. Specifically, building predictive logic-based models entails two primary steps: building a detailed interaction network and translating the interaction network into a logic network. The development of a logic network includes the careful derivation of the logic functions that will drive the network's dynamics.

Build an interaction network by surveying the literature

The first step in developing a logic model is to construct an interaction network of the system to be modeled (see, for example, **Figure 2.2A**). To do so will typically require a thorough literature or database search. This is a critical step and should be performed by, or in close collaboration with, someone who is well acquainted with the biology of the system. Once the interactions involved in the network have been identified, it is often desirable to perform a node reduction to reduce network complexity, especially for very large networks. The formulation of any theoretical model requires sound judgments about which approximations are appropriate for simplifying model complexity without losing essential elements of the underlying mechanism(s) [124]. This is certainly true of logic models where decisions must be made about whether some complex interactions can be lumped into a smaller subset of nodes and interactions

(**Appendix A**). In general, the objective is to use the simplest network possible that still agrees with known experimental data. This may be done manually or with the assistance of computational algorithms [50, 125, 126]. If automated tools are used, it is useful to validate that the reduced network generated includes a suitable amount of complexity for the system and problem considered.

A common manual approach is to eliminate redundant linear regulations. For example, in **Figure 2.2A**, nodes **B** and **C** are both activators of **E** because an even number of inhibitions produces an activating regulation. In contrast, **E** is an inhibitor of **B** because there is one inhibitory regulation between **E** and **B**. A reduced version of this network is presented in **Figure 2.3.** As shown in **Appendix A** both forms of the network produce the same qualitative output.

Translate the interaction network into a logic network

Ultimately, the interaction network must be translated into a set of logic functions (also referred to as transfer functions or logic gates) that will be used to transfer information (or signals) between nodes each time the model is updated. Logic functions often contain one or more Boolean operators. The ***AND*** and ***OR*** operators are used to define how distinct signals acting on the same node (which may be stimulatory and/or inhibitory) will be processed. The ***NOT*** operator is used to negate the state of the node it precedes. The derivation of logic functions is discussed in detail in the next section.

A justification from the literature (or evidence from experimental data) should be provided to support each logic function. When individual logic functions qualitatively agree with experimental data, it is more likely that the model, as a whole, will be predictive. Ideally, a table or appendix summarizing each logic function's rationale will be included with all published models (see **Table 2.1** for an example) [48, 50, 51, 104].

From an interaction network diagram alone, it is not possible to infer how multiple signals acting on the same node should be processed. Therefore, the use of descriptive logic network diagrams is recommended to graphically depict a logic model. The information contained in the full set of logic functions should be equivalent to the information contained in a logic network diagram (compare

Figure 2.2B to **Figure 2.2C** and **Figure 2.3B** to **Figure 2.3C**). The use of descriptive diagrams to represent a biological network is not a novel concept [92, 127]. Albert and colleagues [48] proposed the use of "pseudo-nodes" and "complementary pseudo-nodes" to clarify the functional nature of edge interactions in a logic network diagram. In a large-scale Boolean model of EGFR and ERB2 signaling, Samaga et al. [119] used a graphical representation where ***AND*** interactions were depicted as small blue circles and all other interactions were assumed to be ***OR*** interactions. More recently, Morris et al. [97] presented logic network diagrams that used "logic gate" notation similar to that used in an engineering diagram.

Throughout this review, we have adopted a modified version of the notation used by Morris et al. [97] Our notation graphically illustrates how multiple edges regulating the same node will be integrated by explicitly identifying where ***AND*** and ***OR*** operators are used in the network. In addition, all activating interactions in our logic network diagrams are indicated with a black arrow and all inhibiting interactions are indicated with a red line and a blunt edge. Regardless of the graphical method used, the use of diagrammatic logic networks is strongly encouraged to remove ambiguity from interaction networks.

Deriving logic functions: the importance of characterizing each interaction

When translating a set of interactions into a logic model, the implicit assumptions underlying all logic functions must be carefully considered. We recommend the construction of truth tables for each logic function to confirm the logical output of each function is in agreement with experimental data (or that of a hypothesized regulatory mechanism). A truth table provides the logical output of all possible combinations of input values a logic function may receive. In Boolean models with only two discrete states, there are 2^r possible combinations of regulatory inputs in a truth table, where r is the number of regulators (or edges) leading into the regulated node.

For nodes with one regulator, derivation of the logic function is straightforward: the next state of the regulated node is fully determined by the current state of its only regulator (**Figure 2.4A**). An example of the two ways a

single molecule can regulate another molecule in a Boolean model is presented in **Figure 2.4A**, along with corresponding truth tables. In the case where **C** is activated by **A**, the logic rule is represented as $\mathbf{C^{t+1} = A^{t}}$, which means the value of **C** in the next arbitrary time step (t+1) will be the current value of **A**. In the case where **C** is inhibited by **B**, the rule is represented as $\mathbf{C^{t+1} = \textit{NOT}\ B^{t}}$, which means the value of **C** in the next arbitrary time step (t+1) will be the inverse of the current value of **B**. Importantly, because **C** will always be ***ON*** whenever **B** is ***OFF*** in this inhibition example, the implicit meaning of this function is that of constitutive activation of **C** in the absence of **B**. If this turned out to be an inappropriate assumption for the of regulation of **C**, then additional activators of **C** would need to be added to the model in order to generate a more complex and accurate regulatory logic function for **C**.

For nodes that have multiple regulators, the development of a logic function can be more challenging. In a logic model, there are three possible mechanisms by which two nodes can regulate another node: regulation by two activators (**Figure 2.4B**), regulation by an activator and an inhibitor (**Figure 2.4C**), or regulation by two inhibitors (**Figure 2.4D**). The simple 3 node interaction networks presented in the first column of **Figure 2.4B-D** clearly indicate that 2 nodes regulate another node but do not provide precise information about how the signals will be integrated to produce a response in the regulated node. In contrast, the second and third columns of **Figure 2.4B-D** provide logic network diagrams and truth tables illustrating the logical output produced when an ***AND*** or ***OR*** operator is used, respectively. In general, the use of an ***AND*** operator with two regulators results in the regulated node turning ***ON*** in one of the four possible input conditions. In contrast, the use of an ***OR*** operator with two regulators results in the regulated node turning ***ON*** in three of the four possible input conditions. Although the examples provided are simple, there are important underlying assumptions that should be emphasized.

AND NOT: inhibitor wins

When an activator and an inhibitor are joined by an ***AND*** operator (**Figure** 2.4**C**: ***AND*** column), the function is referred to as an ***AND NOT*** function. In the

logic function $\mathbf{C^{t+1} = A^t}$ ***AND NOT*** $\mathbf{B^t}$, the inhibitor is dominant because the state of **C** is ***OFF*** in three of the four possible input conditions. The presence of the inhibitor also trumps the presence of the activator when both are ***ON*** (refer to the last row in the truth table).

OR NOT: activator wins

Similarly, when an activator and an inhibitor are joined with an ***OR*** operator (**Figure 2.4C**: ***OR*** column), the function is referred to as an ***OR NOT*** function. In this case, the activator is dominant because the state of **C** is ***ON*** in three of the four input conditions. The presence of the activator trumps the presence of the inhibitor when both are ***ON*** (refer to the last row in the truth table). Importantly, when both the activator and inhibitor are ***OFF***, **C** will turn ***ON*** (refer to the first row in the truth table). The implicit meaning of $\mathbf{C^{t+1} = A^t}$ ***OR NOT*** $\mathbf{B^t}$, therefore, is that **C** becomes activated when both regulators are absent or below their functional threshold, even if **C** was ***OFF*** in the previous time step. At first glance this may seem counter intuitive and biologically implausible. However, what the ***OR NOT*** function actually simulates is the condition where the regulated node is ubiquitously expressed at a functionally active level such that it can only be deactivated by the presence of a direct inhibitor and, importantly, this direct inhibitor can be overridden by an activator (refer to the truth table). Later, we will consider a biological example where ***OR NOT*** is the correct function to model a biological interaction.

To be OR NOT to be?

As illustrated in **Figure 2.4** and described above, when more than one node regulates the same target node, the ***AND*** and ***OR*** logical operators produce very different outcomes. It is important to understand the underlying assumptions of all logic function included in a logic model to ensure they provide a reasonable biological approximation [123]. Of course, the exact nature of an interaction may not be known (e.g., what is necessary and what is sufficient for a molecular activation). In these cases, logic models can also be used to test hypothetical interactions and compare if results match experimental data.

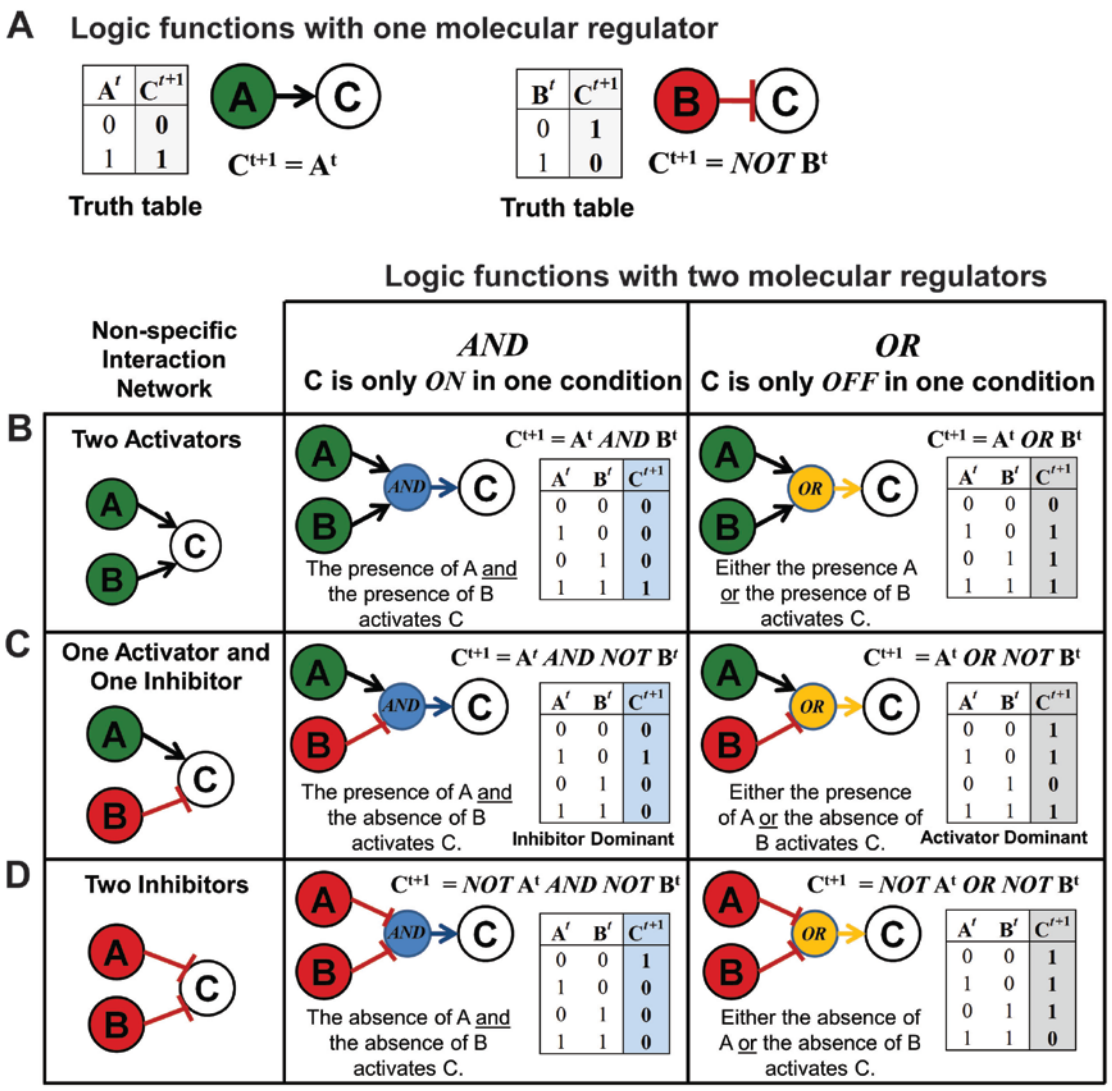


Figure 2.4. Logic Functions. Each node in a logic model has a corresponding logic function that controls its regulation each time the model is updated. (**A**) Derivation of logic functions for nodes regulated by only one node: the next state of the regulated node (at time t + 1) is fully determined by the current state of the regulating node (at time t). $\mathbf{C}^{t+1}$ indicates the value of **C** at the next time step. (**B**–**D**) When multiple nodes regulate the same node, derivation of logic functions is more complex. Three node networks are presented where **C** is regulated by two nodes: **A** and/or **B**. In the left hand column, typical representations of interaction diagrams are presented with regulation by (**B**) two activators, (**C**) an activator and an inhibitor, or (**D**) two inhibitors. The corresponding logic functions when an ***AND*** or an ***OR*** relationship is used, respectively, are presented in the next two columns. In each example, a more precise logic network is depicted that clearly indicates how the two nodes will be integrated to produce a response in **C**. A truth table is provided that illustrates the output $\mathbf{C}^{t+1}$ generated for all possible input conditions. 0 indicates ***OFF*** and 1 indicates ***ON***. For clarity, activating nodes are colored green and inhibiting nodes are colored red. Blue circles indicate ***AND*** interactions and yellow circles indicate ***OR*** interactions.

2.5 Logic models have one or more fixed states known as attractors

The network state of a logic model can be uniquely represented in a variety ways, including by binary or decimal notation (**Figure 2.2D**). In synchronously updated Boolean logic models (where all nodes are updated at the same instant each time step), each state deterministically gives rise to another state according to the model's logic functions. Eventually, all states will settle into one or more stable states known as attractors. If an attractor consists of a series of states that oscillate in a cycle, the attractor is called a cycle attractor or a limit cycle. If an attractor consists of a single fixed state (which will be the case when a state always gives rise to itself), the attractor is called a fixed point attractor [93, 128]. Examples of each type of attractor are presented in **Figure 2.2E** and **Figure 2.3D**. In asynchronously updated Boolean logic models (where a single node is selected randomly and updated instantly), the next state of the network is non-deterministic. Nevertheless, the network will eventually settle into an attractor regardless of the initial state of the network [92, 103, 123].

In small to moderate sized network models that have been well constructed, attractors often empirically agree with biological phenotypes [93, 117]. In the model previously discussed of yeast cell cycle regulation developed by Li et al. [117], for example, there were 11 nodes which generated 2048 possible network states and 7 attractors. The vast majority of all possible states (1764 out of 2048) settled into a fixed point attractor that represented the stationary phase of the cell cycle. The authors found that the trajectory of states (also referred to as a basin of attraction [128]) leading to this point attractor followed expected molecular changes observed during cell cycle progression.

In any binary Boolean logic model consisting of n nodes, there are 2^n possible network states. In the hypothetical 12 node network presented in **Figure 2.2** there are 4096 possible states, and in the synchronous form of this model there are 10 attractors (**Figure 2.2E**). In a 21 node model we recently developed, there were 2,097,152 possible states that settled into 1 of 52 possible attractors (data not shown). Clearly, as the network size increases, it becomes difficult to draw biological inferences from the full attractor space. In addition, for

large-scale networks, the enumeration of all attractors quickly becomes computationally intractable owing to the exponential relationship between the number of nodes in the network and the number of possible network states [129]. One approach for analyzing large attractor spaces is to measure the robustness of each attractor [130-132]. Unfortunately, these methods are generally more useful to a theoretician than a biologist. Another approach for analyzing large state spaces with many attractors is to use asynchronous updating methods [49, 50]. These methods can generate a probability that a given network node will be ***ON*** or ***OFF*** under a particular set of conditions. Because asynchronous methods represent a repeated sampling of many different timescales [103, 123, 133], they are also useful for modeling the heterogeneity of a population of cells [50, 134]. Asynchronous methods can also facilitate *in silico* molecular perturbations (such as knock-downs or constitutive activations) to generate output that is readily comparable to biological data [50].

Finally, not all stable attractors in a deterministic synchronous logic model remain stable under non-deterministic asynchronous update conditions. The random perturbations associated with an asynchronously updated model result in the disappearance of some synchronous attractors when the assumption of timescale uniformity is removed [93, 103, 128]. Thus, the asynchronous method is more likely to identify attractors that are robust to the typical stochastic variations observed in molecular interactions [135, 136] under physiological conditions.

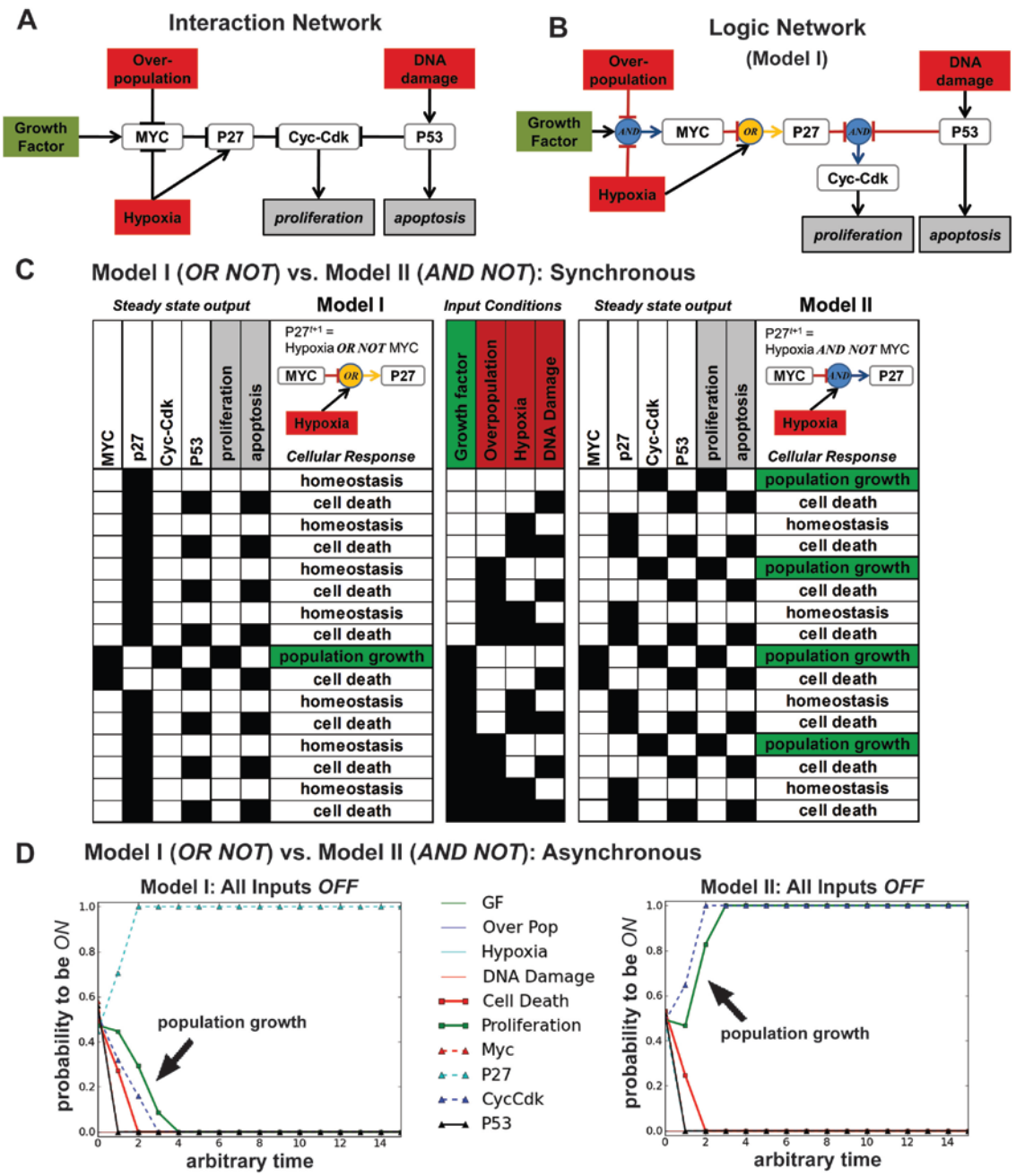


Figure 2.5. A Model of Normal Cell Proliferation and Apoptosis: *OR NOT* vs. *AND NOT*. (**A**) An interaction network diagram depicting regulation of cellular proliferation and apoptosis based on a model published by Ribba et al. [87]. **Overpopulation**, **Growth Factor**, **Hypoxia**, and **DNA Damage** are the input nodes and **proliferation** and **apoptosis** are the output nodes. The four internally regulated nodes are colored white. (**B**) Logic network describing the logic functions used in **Model I** (see also **Table 2.1**). (**C**) The 10 nodes in this network produce a total of 2^{10} (1024) possible states, each of which settles into 1 of 16 possible attractors (identified as steady state output in the table), under a synchronous updating scheme. The 16 possible input conditions are displayed in the center table. On the left, **Model I** output is presented (with **P27** regulated by ***OR NOT***) and, on the right, ***Model II*** output is presented (with **P27** regulated by ***AND NOT***). (**D**) Comparison between asynchronous attractors produced by **Model I** and **Model II** when all input nodes are ***OFF***. This input condition corresponds to the first row in the synchronous scheme's table. An asynchronous random order update was repeated 200 times for 15 time steps. Each curve shows the average value of a node after 200 simulations for each time point. Blue circles indicate ***AND*** interactions and yellow circles indicate ***OR*** interactions.

2.6 Biologically motivated examples

A model of normal cell proliferation and apoptosis

In **Figure 2.5**, we present a 10 node network of the regulation of cellular proliferation and apoptosis. This model is adapted from the Boolean model used by Ribba et al. to control cell division and cell death in a multiscale model of colorectal cancer [87]. For simplicity, our version of the model eliminates linear regulations. In the original Ribba model, for example, P53 activated BAX and BAX, in turn, activated apoptosis. BAX was eliminated in our model so that P53 is a direct activator of apoptosis. In this example, the linear node reduction results in the same qualitative network behavior and output. The interaction network for this model is presented in **Figure 2.5A**, and the logic network is depicted in **Figure 2.5B**. There are 4 input nodes representing the signals a cell may respond to in the model: **Growth Factor**, **Over-population**, **Hypoxia**, and **DNA Damage**. The output nodes are **proliferation** and **apoptosis**. The 4 internally regulated nodes are **MYC**, **P27**, **Cyc-CDK**, and **P53**. The logic functions used in this model are listed in **Table 2.1** along with a biological justification for each function and a statement of any relevant assumptions.

In this model there are 1024 (2^{10}) possible network states and 16 (2^4) possible input conditions (**Figure 2.5C**). For the logic network referred to as Model I (**Figure 2.5B**), under synchronous updating each of the 1024 possible states eventually settle into one of 16 fixed point attractors corresponding to one of three biological states (**Figure 2.5C, left**). In Model I, when all input signals are ***OFF***, **apoptosis** and **proliferation** are also ***OFF*** in the attractor**,** indicating cellular homeostasis. When all four input signals are ***ON***, the combination of these signals leads to **apoptosis *ON*** and cell death (first row in the table). The only input condition leading to **proliferation *ON*** in Model I is **Growth Factor *ON*** and all other inputs ***OFF***.

In another version of this model referred to as Model II, the logical operator controlling the regulation of **P27** by **MYC** inhibition and **Hypoxia** activation is changed to ***AND NOT*** from ***OR NOT*** (**Figure 2.5C**, right). This seemingly small change produces very different output. In Model II, under

synchronous updating when all input signals are ***OFF***, all states settle into an attractor that leads to **proliferation *ON*** and population growth, which is not the behavior expected from a normal cell. Moreover, there are now 4 input conditions that give rise to **proliferation *ON***, including the state where all input signals are ***OFF***. In this network example, the use of ***OR NOT*** for the regulation of **P27,** rather than ***AND NOT*****,** is essential for obtaining the expected readout.

Under asynchronous updating, Model I and II settle into the same attractors each model produced under synchronous updating (**Figure 2.5D**). Empirically, this will be the case when only point attractors are possible under synchronous updating because no regulatory feedback loops are included in the network. For more complex networks with multiple feedback loops, the attractor space produced by synchronous and asynchronous updating methods will not always be identical.

Oscillations in Boolean models

Logic models have been used to model oscillations in a number of biological systems [48, 106, 109]. Oscillations play important roles in many biological processes including the cell cycle, circadian rhythms, developmental processes, and the cellular response to stress [137]. It is easy to generate sustained oscillations (which are, by definition, a cycle attractor) with a synchronous logic model. All that is required is the presence of a negative feedback between one or more nodes. Even a single node can generate oscillations (**Figure 2.6A**) [138]. In contrast, generating sustained oscillations in chemical systems with an ODE model requires at least three distinct species represented by two or more ODEs [139, 140]. Physically realistic sustained oscillations are possible with only two chemical species when delay differential equations are used because these equations introduce an explicit time delay into the system [43, 141]. It must be emphasized that simply generating oscillations in a mathematical model of any type does not imply the underlying mechanism driving the oscillations in the model is equivalent to that driving the experimentally observed oscillations.

Given the ease with which oscillations can be generated with a logic model, it is essential to ensure that the implicit assumptions underlying the logic functions in a model are appropriate for the oscillatory system modeled. The protein P53 has been dubbed the "guardian of the genome" for its role in maintaining genome integrity and tumor suppression in normal cells [142]. P53 plays critical regulatory roles in both cell cycle progression and apoptosis. In response to stress, such as DNA damage from ionizing radiation, P53 protein is known to mediate the transcription of MDM2, which targets P53 for degradation via ubiquitination [45, 137, 143, 144]. In response to DNA damage, this antagonistic relationship induces important cellular oscillations in P53 and MDM2 expression, which have been described as a digital behavior [137]

.

Table 2.1. Justifications for logic functions used in Figure 2.5, Model I.

Logic Rule	Justification	Source(s)
MYC^{t+1} = Growth Factor ***AND NOT*** Overpopulation ***AND NOT*** Hypoxia	Both overpopulation and hypoxia inhibit MYC expression. Hypoxia initiates the TGF-Beta pathway and promotes phosphorylation of SMAD's MH2 domain, which inhibits MYC expression. Overpopulation activates the Wnt signaling pathway, releasing the APC destruction complex and promoting degradation of β -catenin, a transcriptional activator of MYC. In contrast, growth factors activate MYC transcription via RAS signaling**.** Our logic rule assumes that MYC activation (via growth factor induced signaling) is only possible when neither inhibitor (Overpopulation and Hypoxia) is active.	[145-148]
$P27^{t+1}$ = Hypoxia ***OR NOT*** MYC^{t}	Hypoxia activates P27, a potent cell cycle inhibitor, by initiating the TGF-β pathway and promoting phosphorylation of SMAD's MH2 domain, which plays a role in maintaining P27 expression. MYC, when active, inhibits transcription of P27 by binding to its promoter**.** We assume there are endogenous levels of active P27 present in the absence of hypoxia. The ***OR NOT*** logic rule ensures that when present, Hypoxia's activation of P27 dominates MYC's inhibition.	[145, 146, 149]
$Cyc\text{-}Cdk^{t+1}$ = ***NOT*** $P27^{t}$ ***AND NOT*** $P53^{t}$	Cyclin dependent kinases (Cyc-Cdk) are inhibited by both P27 and P21. P21 is activated by P53 in response to DNA damage signals. For simplicity we have excluded P21 from this regulation.	[150, 151]
$P53^{t+1}$ = DNA Damage	P53 is transcriptionally activated in the presence of DNA Damage.	[137]
$Proliferation^{t+1}$ = $Cyc\text{-}Cdk^{t}$	Cyclin dependent kinases (Cyc-Cdk) are required for cell progression in the G1 phase. They act to inhibit the Rb protein by phosphorylation, which is an inhibitor of cell cycle progression and proliferation.	[150, 151]
$Apoptosis^{t+1}$ = $P53^{t}$	P53 transcriptionally activates BAX, which is an upstream activator of apoptosis.	[152]

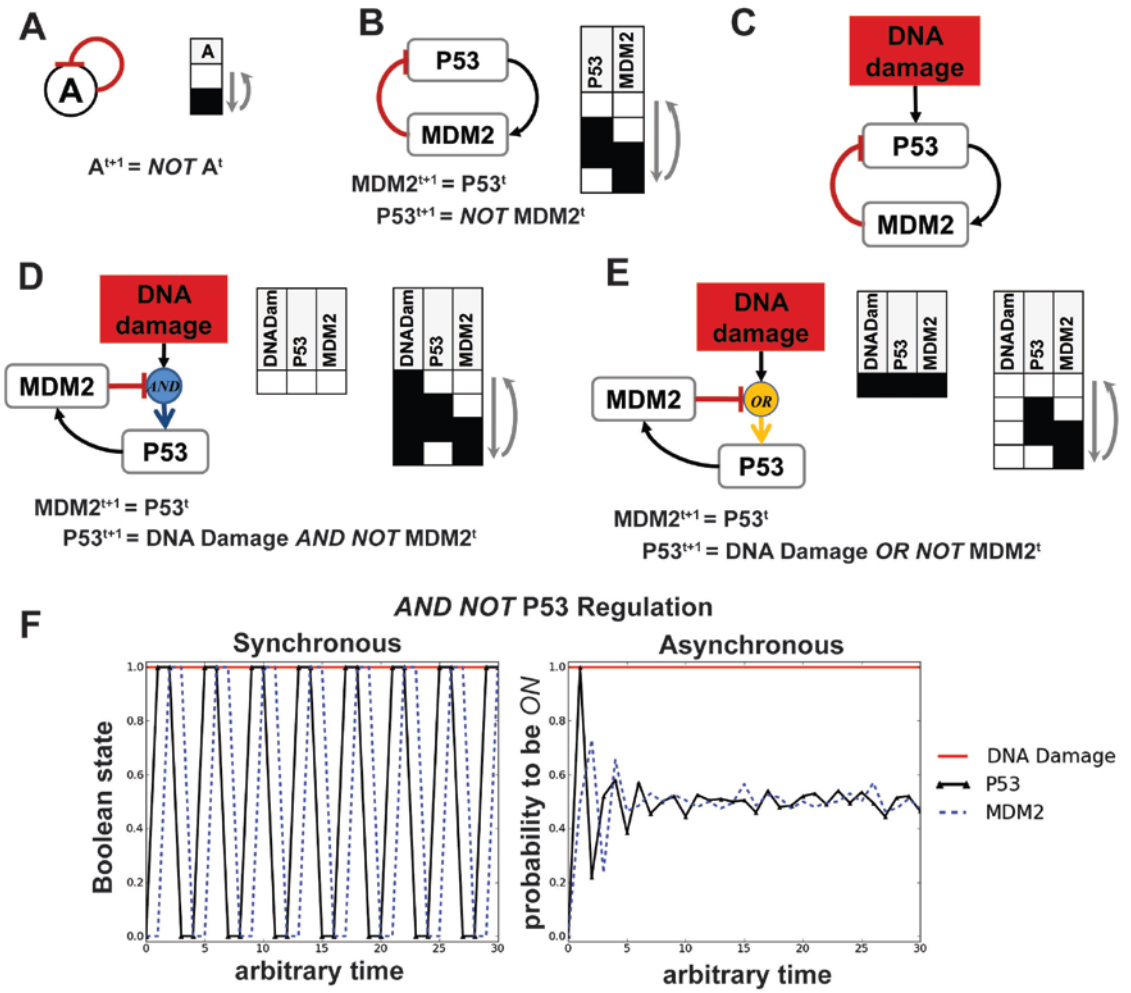


Figure 2.6. Regulatory feedback and oscillations in logic models. (**A**) A single node will produce an oscillating attractor if a negative auto-feedback is present. (**B**) Two nodes will produce an oscillating attractor if a negative feedback is present. In this example **P53** activates **MDM2**, which inhibits **P53**. (**C**) A more realistic representation of the P53/MDM2 circuit includes activation of **P53** by **DNA Damage**. This representation of the network does not identify whether an ***AND NOT*** or an ***OR NOT*** should be used. (**D**) When **P53** is regulated by an ***AND NOT*** function, two synchronous attractors are possible. (**E**) When **P53** is regulated by an ***OR NOT*** function, two synchronous attractors are possible. (**F**) Based on biological data, the **AND NOT** appears to be the correct regulation because the presence of **DNA Damage** induces functional levels of **P53** that give rise to oscillations in the attractor. In the absence of **DNA Damage**, all nodes in the attractor are below their functional activation threshold (***OFF***). In the plots, a synchronous update scheme and an asynchronous update scheme (with 200 simulations) are compared for the logic model presented in (**D**) when **DNA Damage** is ***ON*** and the other 2 nodes are ***OFF*** at the start of the simulation. Blue circles indicate ***AND*** interactions and yellow circles indicate ***OR*** interactions.

If we first consider the simple 2 node network in **Figure 2.6B** consisting of only **P53** and **MDM2** nodes, we can see that, because only one edge leads into each node, the interaction network is equivalent to its representation as a logic network (using the logic network notation described previously). In this simple network, all four possible states make up the limit cycle attractor. Thus, if no

other inputs are added to this network, it will perpetually oscillate between these four states. A slightly more realistic model includes adding a **DNA Damage** node as the input signal (**Figure 2.6C**). Because two nodes now regulate **P53** (an activator and an inhibitor), we must decide whether ***AND NOT*** or ***OR NOT*** (**Figure 2.6E**) is appropriate for the regulation of **P53**. In the ***AND NOT*** case, four of the eight possible network states settle into a fixed point attractor where everything is ***OFF***. The remaining four states settle into a limit cycle where **P53** and **MDM2** oscillate and **DNA Damage** is fixed to ***ON*** (**Figure 2.6D**). In contrast, in the ***OR NOT*** case, four of the eight possible states settle into a point attractor where everything is ***ON*** and the remaining four states settle into a limit cycle where **P53** and **MDM2** oscillate and **DNA Damage** is fixed to ***OFF*** (**Figure 2.6D**). Because there is nothing in the network to regulate **DNA Damage**, it will never oscillate in either version of model.

P53 and MDM2 proteins are known to exist at low endogenous levels in the absence of DNA damage [137]. We, therefore, consider them to be at levels below their functional threshold in the absence of DNA damage. However, when DNA damage is present, P53 becomes functionally activated [137], which triggers functional MDM2 expression levels and oscillations. Thus, we conclude that the ***AND NOT*** function is appropriate for this regulation (**Figure 2.6D**). All model results presented in **Figure 2.6D-E** relied on a synchronous updating scheme. In **Figure 2.6F**, the synchronous and asynchronous updating schemes are compared for the ***AND NOT*** form of the model using the following initial conditions: **P53** ***OFF***, **MDM2** ***OFF***, and **DNA Damage** ***ON***. In the synchronous approach, it can clearly be seen that **DNA Damage** stays constant, while **P53** and **MDM2** oscillate out of phase between 0 and 1 (or ***OFF*** and ***ON***). In the asynchronous model, which represents and average of 200 simulations using a random update order, over time both **P53** and **MDM2** are roughly 50% likely to be ***ON***. This is the expected probability when all 200 simulations settle into a limit cycle attractor where **P53** and **MDM2** oscillate, which is also suggestive of the average signal across a population of cells exposed to DNA damage.

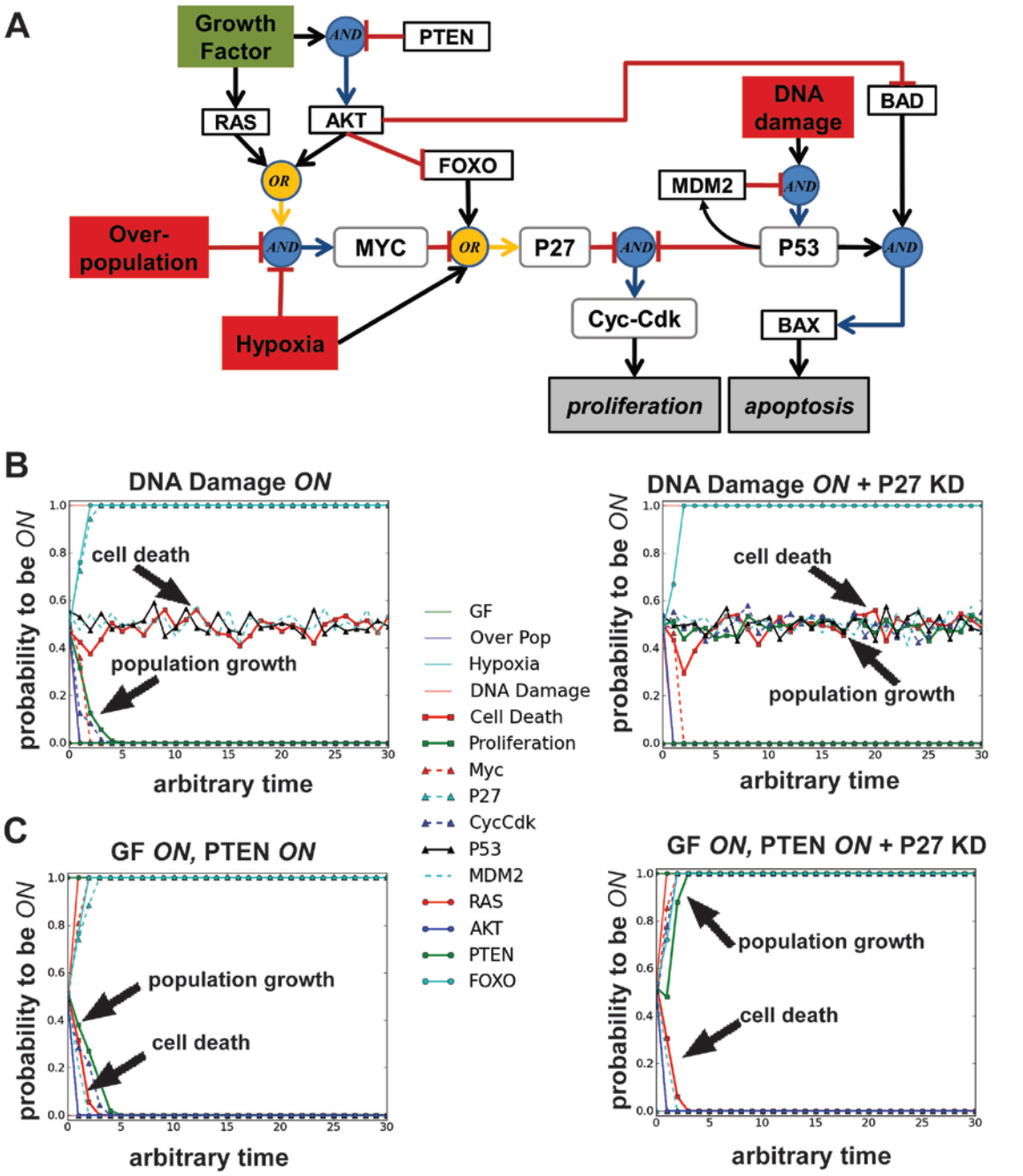


Figure 2.7. An Expanded Model of Normal Cell Proliferation and Apoptosis. (**A**) A logic model diagram of a more detailed version of **Model I** in **Figure 2.5**. Specifically, growth factor activation of MAPK and PI3K signaling pathway is incorporated via RAS and AKT, respectively. In addition, PTEN's downstream inhibition of AKT is included as is AKT's influence on cell cycle progression and apoptosis. (**B**) Control and *in silico* perturbations of this model were performed using asynchronous simulations. Comparison between asynchronous attractors with and without **P27** knock down (KD) for the initial condition: **DNA Damage *ON*** and all other nodes randomized. (**C**) Comparison between asynchronous attractors with and without **P27** KD for the initial condition: **Growth Factor** (**GF**) ***ON***, **PTEN *ON***, and all other nodes randomized. All simulations with **P27** KD began with **P27 *OFF*** and kept it fixed to this state for the duration of the simulation. All asynchronous simulations were performed 200 times to get the probability each node would be ***ON*** under the test conditions. Blue circles indicate ***AND*** interactions and yellow circles indicate ***OR*** interactions.

Expanding the model of normal cell proliferation and apoptosis

In the previous model of cellular proliferation and apoptosis (**Figure 2.5**) only fixed point attractors were possible because no regulatory feedback loops were present in the network model. If the feedback loop in **Figure 6D** is added to the network in **Figure 5,** then the 8 fixed point attractors with **DNA Damage *ON*** become cycle attractors with **P53** and **MDM2** oscillations (data not shown). In addition to adding an **MDM2** node, we expanded Model I (**Figure 2.5**) to include 5 additional nodes (**Figure 2.7A**). An **AKT** node was included because AKT related signaling plays an important role in activating cellular growth and inhibiting apoptosis. In addition, PTEN, a powerful tumor suppressor and upstream inhibitor of AKT activation was included [153]. RAS, FOXO, BAD, and BAX were also added. In **Figure 2.7B**, we performed asynchronous simulations using two distinct conditions: (1) **DNA Damage *ON*** and all other inputs ***OFF*** and (2) **Growth Factor *ON***, **PTEN *ON***, and all other inputs ***OFF***. We also perturbed each of these conditions by preventing **P27** from turning ***ON*** in the simulations, which effectively served as an *in silico* knock down (KD) of this node. In the **DNA Damage *ON*** condition, the result is as expected: all states settle into attractors where there is 0% probability for population growth (**proliferation**) and approximately 50% probability for cell death (**apoptosis**). This intermediate cell death probability, which results from the fact that **P53**, **MDM2**, and **apoptosis** nodes are oscillating in the attractor, is representative of an average between the oscillating states. A very different outcome is found, however, when the **DNA Damage *ON*** condition is tested with the **P27** KD. In this case, **proliferation** also has approximately 50% probability to be ***ON*** (**Figure 2.7B**).

When both ***Growth Factor*** (which should trigger proliferative signals) and **PTEN** (which should suppress proliferative signals) are ***ON***, then cellular homeostasis (or quiescence) occurs because all states end up in an attractor where both population growth (**proliferation**) and cell death (**apoptosis**) have a 0% probability to be ***ON***. When the same conditions are tested with the **P27** KD, however, population growth is 100% likely to be ***ON*** in the attractor (**Figure 2.7C**). While these results are not surprising given the importance of P27 in

regulating cell cycle progression and proliferation, they do demonstrate how models of this type may be used to make predictions related to network dynamics that can then be verified experimentally.

2.7 The use of Boolean models in dynamical systems

We have seen how Boolean models can be used to simulate network dynamics and have also discussed that Boolean models are capable of qualitative agreement with more mechanistically precise ODE models. It must be stressed, however, that a Boolean model will not always be appropriate for modelling network dynamics. In complex biochemical pathways, the time evolution of the concentration of interacting species is governed by nonlinear equations representing feedback loops in which the output of a pathway is not proportional to its input [154]. Examples of network behavior that a two-state Boolean model cannot approximate include retroactive signalling [21, 41], load-induced modulation [155], and bifurcations associated with nonlinear equations [156]. In the case of retroactivity and load-induced modulation, which involve upstream signal propagation in covalently modified signalling cascades and are discussed in **Chapter 4**, a two-state Boolean model is too qualitative to predict this behavior because these signalling processes arise from enzyme sequestration mechanisms that are out of reach of two-state Boolean models. In the case of bifurcations, which occur when the qualitative behavior of the solution of a nonlinear system changes as a parameter changes, two-state Boolean models are unable to predict bifurcations because they lack parameters. It is possible that multistate or continuous piecewise logic models (which are parameter driven) are capable of predicting bifurcations. To our knowledge, however, this has not yet been explored systematically.

In a recent paper by Batchelor et al [157], the mechanisms regulating P53 response to different perturbations were investigated. The authors employed ODE models as part of an analysis into the amplitude, duration, and frequency of individual p53 pulses in response to varying amounts of ultra violet radiation.

While a synchronous Boolean model would not be able to elucidate mechanisms driving the degrees of response in a system like this, multistate fuzzy logic [47] and continuous piecewise logic models [49] have been used for similar purposes in other systems. In addition, asynchronous Boolean models may be used for these types of responses. In the model of guard cell aperture closure previously discussed [50], the authors chose to use an asynchronous Boolean model to generate a probability of stomate closure because stomate aperture responses are known to be graded and cannot be represented as simply open or closed. While the predictive power of an ODE model is preferred for a dose dependent response, an ODE model may not be practical for modelling such responses in large networks for computational and parameter space reasons. In such cases when a binary response is not sufficient, the use of more complex logic-based methods, such as asynchronous, multistate, fuzzy, or piecewise models, may be a reasonable alternative.

2.8 Concluding theoretical remarks

Logic-based models are predictive tools that can be leveraged in the absence of reliable parameter information or mechanistic details needed for more quantitatively precise methods, such as ODE models. Importantly, the predictive power of logic methods is dependent on the nature of the logic network model constructed. In this review, we have pointed out important factors to consider when building predictive logic-based models. We have emphasized the importance of using descriptive logic network diagrams and provided biologically motivated example networks. Most significantly, we have emphasized the need to properly characterize the nature of all interactions in the network and to understand the implicit meaning of logic functions used to integrate multiple input signals.

As we have seen, the use of ***AND*** and ***OR*** logical operators produce very different results for the same input conditions (**Figure 2.4B-D**, **Figure 2.5C-D** and **Figure 2.6D-E**). We strongly encourage the creation of truth tables to verify that the output of each logic function is reasonable and in qualitative agreement with experimental data, if available. When the nature of the interaction modeled

by a logic function is not known (e.g., whether an activator will trump an inhibitor, if both are active, or vice versa), then the logic model can be used to test hypothesized mechanisms for the uncertain interaction. The use of “incomplete truth tables”, a computational approach for analyzing the effect of logical uncertainty in a logic network, has also been proposed for these cases [158].

Despite their advantages, resistance to the use of logic-based models in biology exists. Some resistance is related to the idea that a molecule’s state can be reduced to discrete ***ON*** and ***OFF*** values. In actuality, experimental molecular states are often qualitatively described in binary terms. Genes may be characterized as up-regulated or down-regulated in microarray experiments and proteins are often referred to as activated or inactivated to indicate their functional state. Given the stochastic variation in gene and protein expression across cells, biological molecular networks are remarkably robust [159]. The presence of growth factors in the local environment, for example, will almost invariably result in the induction of proliferative pathways within a population of cells, despite the heterogeneity in the molecular expression across individual cells in the population. This deterministic output from a given cellular input has been compared to cellular digital computation [160]. Fundamentally, the basis of digital readouts are 0’s and 1’s – at least at the computational level. Another point of concern with logic-based Boolean models is that time is unrelated to physiological time and can provide only a qualitative chronology of molecular activations [93]. While this is true, Boolean models can provide qualitative predictive values, which allow biomedical scientists to gain unique insights in molecular network dynamics that may otherwise be out of reach.

For those interested in using logic models to study large networks, the use of asynchronous updating is generally recommended [92, 128]. A variety of algorithms exist for introducing asynchronous updates in a logic model [128]. For most purposes, the repeated random order asynchronous [128] update method (which is similar to a statistical Monte Carlo simulation) will be sufficient. This is the algorithm used for the asynchronous simulations in this review. Some attractors found with the simpler synchronous updating scheme may be artifacts

of uniform timescales. In contrast, an asynchronous scheme introduces stochastic variation in timescales. Moreover, asynchronous methods can produce qualitative readouts that are more representative of biological readouts (**Figure 2.5D** and **Figure 2.7B-C**) and can easily facilitate *in silico* perturbations, such as knock downs and constitutive activations.

We view logic models as complementary to other network analysis methods in systems biology and consider them to be an important tool for making biological inferences about the dynamics of intracellular networks. A number of software tools for logic-based network analysis are available [49, 126, 158, 161]. The appropriate software tool to use will depend on the nature of the network model and objectives of the analysis [97]. For asynchronous simulations and *in silico* molecular perturbation studies, we recommend BooleanNet [49] which is a relatively easy to use open source tool developed in Python. Needless to say, the implementation of logic-based models requires computational and mathematical proficiency. As a consequence, collaboration between integrative biologists and computational scientists will play a pivotal role in the successful development, analysis, and interpretation of logic-based models.

Importantly, logic-based models are also a powerful approach for constructing models of biological networks that can ultimately be integrated into multiscale models — models that consider the integration between different scales and phenomena in a biological system or process — to provide an integrative view of biological systems [1]. In the literature, multiscale models of cancer growth have been developed that account for the cellular, genetic, and environmental factors regulating tumor growth [87, 162]. These models have implemented genetic and signaling networks as Boolean models to regulate cell cycle progression where the response to signals from the intracellular gene network determines whether a cell will proliferate or die and, therefore, directly influences the cellular and the extracellular tissue level of the model.

In conclusion, it is never feasible to create a model that is an exact replica of a complex system and, as a consequence, compromises must be made between the predictive power of a model and the complexity of the model. The

discrete nature of a Boolean model sacrifices quantitative dynamics for qualitative dynamics. In exchange, a parameter-free modeling framework can be used to investigate complex intracellular networks.

Chapter 3
A Discrete Logic Model of the PI3K Signaling Pathway

Portions of the discussion section of this chapter were previously published by the author in *Advances in Experimental Medicine and Biology* [1].

3.1 Background

The phosphatidylinositol 3-kinase (PI3K) pathway is highly dysregulated in human cancers and plays important regulatory roles in glucose metabolism [163]. AKT, a protein kinase that plays a central role in the pathway, is a master regulator of cellular growth and survival. AKT (also known as PKB) promotes survival by suppressing apoptosis via inhibition of BAD [164], by stimulating glucose metabolism in several direct and indirect ways, and by controlling cell cycle progression via MYC regulation [165]. AKT's regulation of glucose metabolism includes the direct phosphorylation of phosphofructokinase-2 (PFK2), an enzyme that promotes fructose-2,6-bisphosphate's allosteric regulation of the glycolytic enzyme phosphofructokinase-1 (PFK1) [166]. ATP citrate lyase is both a direct target [167] and downstream target of AKT [168]. In addition, AKT promotes the expression of glucose transporter proteins (GLUT), lactate dehydrogenase, and hexokinase [153] (**Figure 3.1**).

PI3K is a lipid phosphatase consisting of regulatory p85 and catalytic p110 subunits when fully active. PI3K acts as a key upstream regulator of AKT because of its tight regulation of the activity of mammalian target of rapamycin complex 2 (mTORc2) and phosphatidylinositol (3,4,5)-triphosphate (PIP3), which phosphorylate AKT at ser473 and thr308, respectively. PI3K is required for mTORc2 activation via ribosomal binding [169, 170] and PI3K activates PIP3, which recruits AKT to the cellular membrane where it is phosphorylated at thr308

by PDPK1 [163]. In addition to activation by PI3K, PIP3 is strongly inhibited by the tumor suppressor phosphatase and tensin homolog (PTEN) [153, 171, 172].

Given the importance of the PI3K pathway in promoting cell survival and inhibiting cell death, the PI3K pathway is viewed as an extremely promising target for anti-cancer interventions [18-20], and a number of clinical trials featuring PI3K inhibitors are underway [173, 174]. Despite promising pre-clinical data, off-target effects and the induction of drug resistance remain common problems associated with targeted molecular inhibitors [13, 14]. As discussed in **Chapter 1**, it is increasingly likely that the high-incidence of off-target effects and the emergence of therapeutic resistance from molecular inhibitors are related to the complexity of intracellular molecular networks. New approaches for studying complex intracellular networks and generating testable predictions related to off-target effects are needed. In order to build models of this type, the underlying networks they are based upon must be validated with experimental data.

In this chapter, I present a logic network model that links growth factor signaling to the PI3K pathway and glucose metabolism. The objective of the model is to elucidate the non-linear regulation between PI3K/AKT signaling and glucose metabolism. A logic network representing a canonical view of normal PI3K signaling was built. This model was validated by comparing experimental data from the MCF10A cell line to model output under six experimental conditions, including treatment with pathway specific inhibitors and stimulation with epidermal growth factor (EGF). To help understand the nature of PI3K pathway dysregulation as breast cancer progresses, *in silico* oncogenic perturbations were made to the normal network model and the generated output was compared to experimental data from the three cancer cell lines under the same six experimental conditions.

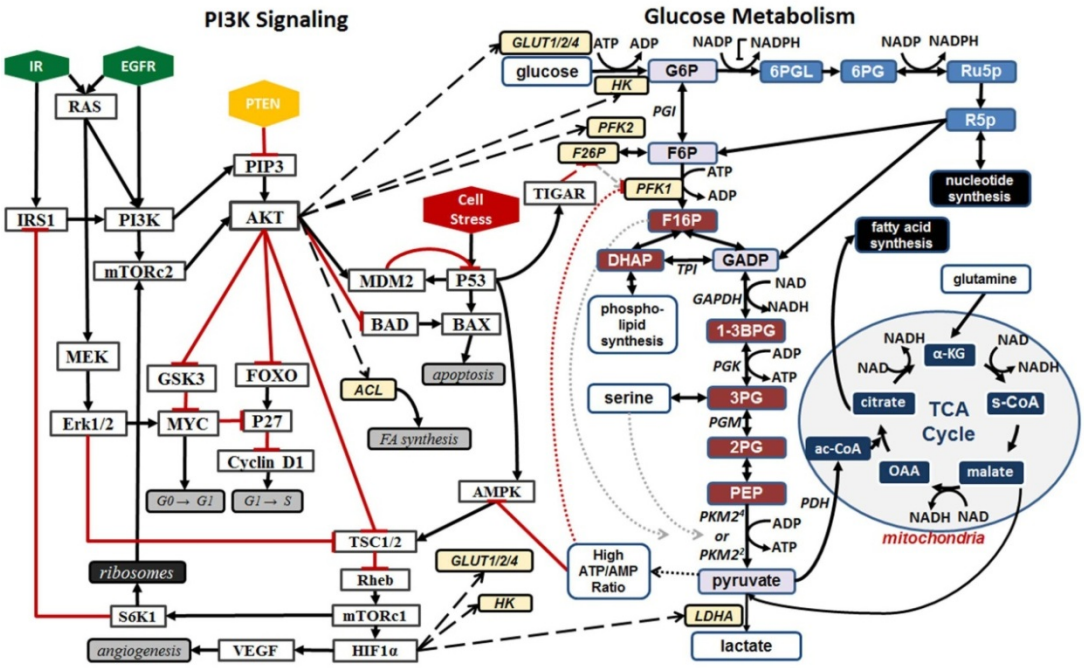


Figure 3.1. Links between the PI3K/AKT pathway and glucose metabolism. Black arrows in the signal transduction interaction network represent activation and red lines with blunt edges represent inhibitions. Black arrows in glucose metabolism represent enzyme catalyzed conversions from substrate to product. Bold dashed lines represent feedback to metabolism originating from the PI3K pathway. Thin dashed lines represent feedback originating from glycolysis.

3.2 Methods

Western blot analysis

Cells were serum starved overnight then incubated for 1 h with or without 50 µM of LY294002 inhibitor (Cell Signaling) or 25 µM of U0126 inhibitor (Cell Signaling). After 1 h, cells were treated with or without 10 ng/mL EGF stimulation for 10 minutes. Proteins were immediately harvested from cell cultures. Antibodies against phospho-AKT (ser473), phospho-AKT (thr308), AKT (pan), phospho-GSK3β, GSK3β, phospho-PTEN (Ser380), and PTEN were obtained from Cell Signaling.

Logic network development and analysis

A detailed literature search was performed to identify the important regulatory components of the PI3K pathway as well as their interacting partners. **Table 3.1** summarizes some key interactions identified, which represent edges in an interaction network (**Figure 3.1**). The nature of each interaction was carefully considered before deriving a set of logic rules. In **Chapter 2**, an overview of the

general methodology used to curate the network as well as an introduction to the theory of logic based models is presented. A random order asynchronous update scheme, which is similar to a Monte Carlo simulation, was used for all simulations (see **Chapter 2**). All nodes not explicitly assigned a value at the start of a simulation were assigned random values in each of the 100 network simulations performed for any set of initial conditions. Each simulation of 100 networks ran for a pre-defined number of time steps (long enough for a logical steady state, also known as an attractor [93], to be reached). The output of a set of asynchronous simulations was the probability that a node stabilized to ***ON*** in the attractor for a given set of initial conditions. The control condition was simulated by fixing the **Growth Factor** to ***OFF***. EGF stimulation was simulated by fixing ***Growth Factor*** to ***ON***. Inhibition by LY was simulated by fixing **PI3K** to ***OFF*** and inhibition of U0 was simulated by fixing **MEK** to ***OFF***.

Model validation and cost function penalties

To compare model readout with experimental data, it was necessary to first discretize the experimental western blot data. Values of ***ON***, ***OFF***, or ***partial*** were assigned to indicate the degree of AKT thr308, AKT ser473, ERK, GSK3β, and PTEN activation. A qualitative scoring system was devised to penalize differences between *in silico* read outs and *in vitro* readouts of AKT and ERK. If the *in silico* experimental readout qualitatively matched the *in vitro* readout for a given treatment condition, then no penalty was imposed. Otherwise, a cost penalty function was applied. It is important to note that logic modeling represents a qualitative methodology and attempts at precise quantitation can be misleading. As a consequence, a simulated response of 100% likely to be ***ON*** for an expected ***partial*** experimental response was treated as more correct than a simulated response of 100% likely to be ***OFF*** for an expected ***partial*** experimental response.

Discretized western blot data were assigned numeric values of 1.0 for ***ON***, 0.0 for ***OFF***, and 0.5 for ***partial*** and are referred to as expected values (EV). Simulated values (SV) refer to the probability that a node will be ***ON*** in the logic model attractor. Briefly, if the EV was 0.0 or 1.0 (i.e., 100% likely to ***OFF*** or ***ON***,

respectively), then the cost penalty imposed was the absolute difference between EV and SV. If the EV was 0.5 (i.e., ***partial***), then SVs between .2 and .8 were not penalized because they were considered to be qualitative partial responses. If the EV was 0.5 and the SV < 0.2, then a penalty of 0.5 was imposed. If the EV was 0.5 and the SV > 0.8, however, a smaller penalty of 0.25 was imposed.

3.3 Results

A detailed literature survey of molecular interactions associated with the PI3K pathway was completed (**Table 3.1**) from which an interaction network was constructed that summarized the regulatory links between glucose metabolism and PI3K signaling (**Figure 3.1**). The information provided in **Table 3.1** and **Figure 3.1** is not sufficient to infer logic functions, however. As discussed in **Chapter 2,** it is essential to carefully derive each logic function before constructing a logic network. After further analysis, logic functions were derived and a first iteration logic network was constructed to represent a canonical view of normal PI3K signaling (**Figure 3.2**). Generally, when constructing a logic network, the objective is to build the simplest network possible that still agrees with experimental data [50, 91].

The logic network model was validated by comparing *in silico* and *in vitro* readouts of AKT and ERK phosphorylation in the presence or absence of pathway inhibitors. Western blots were performed on four cell lines used in an *in vitro* model of breast cancer progression (**Figure 3.3**). For *in vitro* analysis, cells were treated with pharmacological inhibitor U0126 (U0), a direct inhibitor of MEK [175], or LY294002 (LY), a direct inhibitor of PI3K [176]. Six conditions (5 treatments and 1 control) were performed *in vitro* (**Figure 3.3**) and *in silico*: (1) control (no LY, no U0, and no EGF) (2) EGF alone, (3) LY alone, (4) EGF + LY, (5) U0 alone, and (6) EGF + U0. For *in silico* analysis, control conditions were simulated by fixing the **Growth Factor** input node to ***OFF***. The presence of EGF was simulated by fixing ***Growth Factor*** to ***ON***. The presence of LY was simulated by fixing **PI3K** to ***OFF*** for the duration of the simulation. Likewise, the presence of U0 was simulated by fixing **MEK** to ***OFF*** for the duration of the simulation.

Table 3.1. Table of key interactions identified in the PI3K signaling pathway.

Affector	Interaction	Target	Source
AKT	(-) phosphorylation	TSC-1/2	[153, 171, 177, 178]
AKT	(+) phosphorylation	MDM2	[45, 144, 179]
AKT	(-) via ATP/ADP levels	AMPK	[153, 166]
AKT	(+) via cytochrome c levels and apoptotic signals	HK	[166, 180]
AKT	(+) phosphorylation	PFK-2	[166]
AKT	(-) serine phosphorylation, 14-3-3 binding	Raf	[181, 182]
AMPK	(+) direct phosphorylation	TSC-1/2	[153, 171, 177, 183, 184]
Cell Stress	(+) mRNA expression	p53	[45, 144, 185]
cMyc	(+) mRNA expression	LDH-A	[186, 187]
EGFR	(+) recruits to membrane and phosphorylation	GAB1	[188-190]
EGFR	(+) binding and phosphorylation	SHC	[44, 188, 191]
EGFR	(+) direct binding	GRB2/SOS	[44]
ERK	(-) phosphorylation	TSC-1/2	[177, 192, 193]
ERK	(-) via. serine phosphorylation of GAB1	Pi3K	[194]
ERK	(-) phosphorylation	SOS	[195-197]
F-1,6-BP	(+) via FBPase-1	F-6-P	[198]
F-2,6-BP	(-) allosteric inhibitor	FBPase-1	[166, 185, 198]
F-2,6-BP	(+) potent allosteric activator	PFK-1	[185, 198]
F-2,6-BP	(+) via FBPase-2	F-6-P	[185]
F-6-P	(+) via PFK-1	F-1,6-BP	[198]
F-6-P	(+) via PFK-2	F-2,6-BP	[166]
GAB	(+) amplifier	Ras-GTP	[199]
GAB1	(+) membrane relocation	GAP	[188]
GAB1	(+) phosphorylation (p85)	PI3K	[189]
GAB1	(+) recruitment and phosphorylation	SHP2	[188]
GAP	(-) phosphatase for GTP	Ras-GTP	[177, 200]
GRB2/SOS	(+) catalyzes the GDP->GTP reaction	Ras-GTP	[196, 201]
GrowthF1	(+) direct	PI3K	[45, 144]
IR	(+) phosphorylation of Tyr895	IRS1	[202]
IR	(+) read another paper to find out…	GAB1	[203]
IR	(+) phosphorylation	Shc	[204]
iRS1	(+) physically associated with Pi3K	Pi3K	[171]
IRS1	(+) binding via SH2 domain of GRB2	GRB2/SOS	[44, 201, 202, 205]
IRS1	(+) phosphorylates (tyrosine)	Shc	[201]
IRS1	(+) binding via. SH2 domain of SHP2	SHP2	[205]
MDM2	(-) delayed degradation via ubiquinations	p53	[45, 144]
MEK	(+) phosphorylation	ERK	[191]
mTORc1	(-) via S6K, inverse inhibitor	iRS1	[153, 171, 177, 180]
mTORc1	(-) inhibits assembly of other mTOR	mTORc2	[153]
mTORc1	(+) via HIF1-a	LDHA	[166, 187]

Affector	Interaction	Target	Source
mTORc2	(+) phosphorylation (Ser473)	AKT	[153, 171, 206, 207]
p53	(+) mRNA expression	AMPK	[177]
p53	(+) mRNA expression	MDM2	[45, 144]
p53	(-) via PTEN, which dephosphorylates	PIP3	[45, 144]
p53	(+) mRNA expression	TIGAR	[5, 185]
p53	(-) mRNA expression	PGM	[5]
PDK1	(+) phosphorylation (Thr308)	AKT	[192]
Pi3K	(+) via ribosomes	mTORc2	[153, 170, 171, 177, 206]
Pi3K	(+) induces phosphorylation	PIP3	[45, 144, 200]
PIP3	(+) direct	AKT	[45, 144, 153, 200]
PIP3	(+) binding of PH domain	GAB1	[44, 189, 190]
PIP3	(+) recruits to membrane	PDK1	[192]
Raf	(+) phosphorylation	MEK	[191]
Ras-GTP	(+) via Raf, MEK, MAPK pathway	ERK	[177, 200]
Ras-GTP	(+) physically associated with Pi3K	PI3K	[180, 191, 200]
Ras-GTP	(+) membrane recruitment and phosphorylation	Raf	[191]
SHC	(+) creates docking site for SH2 domain of Grb2	GRB2/SOS	[191]
SHP2	(-) dephosphorylation	GAB1/PI3K	[188, 189, 208]
SHP2	(+) dephosphorylation	GAB1/GAP	[188]
SHP2	(-) via. IRS1	PI3K	[203]
TIGAR	(-) phosphatase activity	F-2,6-BP	[5, 185, 186]
TSC-1/2	(-) inverse inhibitor	mTORc1	[153, 171, 177, 180]

Table 3.2. PI3K pathway related mutations identified in three breast cancer cell lines. Genes with mutations are indicated by "Y".

Gene	MCF7	MDA-MB-231	SUM-149	Source
BRAF		Y		[209]
P53		Y	Y	[210-212]
PIK3CA	Y	Y	Y	[209, 213, 214]
PTEN			Y	[209, 215]
RAS		Y	Y	[209]

3.3.1 Western blot analysis of each cell line

Based on phospho-western blot analysis (**Figure 3.3**), the activation state of AKT, ERK, PTEN, and GSK3β (a downstream target of AKT) was assessed in each cell line. The activation state of each protein was assigned one of three possible values based on the western blot analysis: ***ON***, ***OFF***, or ***partial*** activation. The status of PTEN was probed because of its role as a potent tumor suppressor that modulates AKT signaling by directly inhibiting PIP3, which mediates the phosphorylation of AKT at thr308 [153]. In **Figure 3.3C**, SUM-149 cells clearly lack PTEN protein expression, which is in agreement with other studies [209, 215] as well as our own qPCR analysis (data not shown). The remaining three cell lines, however, express phosphorylated PTEN under all six experimental conditions, and none of these cells exhibit phosphorylation of AKT on the thr308 residue (**Figure 3.3A-B,D)**. In contrast, SUM-149 cells exhibit some phosphorylation on this residue under all experimental conditions except when EGF and LY are applied together (**Figure 3.3C**).

The status of GSK3β was probed because GSK3β inhibition has been linked to AKT activation [216]. Our data suggest that GSK3β is a robust indicator of EGF stimulation and not necessarily AKT activation, however (**Figure 3.3**). In our experiments GSK3β was always ***ON*** in conditions where EGF stimulation occurred, regardless of AKT status or the presence of a pathway inhibitor.

3.3.2 PI3K pathway mutations in each cell line

To help identify differences in the PI3K pathway across the four cell lines, the existence of mutations in genes involved in PI3K signaling was investigated in the four cell lines (**Table 3.2**). In addition to lacking PTEN gene expression [209, 215], SUM-149 cells have a known P53 mutation (M237I) that appears to disrupt normal conformation and induces loss of functional activity [211, 212]. SUM-149 cells also have a known synonymous single nucleotide polymorphism (SNP) mutation (81T>C) in HRAS, an isoform of RAS [209]. While synonymous mutations are generally not thought to induce disease, this particular mutation was identified as an indicator of bladder cancer risk in patient samples [217]. In addition, there is some evidence in the literature that mutations which do not

change the amino acid sequence are still capable of inducing aberrant behavior [218, 219]. For this reason, the SNP mutation was included in **Table 3.2** as a potential source of differential signaling across the cell lines.

Both MDA-MB-231 and SUM-149 cells have the same SNP (1173A>G; (I391M) in PIK3CA [209], which encodes the p110 catalytic subunit of PI3K. No data has been reported to suggest this mutation is activating or disease inducing in breast cancer, however [209, 213, 214]. MDA-MB-231 cells also have a known mutation in BRAF (1391G>T) and KRAS (38G>A), which are considered oncogenic mutations [209]. MDA-MB-231 cells have a mutation in P53 (R280K) that severely effects P53 localization and promotes its rapid degradation by MDM2 [210, 211]. Finally, the only PI3K pathway related mutation identified in non-metastatic MCF-7 cells is a mutation in PIK3CA (E545K), which is considered an oncogenic mutation [209].

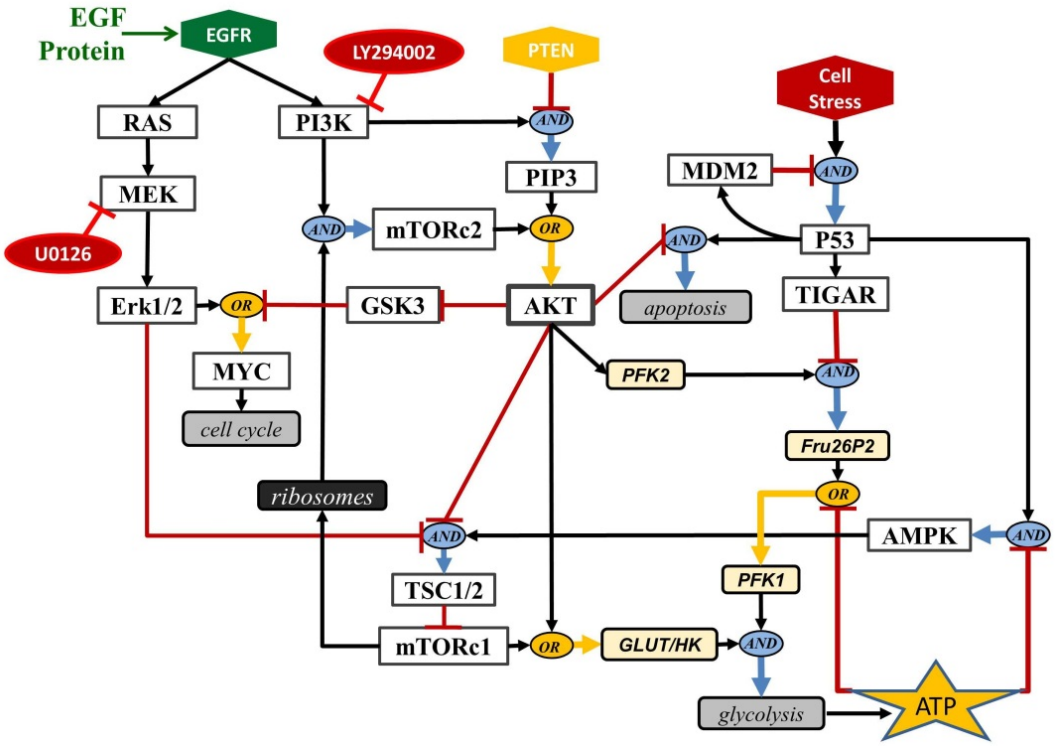


Figure 3.2. Logic network of PI3K signaling. Yellow circles indicate ***OR*** logic functions and blue circles indicate ***AND*** logic functions. LY294002 inhibition of **PI3K**, U0126 inhibition of **MEK**, and EGF stimulation of **EGFR** are indicated.

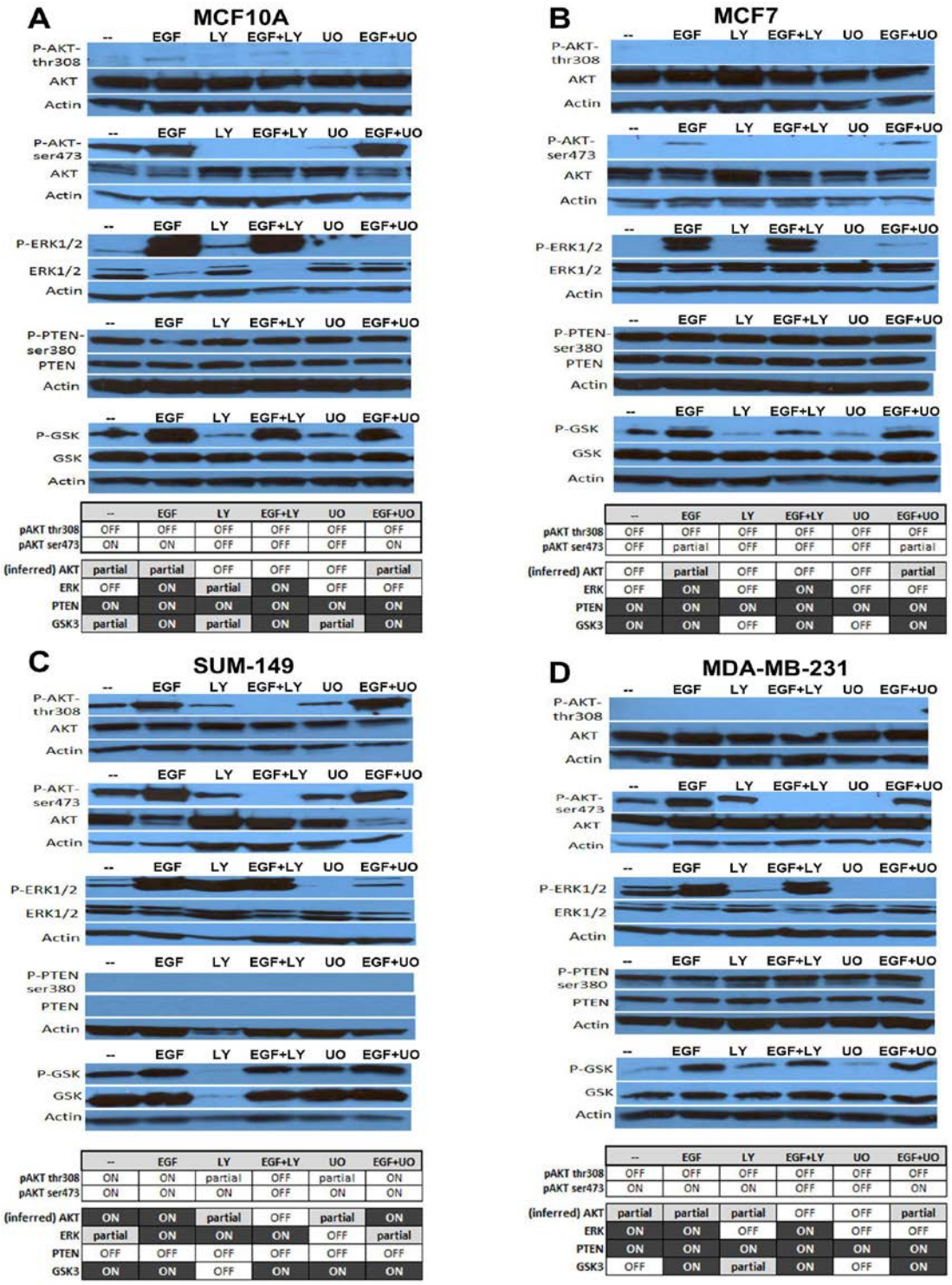


Figure 3.3. Western blot analysis. Results of western blot analysis of AKT, ERK1/2, PTEN, and GSK3β under different experimental conditions are presented for (**A**) MCF10A, (**B**) MCF7, (**C**) SUM-149, and (**D**) MDA-MB-231 cells. All cells were grown in serum free media overnight then treated with or without U0126 (U0) or LY294002 (LY) for 1 h, followed by 10 m with or without EGF stimulation under the following conditions (1) control (--); (2) EGF alone; (3) LY alone; (4) EGF + LY; (5) U0 alone; (6) EGF + U0. The data was then discretized to one of three states (***ON***, ***partial***, or ***OFF***) which is summarized in table form for each cell line. For AKT, which is activated via phosphorylation at two distinct residues, the discretized state of AKT was inferred based on the phospho-western blot of both residues.

3.3.3 PI3K logic model

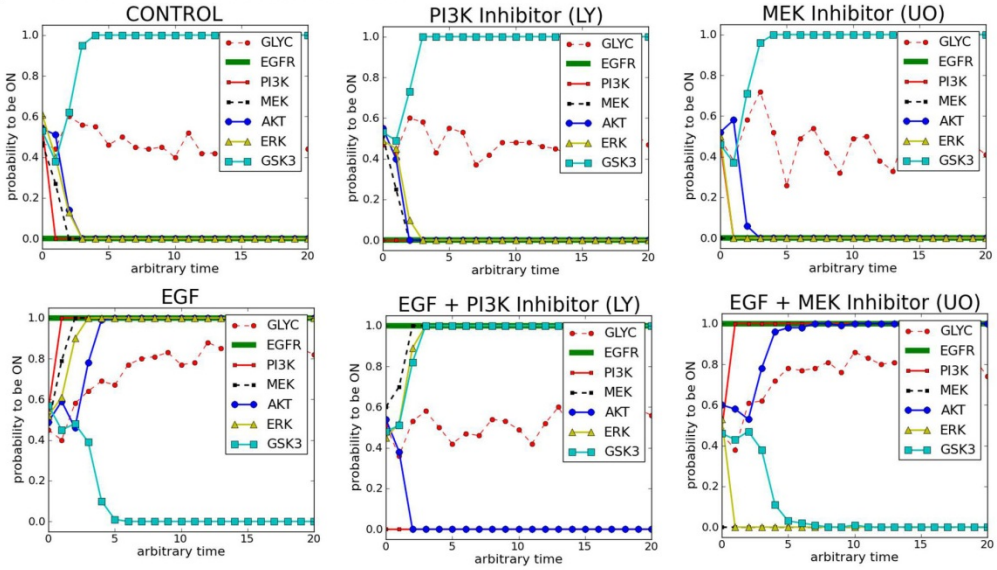
For logic model analysis, the activation state of AKT and ERK under each experimental condition was compared to the *in silico* probability that **AKT** and **ERK** were ***ON*** in the logic attractor after 100 simulations using a random update order (refer to **Chapter 2** for a detailed explanation of this approach) for each experimental condition. AKT and ERK were chosen as the primary *in silico* and *in vitro* readouts for three main reasons. First, AKT activation is tightly controlled by PI3K and ERK activation is tightly controlled by MEK activation. In normal cells, inhibiting PI3K with LY is expected to inhibit AKT, and inhibiting MEK with U0 is expected to inhibit ERK. Second, AKT and ERK signaling are known to be important for promoting proliferation and survival. Finally, AKT and ERK are frequently used as readouts of PI3K and mitogen-activated protein kinase (MAPK) pathway activation in the literature, which allows for comparison of our results to other studies in the literature [194, 220, 221].

After a round of simulations was completed, the simulated activation states of AKT and ERK were compared to their experimentally determined states for each treatment condition. A cost function was applied that evaluated how the simulated results compared to experimental results under each of the six treatment conditions. The cost function assessed a penalty score based on how different the simulated AKT and ERK values were to the expected experimental values (see **Methods** for more details).

To test the PI3K pathway logic model in **Figure 3.2**, simulations were run to mimic a normal (non-cancerous) network with PTEN ***ON*** under all six experimental conditions. The states of all other nodes were randomized at the start of each simulation but were not otherwise directly perturbed. This effectively meant that all other nodes had a 50% probability to be ***ON*** at the start of each of the 100 simulations run under these conditions. The MCF10A normal-like cell line compared reasonably well to this network (**Figure 3.4C**) generating a relatively low cost penalty of 1.5. The non-metastatic cancer cell line, MCF7, also performed well compared to this “normal-like” model, generating a penalty of only 0.5. Given that only one oncogenic mutation involved in PI3K signaling was

found for this cell line in the literature (**Table 3.2**), this result is not surprising. Compared to the normal-like network conditions, however, the metastatic cell lines fared much worse, with SUM-149 and MDA-MB-231 cells generating cost penalties of 4.0 and 2.5, respectively (**Figure 3.3C**).

A **PI3K network simulations with PTEN active**



B ***in silico* state (probability to be *ON*)**

	--	EGF	LY	EGF+LY	UO	EGF+UO
AKT	0.00	**1.00**	0.00	0.00	0.00	**1.00**
ERK	0.00	**1.00**	0.00	**1.00**	0.00	0.00

C **Cost penalties for difference between simulation and experiment**

MCF10A

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	0.50	0.25	0.00	0.00	0.00	0.25	*1.00*
ERK	0.00	0.00	0.50	0.00	0.00	0.00	*0.50*

1.5

MCF7

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	0.00	0.25	0.00	0.00	0.00	0.25	*0.50*
ERK	0.00	0.00	0.00	0.00	0.00	0.00	*0.00*

0.5

SUM149

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	1.00	0.00	0.50	0.00	0.50	0.00	*2.00*
ERK	0.50	0.00	1.00	0.00	0.00	0.50	*2.00*

4.0

MDA-MB-231

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	0.50	0.25	0.50	0.00	0.00	0.25	*1.50*
ERK	1.00	0.00	0.00	0.00	0.00	0.00	*1.00*

2.5

Figure 3.4. Results of PI3K logic model simulations under normal-like conditions. The six experimental conditions were simulated in a normal-like network model where PTEN was set to ***ON***. For each experimental condition, the network was perturbed to match the *in vitro* experimental conditions as described in the **Methods**. (**A**) The simulated results of each condition via asynchronous update are presented along with (**B**) a summary of the simulated results for **AKT** and **ERK** under each condition. (**C**) A summary of the difference in *in silico* simulated results and the experimental results for each condition are presented as cost penalties. A penalty of 0.00 indicates that the simulation and experimental results were qualitatively identical.

I next perturbed the network with two different oncogenic *in silico* mutation states: (1) PTEN ***OFF*** and (2) PTEN ***ON*** and P53 ***OFF***. Surprisingly, the simulated results for both of these network perturbations were qualitatively the same as that of the normal PTEN ***ON*** state presented in **Figure 3.4** (data not shown). Suppressing the activity of PTEN, which is a potent tumor suppressor and upstream regulator of AKT, is expected to significantly affect the activity of AKT as well as ERK. As a consequence, the observation that both oncogenic perturbations produced results that were qualitatively the same as the results produced by the normal-like network for all six treatment conditions indicates that the logic network in the current form (**Figure 3.2**) is highly robust to molecular perturbations.

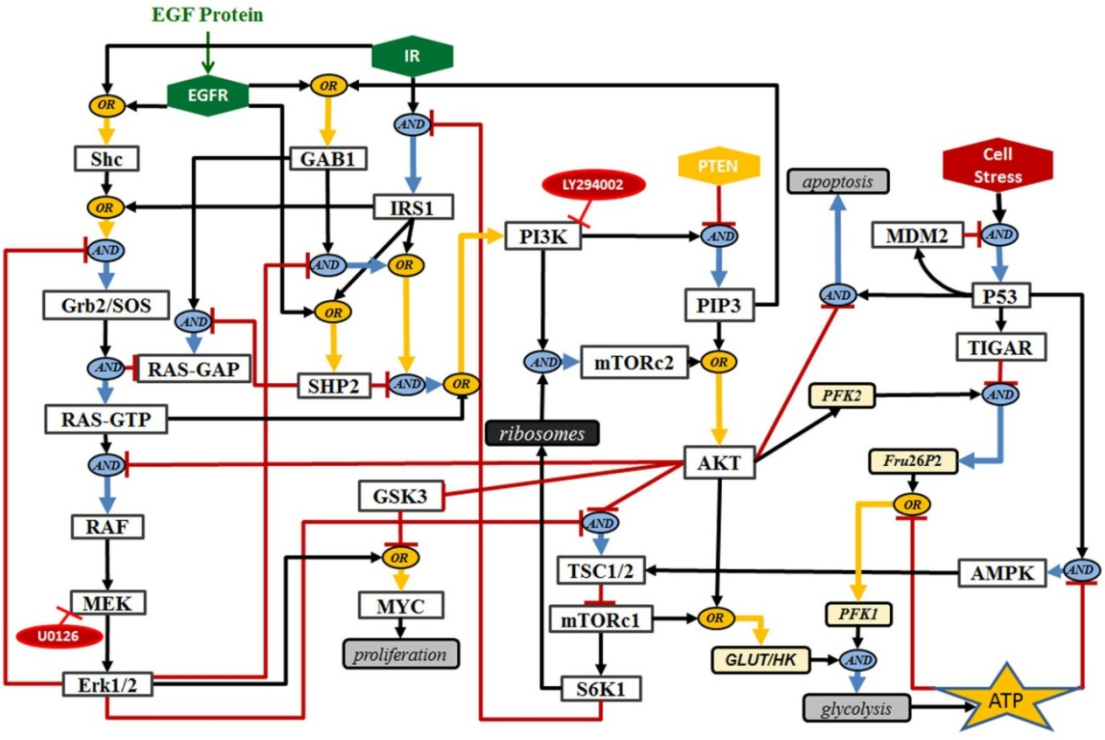


Figure 3.5. Logic network of PI3K/MAPK cross-talk. Yellow circles indicate ***OR*** logic functions and blue circles indicate ***AND*** logic functions. LY294002 inhibition of **PI3K**, U0126 inhibition of **MEK**, and EGF stimulation of **EGFR** are indicated.

3.3.4 PI3K and MAPK cross-talk logic model

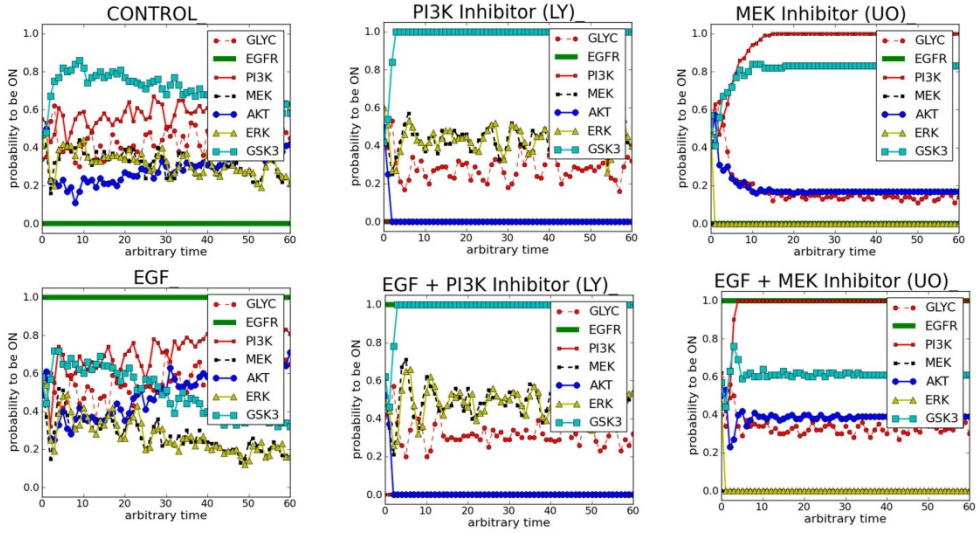
Based on the initial simulations results, it was hypothesized that some regulatory feedback connections between the PI3K and MAPK pathways that were not included in the first iteration of the model may be essential for the regulation of both AKT and ERK. The cross-talk between MAPK and PI3K signaling is complex and a number of regulatory feedback connections between the MAPK and PI3K pathways have been proposed to explain experimental evidence of cross-talk between these two pathways. The most significant examples of cross-talk between these pathways cited in the literature involve the activation of AKT after MEK or ERK inhibition [222-224] and the activation of ERK after AKT knockdown [220] or inhibition [225]. After careful review of the literature, more direct regulatory feedback related to cross-talk with MAPK was included in an expanded logic network model (**Figure 3.5**). This network is referred to as the PI3K/MAPK cross-talk model. The full set of logic rules used for this model, along with a justification of each rule, is summarized in **Table 3.3**.

First, the normal-like network condition with PTEN ***ON*** and all other nodes randomized at the start of the simulation was tested with the expanded PI3K/MAPK cross-talk model. The MCF10A normal-like cell line again compared reasonably well to this network (**Figure 3.6C**), generating a relatively low cost penalty of 1.7, while the cancer cell lines did not compare as well. The time scale of this network model is also observed to be much different than the previous network model. In the previous iteration of the logic network model, all attractors were reached in approximately 5 arbitrary time steps in all treatment conditions (**Figure 3.4A**). In the expanded PI3K/MAPK cross-talk network model, however, the arbitrary time to reach an attractor took much longer when no inhibitor was used in the network (**Figure 3.6A,D**). When either the presence of LY or U0 were simulated, the network reached a steady state attractor in approximately 5 – 10 time steps. In the two conditions without an inhibitor (Control and EGF), however, the time for AKT and ERK to reach a stable value was around 200 time steps, which is referred to as the late response in **Figure 3.6D**. The fact that the larger network would take significantly longer to reach a

steady state under some conditions is not surprising given the addition of several non-linear feedback loops to the network. The observation that the inhibition of PI3K or MEK rapidly induced a steady state attractor (see all conditions other than Control and EGF in **Figure 3.6A**), suggests the extreme importance of PI3K and MEK to the non-linear cross-talk between these two pathways.

Next, the network was tested with the malignant condition of PTEN ***OFF*** and RAS constitutively ***ON*** (**Figure 3.7**). Now (unlike in the first iteration of the model), very different results were obtained compared to the normal-like network conditions. As expected, the PTEN null SUM-149 cells compared best to this network condition. A series of additional oncogenic *in silico* perturbations were performed and are summarized in **Table 3.4** In the table, the conditions that generated the lowest cost penalty for each cell line are indicated in green. Of the tested perturbations, the normal-like MCF10A cells compared most favorably to the network conditions with PTEN ***ON*** and all other nodes randomized, which is representative of normal-like conditions. MCF7 cells compared most favorably to network conditions with PTEN ***ON*** and insulin receptor ***OFF***. This result suggests that IRS1, or a downstream target of IRS1, could be deactivated or otherwise modulated in MCF7 cells. As presented in **Figure 3.7**, SUM-149 cells compared most favorably to network conditions with PTEN ***OFF*** and RAS ***ON***. Surprisingly, metastatic MDA-MB-231 cells compared most favorably to the normal-like conditions (PTEN ***ON*** and all other nodes randomized). This finding may suggest that our current logic model is not sufficiently detailed to account for this cell line's dysregulation. Alternatively, it may suggest that the limited number of perturbations tested (**Table 3.4**) do not fully account for the extreme dysregulation of this cell line.

A **PI3K and MAPK network simulations with PTEN active (early response)**



B ***in silico* state (probability to be *ON*)**

	--	EGF	LY	EGF+LY	UO	EGF+UO
AKT	0.30	0.50	0.00	0.00	0.20	0.40
ERK	0.30	0.30	0.50	0.50	0.00	0.00

C **Cost penalties for difference between simulation and experiment**

MCF10A

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	0.00	0.00	0.00	0.00	0.20	0.00	*0.20*
ERK	0.30	0.70	0.00	0.50	0.00	0.00	*1.50*
							1.7

MCF7

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	0.30	0.00	0.00	0.00	0.20	0.00	*0.50*
ERK	0.30	0.70	0.50	0.50	0.00	0.00	*2.00*
							2.5

SUM149

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	0.70	0.50	0.50	0.00	0.00	0.60	*2.30*
ERK	0.00	0.70	0.50	0.50	0.00	0.50	*2.20*
							4.5

MDA-MB-231

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	0.00	0.00	0.50	0.00	0.20	0.00	*0.70*
ERK	0.70	0.70	0.50	0.50	0.00	0.00	*2.40*
							3.1

D **PI3K and MAPK network simulations with PTEN active (late response)**

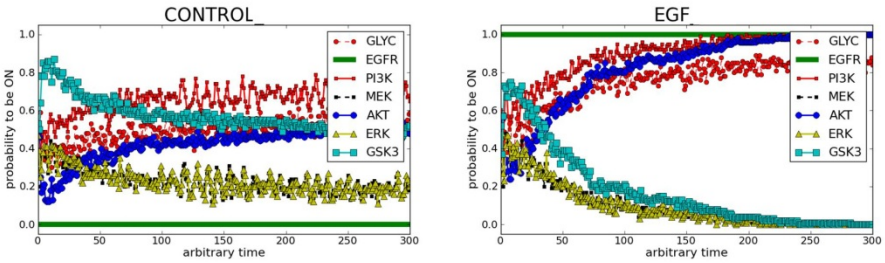
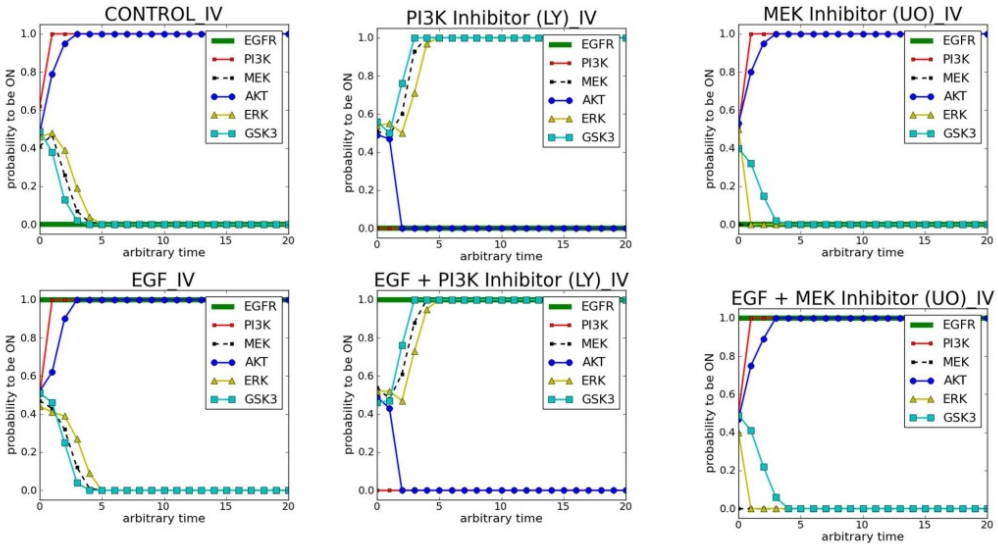


Figure 3.6. Results of PI3K/MAPK cross-talk logic model simulations under normal-like conditions. The six experimental conditions were simulated in a normal-like network model where PTEN was set to ***ON.*** For each of experimental condition, the network was perturbed to match the *in vitro* experimental conditions as described in the **Methods**. (**A**) The simulated results of each condition via asynchronous update are presented along with (**B**) a summary of the simulated results for **AKT** and **ERK** under each condition. (**C**) A summary of the difference in *in silico* simulated results and the experimental results for each condition are presented as cost penalties. A penalty of 0.00 indicates that the simulation and experimental results were qualitatively identical. (**D**) Plots of the Control and EGF simulations in (**A**) extended to 300 arbitrary time steps.

A **PI3K and MAPK network simulations with PTEN *OFF* and RAS *ON***



B ***in silico* state (probability to be *ON*)**

	--	EGF	LY	EGF+LY	UO	EGF+UO
AKT	1.00	1.00	0.00	0.00	1.00	1.00
ERK	0.00	0.00	1.00	1.00	0.00	0.00

C **Cost penalties for difference between simulation and experiment**

MCF10A

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	0.25	0.25	0.00	0.00	1.00	0.25	*1.75*
ERK	0.00	1.00	0.25	0.00	0.00	0.00	*1.25*
							3.0

MCF7

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	1.00	0.25	0.00	0.00	1.00	0.25	*2.50*
ERK	0.00	1.00	1.00	0.00	0.00	0.00	*2.00*
							4.5

SUM149

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	0.00	0.00	0.50	0.00	0.25	0.00	*0.75*
ERK	0.50	1.00	0.00	0.00	0.00	0.50	*2.00*
							2.75

MDA-MB-231

	--	EGF	LY	EGF+LY	UO	EGF+UO	Penalty
AKT	0.25	0.25	0.50	0.00	1.00	0.25	*2.25*
ERK	1.00	1.00	1.00	0.00	0.00	0.00	*3.00*
							5.25

Figure 3.7. Results of PI3K/MAPK cross-talk logic model simulations under non-normal conditions with PTEN *OFF* and RAS constitutively *ON*. The six experimental conditions were simulated in a normal-like network model where PTEN was set to ***ON***. For each of experimental condition, the network was perturbed to match the *in vitro* experimental conditions as described in the **Methods**. (**A**) The simulated results of each condition via asynchronous update are presented along with (**B**) a summary of the simulated results for **AKT** and **ERK** under each condition. (**C**) A summary of the difference in *in silico* simulated results and the experimental results for each condition are presented as cost penalties. A penalty of 0.00 indicates that the simulation and experimental results were qualitatively identical.

3.4 Discussion

The objective of the logic network developed in this work was to construct and validate a predictive network model that can be used to investigate the regulatory relationships between PI3K signaling and glucose metabolism. Before the model can be used to make robust predictions about which signaling molecules exert the greatest degree of control over various components of

glucose metabolism, however, the model must be validated with experimental data. The iterative network construction generated a model that was able to partially recapitulate experimental data in the four cell lines investigated. In general, the network did better at estimating AKT activation state than it did ERK activation. In **Table 3.4**, the numbers shaded green identify the network perturbation condition that most closely match each cell line (i.e., generated the lowest cost penalty). For MCF10A, MCF7, MDA-MB-231, and SUM-149 cells, the total cost penalties for AKT activation were 0.20, 0.10, 0.70, and 0.75, respectively, for each network condition that produced the lowest total cost penalty for each cell line. In contrast, the total cost penalties for ERK activation were 1.50, 1.80, 2.40, and 2.0, respectively, for the same conditions. These results suggest that some direct and indirect regulations of ERK are not correct in the current model.

Both the PI3K and MAPK pathways are highly dysregulated in cancer. Thus far, therapeutic targeting of the MAPK pathway has not been very effective because of so called “de novo” resistance and acquired therapeutic resistance [223]. It has been suggested that RAS to ERK signaling in the MAPK pathway represents an important “escape mechanism” when PI3K is inhibited [224]. Several regulatory cross-talk mechanisms have been suggested to explain the apparent cross-talk between these two pathways. Future work will include a detailed sensitivity analysis of each node in the logic network. Systematically perturbing each node and measuring the change from a baseline network response will allow for the ranking of the relative importance of each node to the overall robustness of the network as well as to any specific functional response (e.g., to ERK activation). A few examples of similar analyses can be found in the literature [226].

Another type of sensitivity analysis that will be important for models of this type is the systematic perturbation of each logic rule as well as each edge in logic network. This latter type of analysis will be particularly important for evaluating the relative importance of proposed feedback regulation between MAPK and PI3K signaling in the literature. In the course of my literature review

of MAPK and PI3K signaling, many proposed feedback mechanisms driving the cross-talk between these two pathways were found. In general, only those that I had a high confidence in, based on available data, were included in the current iteration of the logic model (**Figure 3.5**). Nonetheless, I am skeptical that all feedback loops included in the current model are valid based on a preliminary sensitivity analysis (data not shown). A systematic evaluation of each feedback regulation according to its ability to reproduce experimental data will be important for validating the likelihood that all feedback loops in the current model represent a valid regulatory mechanism.

Glycolytic oscillations

An interesting feature in some of the simulations performed was the presence of glycolytic oscillations (red dashed line in **Figure 3.4A** and **Figure 3.6A**). I am not aware of any studies that have directly reported glycolytic oscillations in cancer cells. The switch from tetrameric to dimeric forms of PKM2 [22, 23, 227], which appears to be regulated by cellular serine and FBP levels, has been proposed as a potential source of oscillations, however [228]. Future experimental validation of this prediction would open up important therapeutic avenues for cancer treatment.

OR regulation of AKT

As discussed in **Chapter 2**, the output of a logic model is very sensitive to the nature of the logic rules constructed. Determining whether to use an ***AND*** or ***OR*** logic rule can have profound effects on the network output (**Figure 2.5**). In **Table 3.3**, I have provided justification for all logic rules used in our network simulations. The activation of AKT was particularly important to understand in order to construct an appropriate logic function for its regulation in the model. PIP3 and mTORc2 both play important roles in activating ATK at thr308 and ser473, respectively. AKT requires phosphorylation of both of these residues for full activation, suggesting that the AKT rule could be written as AKT^{t+1} = PIP3 ***AND*** mTORc2. However, it is possible that AKT retains important downstream function when only one residue is phosphorylated. If this is the case, then the appropriate rule would be AKT^{t+1} = PIP3 ***OR*** mTORc2. PTEN is a powerful

tumor suppressor that inhibits PIP3, preventing the activation of AKT at thr308. There is evidence that PTEN expressing cells retain some PI3K dependent AKT function, suggesting that phosphorylation at ser473 alone may be enough for downstream function [183, 229]. Evidence also suggests that PDPK1 and PDPK2 are required for activation of each residue, respectively. For simplicity we assume that PDPK1 and PDPK2 are both available in the system and that the presence of PIP3, PTEN, and mTORc2 are the rate limiting steps in the activation of AKT. For these reasons, the regulation of AKT is controlled by an ***OR*** logical operator so that either active PIP3 or active mTORc2 can activate AKT in our model. In **Appendix B,** a table that summarizes the cost penalties for molecular perturbations when the logic operator regulating AKT was an ***AND*** instead of an ***OR*** is provided for comparison with **Table 3.4**.

Limitations

A limitation of the model is the lack of isoform level regulation. Experimentally we found differences in AKT1, AKT2, and AKT3 expression in some of cell lines used in this study (data not shown). Characterizing the exact function of these AKT isoforms in the literature is difficult, however. Some important functional differences have been discovered about AKT1 and AKT2 but less is known about AKT3, and little is known about how each isoform affects metabolic regulation. The future inclusion of AKT isoform specific detail is expected to improve the predictive power of the model.

Another limitation of the model is the assumption that PI3K inhibition by LY is absolute. In reality PI3K may still retain a low level of function after inhibition with LY. The use of wortmannin, which has been shown to be a more complete and irreversible inhibitor of PI3K, as an additional control may be helpful [221]. U0126, the MEK inhibitor used in this study, is a very potent inhibitor of MEK [230]. It has also been shown to severely disrupt the mitochondrial membrane, and this effect appears to be independent of its inhibition of MEK [231]. My own experimental work demonstrated that U0126 significantly increased glycolysis and lactate production in the cancer cell lines used in this work (data not shown). PD325901, another potent MEK inhibitor, did

not have the same effect, however. U0126, which is widely used in the literature, is likely still appropriate for studies that only consider molecular changes at the protein level because induced mitochondrial effects appear to be downstream of MEK activity. It is possible, however, that the mitochondrial membrane disruption induced from U0126 generates feedback from metabolism that is ultimately manifested in signaling dynamics. The use of U0126 in future studies that rely explicitly on metabolic readouts (e.g., glucose uptake rates, lactate production rates, or flux analysis) is not appropriate, however.

In addition, it would be instructive to know the P53 activation state for each cell line under the different treatment conditions used in this study. The phospho-western blot experiments were performed under serum starvation and there are data to suggest that, under such conditions, P53 mediates mTORC1 sensitivity to an mTOR/raptor inhibitor. Under nutrient rich conditions, however, the effect of P53 was abrogated along with the sensitivity to the inhibitor [232]. Therefore, it is possible that P53 cell stress responses may be induced in the cell lines studied because of serum deprivation. If that is the case, refining the model to adequately account for the effects of such a response will be important.

3.4.1 Future directions

In the future, the objective is to improve the model presented here so that it can be used to make predictions about the regulatory control of PI3K signaling over glucose metabolism at distinct stages of breast cancer. To effectively build such a model will most likely require the development of a hybrid discreet continuum model. Hybrid models link discrete and continuous models across timescales and are widely used in the engineering and computational sciences. In models of tumor growth, cells can be modeled as discrete entities that respond to continuous models of intracellular and extracellular signals [87, 233-236]. For example, Ribba et al. [87] developed a multiscale model that linked a set of discrete models with continuous models of colorectal cancer growth. The model accounted for the cellular, genetic, and environmental factors regulating tumor growth. It integrated key oncogenes involved in colorectal cancer evolution with a Boolean gene network that was regulated by a discrete cell cycle model. The

response to signals from the intracellular gene network determined whether each cell proliferated or died in the discrete model and, therefore, directly influenced the cellular and the extracellular tissue levels. The combination of discrete and continuous modeling was used to predict the qualitative effect of therapeutic protocols on colorectal cancer and demonstrated that the efficacy of irradiation protocols depend on the type of anti-growth signals to which the tumors are exposed. In another model developed by Singhania et al. [237], a continuous model of the cell cycle was linked to a logic model of genetic pathways regulating critical substrates during cell cycle progression. By combining a continuous ODE model with a discrete Boolean model, the authors obtained a piecewise ODE model system.

In a similar manner, I propose it is possible to combine ODE based models of metabolism with discrete logic-based signaling models, such as the model presented here. While discrete and continuous hybrid models have been used in cancer research for more than 10 years, I am not aware of any that have directly linked metabolism and signal transduction. To successfully implement such a model, timescale separation will need to be carefully considered. As noted in **Chapter 1**, the general paradigm for modeling metabolism assumes that enzyme concentrations do not change during the progression of enzyme-catalyzed reactions. While the assumption may be reasonable for many metabolic models, it will likely not be appropriate to truly understand the complex feedback and control between PI3K signaling and glucose metabolism in cancer.

Table 3.3. Justifications for logic functions used in the PI3K/MAPK cross-talk model in Figure 3.5.

Logic Rule	Rule Justification	Source(s)
AKT^{t+1} = PIP3 ***OR*** mTORc2	PIP3 recruits AKT to the cell membrane where AKT is activated by PDK1. In addition, there is evidence that mTORc2 is essential for full AKT activation (by directly phosphorylating AKT, likely due to PDK2 activity). While AKT requires both PIP3 and mTORc2 for full activation, we assume that partial activation is possible when only one is present. In support of this is data indicating that cells expressing PTEN were still able to activate AKT. We also assume PDK1 and PDK2 are present in the system but do not include them in the model for simplicity.	[153, 172, 183, 229]
$AMPK^{t+1}$ = p53 ***AND NOT*** ATP	AMPK is regulated by both P53 and the ATP/AMP ratio. Declining levels of ATP and rising levels of AMP active AMPK, which is considered a conserved sensor of the energy state within a cell. In addition, expression levels of AMPK have been shown to increase after P53 activation.	[238, 239]
$Apoptosis^{t+1}$ = p53 ***AND NOT*** AKT	P53 transcriptionally activates BAX, which is an upstream activator of apoptosis. AKT also regulated BAD. Both BAD and P53 are positive regulators of BAX, which is a direct regulator of apoptosis in this network. It is assumed that the activation of BAD (via AKT inhibition) and the activation of P53 are required to induce apoptosis in the model.	[152, 164]
ATP^{t+1} = Glycolysis	A net gain of 2 ATP per molecule of glucose is produce in glycolysis when glucose is metabolized to pyruvate	[166, 186, 198]
ERK^{t+1} = MEK	ERK1/2 is directly activated by MEK in the MAPK pathway	[175]
$Fru26P2^{t+1}$ = PFK2 ***AND NOT*** TIGAR	Fru26P2 is the product of PFK2, while TIGAR is potent inhibitor of Fru26P2 levels. We assume that TIGAR is dominant relative to PFK2 in the regulation of Fru26P2 expression (Li et al)	[166, 240]
$GAB1^{t+1}$ = EGFR ***OR*** PIP3	EGFR alone can activate GAB1 and the Ras/MAPK pathway. PI3K can also bind to the PH domain of GAB1. PIP3 can enhance activation of GAB1. The rule is written as EGFR or PIP3 because loss of Gab1's PH domain was not found to not eliminate tyrosine phosphorylation of Gab1 in response to EGF. Therefore PIP3 enhances activation of GAB1, but is not required for activation.	[188-190]
$Glut/HK^{t+1}$ = mTORc1 ***OR*** AKT	Data suggest that AKT up regulates Glucose receptors (GLUT) as well as hexokinase (HK). In addition, data suggest mTORc1 can up regulate GLUT and HK a via HIF1(alpha) regulation. Therefore, either mTORc1 or AKT may induce GLUT and HK in the model, which is treated as a single node for simplicity.	[166]
$Glycolysis^{t+1}$ = Glut/HK ***AND*** PFK1	Data suggests that AKT upregulates Glut receptors and that mTORc1 can also upregulate GLUT receptors a via HIF1(alpha) regulation. Therefore either mTORc1 or AKT may induce GLUT receptors in the model.	[166, 186]
$Grb2/SOS^{t+1}$ = (Shc ***OR*** IRS1) ***AND NOT*** ERK	Activation of Shc recruits the Grb2/SOS complex via SH2 domain binding. IRS1 can also bind to and activate the Grb2/SOS complex via SH2 domain binding. Either Shc or IRS1 are needed to recruit the Grb2/SOS complex, however, which is why that part of the rule is written as an OR. ERK has the ability to dephosphorylate SOS directly, resulting in disassociation of the Gbr2/SOS complex and a decrease in Ras-GTP. Thus, we assume that ERK inhibition is dominant relative to the activators.	[44, 191, 195-197, 201, 202, 205]

Logic Rule	Rule Justification	Source(s)
$IRS1^{t+1}$ = IR ***AND NOT*** S6K1	mTORc1 inhibits IRS1 via S6K1. IRS1 binds to and is activated by insulin receptors (IR). We assume that inhibition is dominant relative to activation.	[197]
$MDM2^{t+1}$ = p53	P53 mediates the transcription of MDM2.	[179]
MEK^{t+1} = Raf	RAF is upstream activator of MEK.	[175, 241]
$mTORc1^{t+1}$ = **NOT** TSC1/2	TSC1 and TSC2 form a heterodimer that inhibits the activity of Rheb. Rheb is a small GTPase required for activation of mTORc1. Thus, in the model we assume that mTORc1 is inactivated by TSC1/2 activity.	[153]
$mTORc2^{t+1}$ = PI3K ***AND*** Ribosomes	Evidence strongly suggests that PI3K is both necessary and sufficient for mTORc2 activation to occur via mTORc2 ribosomal binding. Thus, we assume both PI3K and ribosomes are necessary for mTORc2 activation.	[169, 170]
$p53^{t+1}$ = DNA Damage ***AND NOT*** MDM2	P53 is transcriptionally activated in the presence of DNA Damage and is targeted for degradation by MDM2 ubiquitination (Refer to Chapter 2, Figure 2.6 for additional explanation of this rule)	[137]
$PFK1^{t+1}$ = Fru26P2 ***OR NOT*** ATP	Fru26P2 is activated PFK1 by allosteric regulation and inhibited by high cellular ATP levels. We assume this regulation is an OR NOT regulation because we assume that activator is domination because of evidence part of Fru26P2 allosteric regulation includes decreasing PFK1 affinity for AKT.	[242, 243]
$PFK2^{t+1}$ = AKT	AKT directly activates PFK2 by phosphorylation	[166]
$PI3K^{t+1}$ = ((IRS1 ***OR*** (GAB1 ***AND NOT*** Erk)) ***AND NOT*** SHP2) ***OR*** RasGTP	This is a particularly complicated regulation. GAB1 activates PI3K by recruiting the p85 subunit and ERK negatively regulates PI3K by phosphorylating GAB1 and rendering it unable to recruit p85. This portion of the rule is written as: GAB1 and not ERK. In addition, IRS1 activates PI3K but both GAB1 and IRS1 activation can be regulated by SHP2; therefore, this portion of the rule is written as: [(GAB1 and not ERK) or IRS1] and not SHP2. Finally, Ras-GTP can also activate PI3K, but its activation is not affected by SHP2 or ERK levels, thus overall activation of PI3K can take place via GAB1, IRS1, or Ras-GTP.	[5, 171, 180, 188, 189, 194, 200, 203, 208, 244, 245]
PIP3t + 1 = PI3K and not PTEN	PI3K activates PIP3 at the cell membrane and PTEN dephosphorylates PIP3. We assume the inhibitor is dominant.	[153]
Raf^{t+1} = Ras-GTP ***AND NOT*** AKT	RAS-GTP directly activates RAF and RAF must be phosphorylated and dephosphorylated on specific sites to remain active. Regardless of the activation state of RAF, AKT appears able to phosphorylate RAF and reduce activation levels. Therefore, the rule is written as Ras-GTP and not AKT.	[181, 182] [191]
$RasGAP^{t+1}$ = GAB1 ***AND NOT*** SHP2	GAB1 recruits RasGAP to the membrane and SHP2 down regulates this interaction. Without SHP2, GAB1 is able to recruit RasGAP to the membrane. In the presence of SHP2, GAB1 is not able to recruit RasGAP to the membrane and RasGTP becomes activated. If GAB1 and SHP2 are both absent, then RasGAP is not recruited to the membrane, which is the primary justification for the AND NOT function.	[188, 205]
$Ras\text{-}GTP^{t+1}$ = Grb2/SOS ***AND NOT*** RasGAP	This rule is based on the observation that RasGAP binds to Ras and does not allow the exchange of GDP to GTP, therefore RasGAP must be removed AND the Grb2/SOS complex must be present to catalyze the exchange of GDP to GTP, hence the rule is Grb2/SOS and not RasGAP.	[188, 191]
$Ribosomes^{t+1}$ = S6K1	Phosphorylation of S6K1 by MTORc1 induces ribosomal biogenesis.	[238]

Logic Rule	Rule Justification	Source(s)
$S6K1^{t+1}$ = mTORc1	Phosphorylation of S6K1 is induced by mTORc1.	[238]
Shc^{t+1} = EGFR ***OR*** IR	Either activated EGFR or activated IR can stimulate RAS. Because Shc activates the downstream Grb/SOS complex, which is essential for RAS activation, we assume that Shc is intermediary between this complex and the upstream receptors in the model.	[44, 188, 191, 201, 204, 205]
$SHP2^{t+1}$ = GAB1 ***OR*** IRS1	SHP2 is crucial for Ras activation in insulin and EGF signaling. Activation of the RAS/MAPK pathway has been observed under insulin/IRS1 activation as well as under EGF/GAB1 activation. Thus, the rule is written as GAB1 or IRS1.	[188, 205]
$TIGAR^{t+1}$ = p53	TIGAR is induced by P53.	[143]
$TSC1/2^{t+1}$ = ***NOT*** AKT ***AND NOT*** ERK ***AND NOT*** AMPK	AKT inhibits TSC1/2 and ERK inhibits TSC1/2 through the activation of RSK1. Phosphorylation (ser1798) of RSK1 inhibits TCS1/2 and results in increased mTORc1 expression. We assume that TSC12 is only on when its activator is present and both its inhibitors are absent. While there are a number of negative regulators of the TSC12 dimer, our model only considers ERK inhibition and AKT inhibition of TSC12.	[193]

Table 3.4. Oncogenic perturbations to the PI3K/MAPK cross-talk model. The table summarizes the difference in *in silico* simulated results and the experimental *in vitro* results by displaying the total AKT, total ERK and total combined cost penalty for each cell line and each condition. The conditions that generated the lowest total cost penalty are displayed in green and the conditions that generated the highest total cost penalty are displayed in red for each cell line. A penalty of 0.00 indicates that the simulation and experimental results were qualitatively identical for all 6 experimental conditions.

Perturbations																
Inputs			Constitutive		MCF10A Penalties			MCF7 Penalties			231 Penalities			149 Penalities		
PTEN	Stress/P53	IR	MTORc2	RAS	AKT	ERK	Total	AKT	ERK	Total	AKT	ERK	Total	AKT	ERK	Total
RAND	*OFF*	*OFF*	--	--	1.00	1.60	2.60	1.25	1.90	3.15	1.25	2.50	3.75	1.40	2.60	4.00
ON	*RAND*	*OFF*	*ON*	--	3.75	2.50	6.25	4.50	2.00	6.50	3.00	3.00	6.00	1.50	4.00	5.50
ON	*RAND/OFF*	*OFF*	--	--	0.50	1.50	2.00	0.10	1.80	1.90	1.00	2.20	3.20	3.10	2.40	5.50
ON	*RAND/OFF*	*RAND*	--	--	0.20	1.50	1.70	0.50	2.00	2.50	0.70	2.40	3.10	2.30	2.20	4.50
OFF	*RAND*	*OFF*	*ON*	--	3.75	2.50	6.25	4.50	2.00	6.50	3.00	3.00	6.00	1.50	4.00	5.50
OFF	*RAND/OFF*	*RAND/OFF*	--	--	1.25	1.60	2.85	2.05	2.00	4.05	1.75	2.80	4.55	1.15	3.10	4.25
OFF	*RAND*	*OFF*	*ON*	*ON*	3.75	2.50	6.25	4.50	2.00	6.50	3.00	3.00	6.00	1.50	4.00	5.50
OFF	*RAND/OFF*	*RAND/OFF*	--	*ON*	1.75	1.25	3.00	2.50	2.00	4.50	2.25	3.00	5.25	0.75	2.00	2.75
ON	*RAND*	*OFF*	*ON*	*ON*	3.75	2.50	6.25	4.50	2.00	6.50	3.00	3.00	6.00	1.50	4.00	5.50
ON	*RAND*	*RAND/OFF*	--	*ON*	1.00	1.25	2.25	1.75	2.00	3.75	1.50	3.00	4.50	1.00	2.00	3.00
ON	*OFF*	*RAND/OFF*	--	*ON*	0.50	1.25	1.75	1.25	2.00	3.25	1.00	3.00	4.00	1.40	2.00	3.40

Chapter 4
Targeted inhibitors can produce off-target effects and activate linked pathways by retroactivity

The content of this chapter was previously published by the author in *BMC Systems Biology* [21].

4.1 Background

Chapters 2 and **3** explored the use of logic based network models as a qualitative and parameter-free network modeling method. As discussed at the end of **Chapter 2**, however, some forms of network behavior cannot be adequately modeled used logic-based methods. In this chapter, an ordinary differential equation (ODE) system is used to explore whether off-target effects can be induced by phenomenon driven by enzyme sequestration, a mechanistic detail that is out of reach of parameter-free logic methods.

Cells propagate information through protein signaling pathways that are part of complex signal transduction networks [34]. The simplest view of cellular signaling entails a cascade of molecular events initiated by the recognition of a stimulus and culminating in the chemical alteration of an effector molecule. In the case of covalent modification by the addition or removal of a phosphate group, a reaction commonly found in signaling cascades, each phosphorylated protein serves as the kinase that activates the next cycle's unphosphorylated protein.

Targeted therapies are used to modulate disease progression by inhibiting a specific protein within a dysregulated signaling pathway [246]. Kinase inhibitors are a class of targeted therapies designed to interfere with a specific kinase molecule. While extremely promising as anti-cancer agents, kinase inhibitors can produce off-target effects by inducing changes in molecules other

than the one specifically targeted. Such off-target effects are generally attributed to non-specific binding or to cross-talk [247].

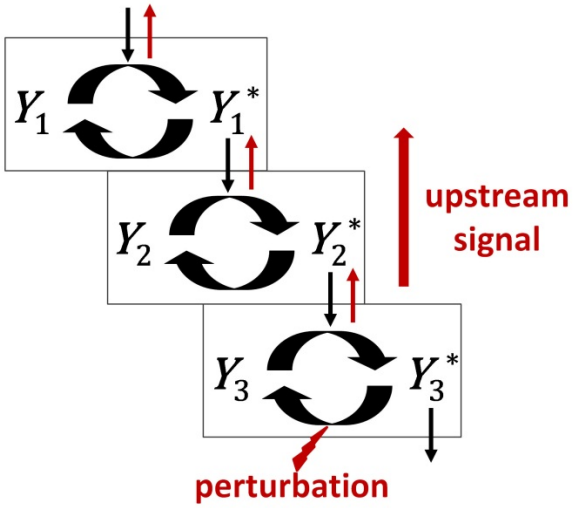


Figure 4.1. Retroactivity arises from enzyme sequestration in covalently modified cascades. A simple signaling cascade is depicted where each sequential cycle represents the activation (denoted by *) and inactivation of a protein Y_i. Y_1* serves as the activating enzyme of Y_2 and Y_2* serves as the activating enzyme of Y_3. The cycles can be thought of as modules where each module's substrate sequesters a key competent of the previous module, limiting the component's ability to participate in the previous module. This sequestration induces a natural change in the preceding module, which may propagate upstream through one or more preceding modules. In this example, a perturbation in the deactivation reaction of cycle 3 induces an effect in cycle 2. If the perturbation takes the form of an increase in the concentration or activity of the enzyme catalyzing the conversion of Y_3* to Y_3, more Y_3 will be available to react with and sequester Y_2*, resulting in less Y_2 substrate availability for the reaction with Y_1*. Thus, a reverse response can propagate upstream to a preceding cycle or cycles. In the schematic, black arrows represent the cell surface to nucleus direction of cellular signaling and red arrows represent the direction of retroactive signaling.

Recent theoretical and experimental studies have demonstrated that covalently modified cascades also exhibit bidirectional signal propagation via a phenomenon termed retroactivity [248-253]. This phenomenon arises because cycles in a cascade are coupled, not only to the next cycle, but also to the previous cycle (**Figure 4.1**). The cycles can be thought of as modules where each module's substrate sequesters a key component of the previous module, limiting the component's ability to participate in the previous module and inducing a natural change in the preceding module. This change may then propagate upstream through one or more preceding modules.

While retroactivity is naturally present in covalently modified cascades, signaling pathways likely have evolved to propagate signals in an optimized

downstream manner. An important consequence of retroactivity, however, is that a downstream perturbation in a signaling cascade can produce an upstream effect without the need for explicit negative feedback connections [248]. Retroactivity may, therefore, play important roles in the dysregulated signaling networks of diseased cells as well as the cellular response to targeted therapies applied to dysregulated signaling networks.

Ventura, Sepulchre, and Merajver [248] demonstrated that increasing the concentration of the inactivating enzyme (e.g., a phosphatase) in the terminal cycle of a cascade can decrease the concentration of the activated protein in the previous cycle [248]. This finding led us to hypothesize that a targeted inhibitor can produce upstream off-target effects via retroactivity that can propagate elsewhere in the signaling network.

Off-target effects associated with targeted therapies are often attributed to cross-talk, which refers to inter-pathway molecular interactions arising because of explicit regulatory feedback connections between two pathways or because two pathways share one or more molecular components. It is well accepted that two pathways sharing one or more components can exhibit cross-talk with respect to a stimulation or perturbation above the shared component(s). If an upstream perturbation occurs in one of the pathways, the perturbation may affect the other pathway via the shared downstream component(s). Such a scenario could lead to specificity problems [254]. Here we propose that perturbations (e.g., from an inhibitor) that occur downstream of a shared component can also induce cross-talk effects without any explicit feedback connections via the following mechanism: the information travels upstream from the site of the perturbation through retroactivity, reaches the common component and then is delivered to the parallel pathway.

To test our hypothesis, we created a computational model that tested the application of a kinase inhibitor in a series of simple signaling networks. The objective of the model was to probe the effect of a targeted inhibitor on retroactive signaling and to test whether retroactivity is likely to contribute to measurable off-target effects under physiological conditions. Specifically, the

model simulated the targeted inhibition of a specific kinase in a series of multi-cycle networks. In all networks, at least two cascades were activated by the same upstream cycle with no explicit feedback connections between them. Our results indicate that within physiologically and therapeutically relevant ranges for all parameters, a targeted inhibitor can naturally induce a steady state off-target effect via retroactivity. Our results also reveal the surprising and crucial result that a downstream kinase inhibitor has the capacity to turn “on” an otherwise “off” parallel cascade when two cascades share an upstream activator.

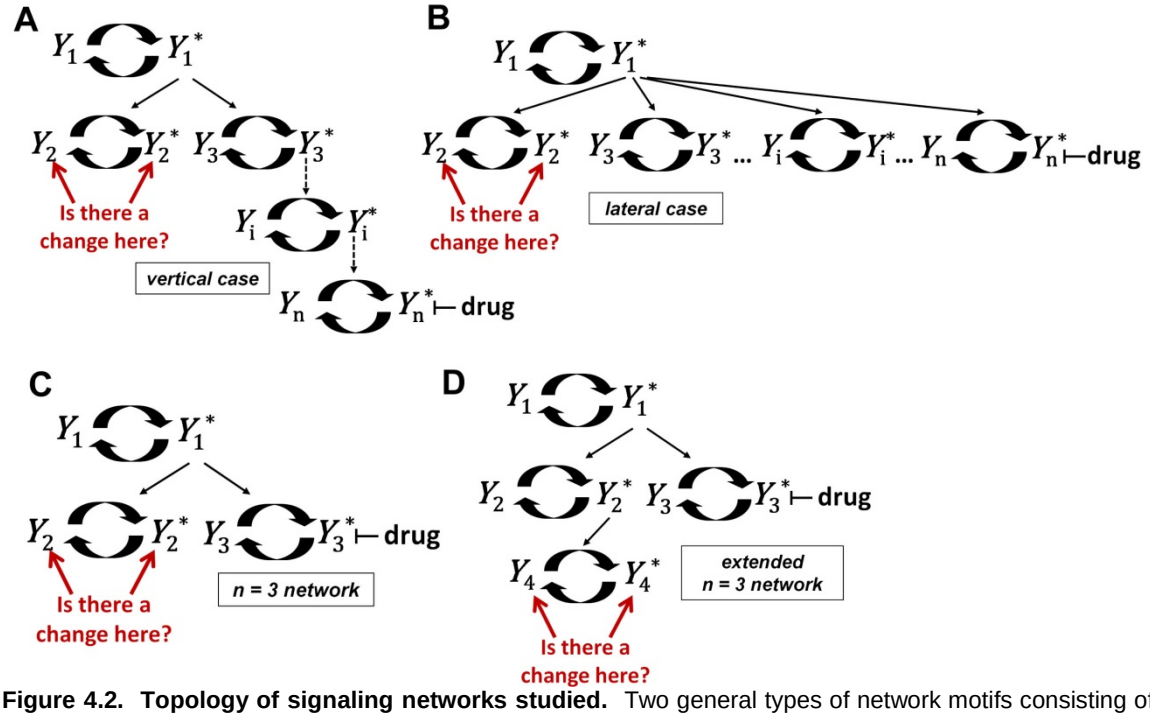


Figure 4.2. Topology of signaling networks studied. Two general types of network motifs consisting of covalently modified cycles were studied: **(A)** the vertical case where the n-th cycle in the right hand cascade is inhibited and **(B)** the lateral case where the n-th single-cycle cascade is inhibited. Y_1* served as the upstream activator for both cycle 2 and cycle 3, which were always in distinct cascades. No additional regulatory connections were included in any network. Off-target effects were monitored in the networks by measuring the steady state concentrations of Y_2 and Y_2* as the concentration of an inhibitory drug that specifically targeted Y_n* was increased. **(C)** The n = 3 network consisted of exactly 3 cycles and was the simplest form of both the vertical and lateral case. **(D)** An *extended* n = 3 network was also studied where a fourth cycle activated by Y_2* was added to the left-most cascade.

Table 4.1. The parameter space of each network consisted of a set of non-dimensional parameters, each with a minimum and maximum allowed value. Each cycle *i*, consisted of 5 dimensionless parameters: E_i, E'_i,K_i, K'_i, and P_i. A final parameter, K_B, applied to the targeted inhibitor. Randomly selected parameter values could not exceed the default ranges listed for each parameter type.

	default range		
parameter	minimum	maximum	*description*
E_i	0.01	100	total kinase to total substrate ratio
E'_i	0.01	100	total phosphatase to substrate ratio
K_i	0.01	100	normalized K_m of kinase reaction
K'_i	0.01	100	normalized K_m of phosphatase reaction
P_i	0.1	10	ratio of the kinase reaction V_{max} to the phosphatase reaction V_{max}
K_B	0.01	100	normalized drug disassociation constant

4.2 Methods

Model development

We designed simple signaling networks to test whether a measurable off-target effect in one cascade can occur when a protein in another cascade is selectively inhibited. In each network studied, cycle i contained the active (phosphorylated) and inactive (unphosphorylated) forms of protein Y_i, where the active form was denoted by Y_i*. For simplicity, we refer to activating and inactivating enzymes in a network as kinases and phosphatases, respectively.

Protein Y_1* served as the activating kinase for all cascades. Cycle 2 and cycle n were always in distinct cascades (**Figure 4.2**). To determine if an off-target effect occurred due to perturbation by the inhibitor, the steady state concentration of the protein in cycle 2 was monitored as the concentration of the drug that specifically targeted Y_n* was increased. A competitive inhibitor was used that directly bound to Y_n*, limiting its ability to participate in the phosphatase reaction of cycle n, but did not change the rate of the phosphatase reaction in cycle n.

Two general network types were considered: a vertical and a lateral case (**Figure 4.2**). The vertical case consisted of two cascades where the inhibited cascade length varied (**Figure 4.2A**). This motif is similar to the upstream activation of JUN and P53 by JNK1/2 in the mitogen-activated protein kinase

(MAPK) pathway [255]. The lateral case was equivalent to a fan-out network topology and consisted of n - 1 single cycle cascades that were all activated by Y_1* (**Figure 4.2B**). This motif is similar to the activation of multiple cascades by p38 [255]. The n = 3 network consisted of exactly 3 cycles and represented the simplest form of both network types studied (**Figure 4.2C**).

The general reaction scheme used was:

$$Y_i + E_{k_i} <\frac{a_i}{d_i}> C_i \xrightarrow{k_i} Y_i^* + E_{k_i}$$

$$Y_i^* + E_{p_i} <\frac{a_i'}{d_i'}> C_i' \xrightarrow{k_i'} Y_i + E_{p_i}$$

$$Y_n^* + D <\frac{k_{on}}{k_{off}}> C_D$$

$$E_{k_i} = \begin{cases} i = 1, & E_{k_1} \\ i = 2, i = 3, & Y_1^* \\ i > 3 \text{ (vertical)}, & Y_{i-1}^* \\ i > 3 \text{ (lateral)}, & Y_1^* \end{cases}$$

where

Y_i^* is the activated protein in the i^{th} cycle

Y_i is the inactivated protein in the i^{th} cycle

E_{k_i} is the kinase enzyme in the i^{th} cycle

E_{p_i} is the phosphatase enzyme in the i^{th} cycle

D is the inhibitory drug

C_i is the Y_i and E_{k_i} complex in the i^{th} cycle

C_i' is the Y_i^* and E_{p_i} complex in the i^{th} cycle

C_D is the Y_n^* and D complex in the n^{th} cycle

and values above and below arrows represent rate constants.

Parameter definitions

In order to reduce the complexity of each network studied, parameters were non-dimensionalized into 4 parameter types as described in **Appendix C**. The allowed value of each parameter type was restricted to the default ranges listed in **Table 4.1**. A summary of the parameters types is provided below.

Subscripts containing k or p indicate parameters associated with a kinase or phosphatase reaction, respectively, and subscripts containing T indicate the

total concentration of a species. $\mathrm{V_{max}}$ and $\mathrm{K_m}$ are the standard Michaelis-Menten constants representing, respectively, the maximum velocity of a reaction (at a given enzyme concentration) and the substrate concentration necessary to achieve $\frac{1}{2}\ \mathrm{V_{max}}$ [32].

(1) total enzyme to substrate ratio of the kinase and phosphatase reaction, respectively, in cycle i:

$$\boldsymbol{E_i} = E_{k_{i\mathrm{T}}}/Y_{i\mathrm{T}} \qquad \boldsymbol{E'_i} = E_{p_{i\mathrm{T}}}/Y_{i\mathrm{T}}$$

(2) normalized $\mathrm{K_m}$ of the kinase and phosphataste reaction, respectively, in cycle i:

$$\boldsymbol{K_i} = \mathrm{K_{m}}_{k_i}/Y_{i\mathrm{T}} \qquad \boldsymbol{K'_i} = \mathrm{K_{m}}_{p_i}/Y_{i\mathrm{T}}$$

where $\mathrm{K_{m}}_{k_i} = \frac{d_i+k_i}{a_i}$ and $\mathrm{K_{m}}_{p_i} = \frac{d'_i+k'_i}{a'_i}$

(3) $\mathrm{V_{max}}$ ratio of the kinase and phosphatase reactions in cycle i:

$$\boldsymbol{P_i} = \mathrm{V_{max}}_{k_i}/\mathrm{V_{max}}_{p_i}$$

where $\mathrm{V_{max}}_{k_i} = k_i E_{k_{i\mathrm{T}}}$ and $\mathrm{V_{max}}_{p_i} = k'_i E_{p_{i\mathrm{T}}}$

(4) normalized disassociation constant of the inhibitor binding to Y_n*:

$$\boldsymbol{K_B} = \frac{k_{off}/k_{on}}{Y_{nT}}$$

$\boldsymbol{E_i}$ and $\boldsymbol{E'_i}$ values less than 1 indicate that the enzyme is less abundant than the substrate. $\boldsymbol{K_i}$ and $\boldsymbol{K'_i}$ values less than 1 indicate that the total available substrate exceeds the concentration needed to reach $\mathrm{K_m}$ and, consequently, the enzymatic reaction operates close to or in the zero order regime [33]. In contrast, $\boldsymbol{K_i}$ and $\boldsymbol{K'_i}$ values greater than 1 indicate that an insufficient amount of substrate exists to reach $\mathrm{K_m}$ and the enzymatic reaction operates in the linear regime [33]. $\boldsymbol{P_i}$ values greater than 1 indicate that the $\mathrm{V_{max}}$ of the kinase reaction exceeds the $\mathrm{V_{max}}$ of the phosphatase reaction and, consequently, the cycle tends

toward the activation reaction. Likewise, $\boldsymbol{P_i}$ values less than 1 indicate that the cycle tends toward the deactivation reaction.

Determination of off-target effects

The concentrations of species Y_i, Y_i*, and the inhibitory drug D were normalized as follows:

$$y_i = \frac{[Y_i]}{Y_{iT}} \quad y_i^* = \frac{[Y_i^*]}{Y_{iT}} \quad \boldsymbol{I} = \frac{D_T}{Y_{nT}}$$

To determine if a detectable off-target effect occurred for a specific set of parameters, changes in the steady values of y_2 and y_2^* were monitored as the model parameters were held fixed but $\boldsymbol{I}$ was varied from 10^{-4} to 10^4. If a change in the steady state value of y_2 or y_2^* occurred that was greater than or equal to a detection threshold of 0.10, an off-target effect in cycle 2 was reported. For numeric reasons, the range used for $\boldsymbol{I}$ was intentionally larger than needed. For a given parameter set, it was numerically more efficient to simulate with a small (10^{-4}) and a large (10^4) value for $\boldsymbol{I}$ and then check for a change in the steady state values of y_2 and y_2^* than it was to simulate with many values of $\boldsymbol{I}$. In fact, the majority of target-effects in our model were observed as $\boldsymbol{I}$ was varied from 0.1 to 10.

When we tested the $n = 3$ network, we obtained the same results when we used either $\boldsymbol{I}$ = 0.0000 or $\boldsymbol{I}$ = 0.0001 (10^{-4}) as the minimum drug concentration. For this reason (and because it would be experimentally challenging to distinguish 0.0000 from 0.0001 *in vivo*), we effectively considered $\boldsymbol{I}$ = 10^{-4} to represent the absence of the drug in the system.

Numerical simulations

For each network tested, a system of ODEs was used to model the rate of change of the reactants. Because we were only interested in changes in steady state values as a function of $\boldsymbol{I}$, we first solved the system by setting the ODEs equal to zero and generating a system of steady state equations. As described in **Appendix C**, the model in this form was the basis for the non-dimensionalization of model parameters.

For numerical reasons, it was more efficient to solve the ODEs over a very long time period rather than solving the system of steady state equations directly. After randomly selecting a set of non-dimensional parameters, the selected values were mapped to corresponding dimensional parameter values (as described in **Appendix C**) and the system of ODEs was solved using the Matlab R2009b ode15s stiff solver from 0 to a maximum of 100,000 arbitrary time units. The majority (~90%) of randomly selected parameter sets obtained steady state within 5,000 arbitrary time units. The units are arbitrary because we began with dimensionless parameters lacking an explicit timescale. Finally, to confirm the numerical steady state solution, the original dimensionless parameters and the final y_i and y_i* variable values were substituted into the analytical steady state equations listed in **Appendix C**. Matlab source code was compiled as a C program and run on Intel Nehalem/i7 Core processors.

Random parameter space exploration

Random parameter selection was performed via latin hypercube sampling (LHS) to provide an efficient and even sampling distribution across the range of allowed values in the parameter space [256-258]. Each parameter space exploration consisted of 5000 randomly selected parameter sets. The number of parameter sets sampled was determined by calculating the percent of off-target effects in q randomly sampled parameter sets for the $n = 3$ network (**Figure 4.2C**). The variation in the percent of off-target effects stabilized when q was greater than or equal to 5000 (**Appendix C, Figure C.1**). The percentage of 5000 randomly selected parameter sets that produced an off-target effect provided a probability that off-target effects would occur in a given network's parameter space.

Numeric perturbation analysis

A modified perturbation method was used to probe which model parameters were most important for producing an off-target effect as a result of the inhibition of Y_n*. Traditional biochemical sensitivity analysis [259] with the dimensionless parameters was not possible because these parameters applied to the steady state equations and not the time dependent differential equations

(**Appendix C**). Instead, we developed a numerical perturbation based method that allowed us to evaluate the parametric sensitivity of off-target effects in a network's parameter space. In the method, the value of a single parameter was randomly selected from a restricted range of values while all other parameter values were randomly selected from the full range permitted by the baseline parameter space. If off-target effects are sensitive to a given parameter, we expect that when values for the parameter under test are randomly selected from a reduced range of values, the percentage of off-target effects produced will differ from the percentage produced when values for the parameter are instead selected from a fixed baseline range. In both cases, all other parameter values are selected from a fixed baseline range so that the only change in the system is a perturbation in the allowed range of the parameter under test.

The reduced ranges used to perturb each parameter were arrived at by partitioning the default range established for each parameter type in **Table 4.1**. The default ranges were divided into smaller perturbation sub-ranges such that the minimum and maximum of a sub-range was an order of magnitude larger than the minimum and maximum of the previous sub-range. Because the $\boldsymbol{E_i}$, $\boldsymbol{E'_i}$, $\boldsymbol{K_i}$, $\boldsymbol{K'_i}$, and $\boldsymbol{K_B}$ parameters had a default initial range of 0.01 – 100.0 (**Table** 4.1), the sub-ranges used to perturb these parameters were: (i) 0.01 – 0.10, (ii) 0.10 – 1.0, (iii) 1.0 – 10.0, and (iv) 10.0 – 100.0. Because the $\boldsymbol{P_i}$ parameters had a default range of 0.10 – 10.0, the sub-ranges used to perturb these parameters were: (i) 0.10 – 1.0 and (ii) 1.0 – 10.0.

A complete numeric perturbation analysis of a parameter space consisted of determining the percentage of off-target effects in 5000 randomly selected parameter sets for each parameter's sub-ranges. In the $n = 3$ network (**Figure 4.2C**) there were a total of 16 parameters (5 parameters per cycle and $\boldsymbol{K_B}$). Three of the parameters ($\boldsymbol{P_1}$, $\boldsymbol{P_2}$, and $\boldsymbol{P_3}$) had 2 perturbation sub-ranges each and the remaining 13 parameters had 4 perturbation sub-ranges each. In this example, the analysis consisted of a total of 59 sets of 5000 simulations (58 sets for each parameter sub-range and 1 set to establish the baseline percentage of off-target effects in the parameter space prior to perturbation).

4.3 Results

The question we wanted to answer with our models was whether a targeted inhibitor is likely to induce a measurable off-target effect due to retroactivity in a non-targeted cascade under physiological conditions. In each network, cycle n, was perturbed by an inhibitor. An off-target effect occurred in the model if, after increasing $\boldsymbol{I}$ (the normalized inhibitor concentration) from 10^{-4} to 10^{4}, a change in the steady state concentration of Y_2 and/or Y_2* occurred that was at least 0.10 of the total Y_2 protein pool. For example, a change of 0.25 in Y_2 and 0.08 in Y_2* would indicate that the steady state values of Y_2 and Y_2* changed by 25% and 8% of the total Y_2 protein pool, respectively, and that a detectable off-target effect occurred in Y_2.

4.3.1 Parameter ranges promoting off-target effects in cycle 2

First, we investigated the $n = 3$ network (**Figure 4.2C**) where Y_3* is targeted by the inhibitor. When the full parameter space (defined in **Table 4.1** and depicted in **Figure 4.3H**) was used, 1.6% of the 5000 randomly selected parameters sets produced an off-target effect in cycle 2. This value was essentially unchanged (1.5%) when we randomly selected 50,000 parameter sets for comparison. To identify the model parameters that were most important for producing a cycle 2 off-target effect, a numerical perturbation analysis was performed (**Figure 4.3A**-**F**). The results of the analysis suggest that the parameters controlling the activity of cycle 3 play a large role in inducing an off-target effect in cycle 2. Not surprisingly, $\boldsymbol{K_3}$ (the normalized $\mathrm{K_m}$ of the kinase reaction in cycle 3) had the greatest effect on off-target effects in this network (**Figure 4.3D**). $\boldsymbol{K_3}$ determines how much sequestration of Y_1* by Y_3 occurs and this is the key mechanism of retroactivity. When $\boldsymbol{K_3}$ was restricted to values greater than 1, the off-target effects in the network were essentially eliminated. In contrast, when $\boldsymbol{K_3}$ was restricted to values between 0.01 and 0.10, the percentage of off-target effects increased to 4.6%. Similarly, $\boldsymbol{K'_3}$ (the normalized $\mathrm{K_m}$ of the phosphatase reaction in cycle 3) also affected the percentage of off-target effects but to a lesser degree than $\boldsymbol{K_3}$ (**Figure 4.3E**).

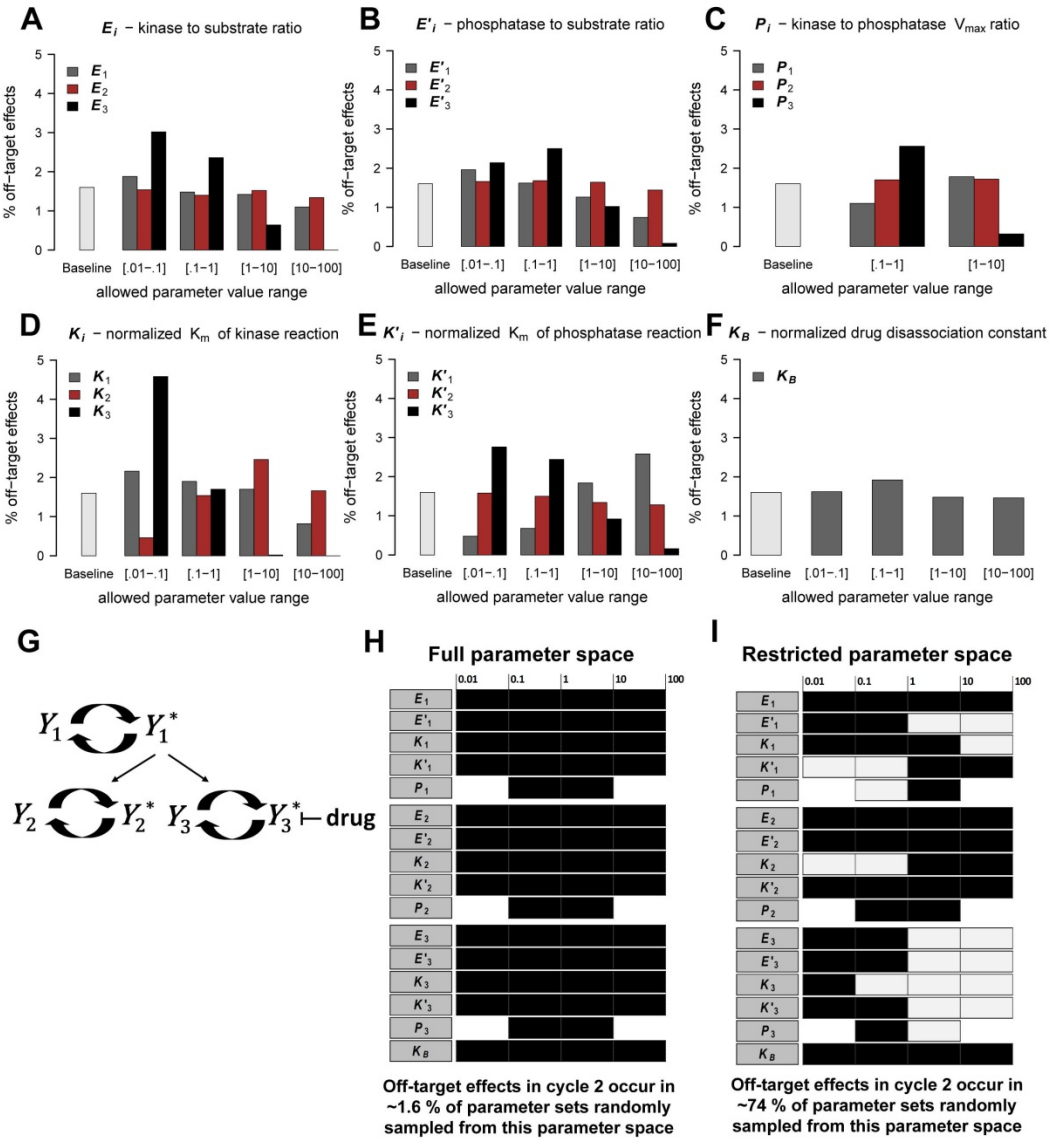


Figure 4.3. A numerical perturbation analysis revealed parameter value ranges that promote off-target effects in the $n = 3$ network. A perturbation analysis of the $n = 3$ network (**G**) was performed where a single parameter's value was randomly selected from a small range of values, while all other parameters were selected from the larger ranges defined in **Table 4.1** The baseline in each plot reflects the percent of off-target effects in cycle 2 in 5000 sampled parameter sets when all parameter values were randomly selected from the ranges defined in **Table 4.1** and depicted in (**H**). All other bars reflect the results of systematically perturbing each parameter (one at a time) using the given sub-ranges (**A**-**F**). Based on this perturbation analysis, a restricted parameter space was generated (**I**) from which ~74% of the sampled parameter sets produced off-target effects in cycle 2. In contrast, only ~1.6% of sampled parameter sets from the full parameter space (**H**) produced off-target effects in cycle 2.

E_3 and E'_3 (the total kinase to substrate and the total phosphatase to substrate ratio, respectively, in cycle 3) also appeared to exert a large degree of control over off-target effects (**Figure 4.3A**-**B**). These results indicate that off-target effects were more likely when the kinase and phosphatase enzymes of

cycle 3 were saturated. $\boldsymbol{P_3}$, the ratio of the V_{max} of the kinase and phosphatase reactions in cycle 3, also affected the percentage of off-target effects (**Figure 4.3C**). When $\boldsymbol{P_3}$ was less than 1, cycle 3 tended toward the deactivation reaction and the percentage of off-target effects increased to 2.56% from 1.6%. Similarly, when $\boldsymbol{P_3}$ was greater than 1, cycle 3 tended toward the activation reaction and the percentage of off-target effects was significantly reduced relative to the baseline (0.32%).

The only parameter associated with cycle 2 that affected the percentage of off-target effects in this network was $\boldsymbol{K_2}$ (the normalized K_m of the kinase reaction of cycle 2). $\boldsymbol{K_2}$ values between 0.01 and 0.10 are expected to produce an efficient kinase reaction because $Y_{2T} >> K_{m_2}$. The results of the numerical perturbation analysis indicated that when $\boldsymbol{K_2}$ was restricted to values in this sub-range, a small percentage of off-target effects was observed (**Figure 4.3D**). In contrast, when $\boldsymbol{K_2}$ was restricted to values between 1.0 and 10.00, the percentage of off-target effects increased relative to the baseline. These results suggest that an off-target effect in cycle 2 is more likely to occur in the $n = 3$ network when the conversion of Y_2 to Y_2* operates in the linear regime because of substrate constraints. This result is somewhat counter intuitive given the fact that we are interested in measuring a response that propagates from cycle 3 to cycle 1 and then down to cycle 2. It is reasonable to expect that an efficient kinase reaction in cycle 2 would be important for recruiting Y_1* to activate Y_2 and generate an effect in cycle 2. If the cycle 2 kinase reaction is less efficient than the cycle 3 kinase reaction, however, more Y_1* will be available to convert Y_3 to Y_3*, ultimately contributing to the sequestration of more Y_3* into a complex with D (**Figure 4.2C**). Such a sequestration may give rise to a detectable upstream effect as a result of the reduced substrate availability in cycle 3.

The cycle 1 parameters with the greatest impact on the percentage of off-target effects were $\boldsymbol{K_1}$ and $\boldsymbol{K'_1}$ (the normalized K_m of the kinase and phosphatase reaction, respectively, in cycle 1) (**Figure 4.3D**-**E**). Larger values of $\boldsymbol{K_1}$ acted to suppress off-target effects, while smaller values produced an increase in off-target effects relative to the baseline (**Figure 4.3D**). The reverse was observed

for K'_1, with higher values producing a higher percentage of off-target effects than smaller values (**Figure 4.3E**). Together, the K_1 and K'_1 results suggest that off-target effects are favored when the cycle 1 phosphatase reaction tends toward inefficiency and the cycle 2 kinase reaction tends towards efficiency. This result is not surprising given that the availability of Y_1^* is essential for the propagation of a signal from cycle 3 to cycle 2.

The value of K_B, the normalized drug disassociation constant, had a very slight effect on the percentage of off-target effects. In general, K_B values greater than 1 produced a slight decrease in the percentage of off-target effects relative to the baseline (**Figure 4.3F**). This result suggests that weaker binding (and larger dissociation constants) promoted fewer off-target effects, as would be expected given the decreased sequestration of Y_3^* that would occur. The change in the percentage of off-target effects induced by restricting K_B values was fairly small compared to the change induced when other model parameter values were restricted (**Figure 4.3**). This result suggests that the activity and efficiency of component cycles in the network may be more important for propagating an off-target effect than the actual kinetics of a targeted therapy.

The results of the above analysis indicated that certain parameter value ranges are more likely to induce an off-target effect in cycle 2 as the drug concentration is increased (**Figure 4.3**). When we restricted the $n = 3$ parameter space by reducing the ranges from which some key parameters were selected (**Figure 4.3I**), the percentage of off-target effects in 5000 randomly sampled parameter sets increased from 1.6% to 73.9%.

A second numerical perturbation analysis was performed using this new restricted $n = 3$ parameter space as a baseline. In general, many of the trends observed in the analysis of the original $n = 3$ parameter space (depicted in **Figure 4.3H**) were observed in the analysis of the restricted $n = 3$ parameter space (**Appendix C**, **Figure C.3**). For example, low K_3 values remained important for producing off-target effects in both parameter spaces. The effects of parameters associated with cycle 2, however, were different in the two parameter spaces. When the original parameter space was tested, K_2 was the only cycle 2

parameter found to substantially affect the percentage of off-target effects (**Figure 4.3D**). In the restricted parameter space, however, some ranges of E_2, E'_2, and K'_2 produced off-target effect percentages that differed substantially from the baseline. For example, E_2 values between 10 and 100 produced off-target effects in 92.1% of sampled parameter sets, the largest percentage of off-target effects observed in any of our analyses (**Figure C.3A**). Because E_2 is the total enzyme to substrate ratio of the kinase reaction (Y_{1T}/Y_{2T}), this result suggests when more total protein exists in cycle 1 compared to cycle 2, off-target effects in cycle 2 are more likely in this network.

While some parameters associated with cycle 2 were able to effect the percentage of off-target effects, the parameters associated with cycle 3 continued to have the greatest effect on off-target effects in the restricted $n = 3$ parameter space. Only parameters in cycle 3 had the ability to effectively eliminate (or substantially reduce) the percentage of off-target effects within specific reduced ranges. Values between 10 and 100 for E_3, E'_3, K_3 and K'_3 produced off-target effect percentages of 0%, 3.24%, 0% and 3.20%, respectively. In addition, P_3 values greater than 1 produced off-target effects in 18.24% of sampled parameter sets which, compared to the baseline of 73.9%, represents a large decrease in off-target effects (**Figure C.3**).

4.3.2 Varying a single parameter can produce a large change in the size of the off-target effect

The magnitude of off-target effects produced by parameter sets randomly sampled from the original $n = 3$ parameter space (**Figure 4.3H**) generally fell between .10 and .30 of the total Y_2 protein pool (**Figure 4.4A**). In contrast, the magnitude of off-target effects produced by parameter sets randomly sampled from the restricted $n = 3$ parameter space (**Figure 4.3I**) were more uniformly distributed across a range of values (**Figure 4.4B**). These results suggest that when conditions in a network are favorable for off-target effects, the size of an off-target effect is highly variable.

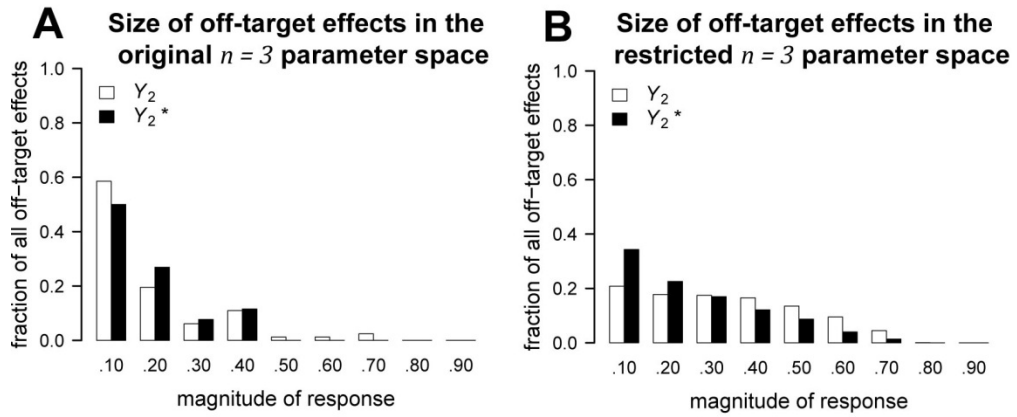


Figure 4.4. Distribution of the size of off-target effects in the *n* = 3 network. Histograms of the size of off-target effects in *n* = 3 network (**Figure 4.3D**) are plotted for two different parameter spaces. The y-axis on each plot represents the proportion of all parameter sets that produced off-target effects in 5000 randomly selected parameter sets. The x-axis on each plot represents the size of an off-target effect in cycle 2 as a proportion of Y_{2T} such that each value indicates a response that was at least as big as the given value but less than the next sequential value. For example, a value of 0.30 indicates that the magnitude of the response was greater than or equal to 0.30 but less than 0.40. (**A**) The majority of off-target effects in the original *n* = 3 parameter space (depicted in **Figure 4.3H**) were less than 0.30. (**B**) In contrast, the distribution of the size of off-target effects in the restricted *n* = 3 parameter space (depicted in **Figure 4.3I**) was more uniform.

We used stimulus response curves to examine how a change in a single parameter value may affect the size of an off-target effect in Y_2* as a function of the normalized inhibitor concentration (**Figure 4.5**). A randomly selected parameter set and a parameter set derived from a *Xenopus* MAPK model [260] (refer **Appendix C** for the derivation of the *Xenopus* parameter values) were used. In each parameter set, either $\boldsymbol{E_2}$ or $\boldsymbol{K_3}$ was varied, while all other parameter values were fixed to the values listed in **Table 4.2**.

The randomly selected parameter set produced a baseline off-target response of 0.19 in Y_2* (**Figure 4.5A**-**B**) and of 0.40 in Y_2 (data not shown). In this parameter set the original $\boldsymbol{E_2}$ value was 32.56 and the original $\boldsymbol{K_3}$ value was 0.04. Increasing $\boldsymbol{E_2}$ to 326.61 substantially decreased the response in Y_2* and decreasing $\boldsymbol{E_2}$ to 3.26 increased the response in Y_2* from 0.19 to 0.27 (**Figure 4.5A**). Similarly, increasing $\boldsymbol{K_3}$ to 0.41 reduced the response in Y_2 * to 0.07 (below the detection threshold) and decreasing $\boldsymbol{K_3}$ to 0.004 increased the response in Y_2* to 0.26 (**Figure 4.5B**).

The parameter set derived from the MAPK *Xenopus* model [260] produced a baseline response of 0.08 (below the detection threshold) in both Y_2* (**Figure**

4.5C-**D**) and Y_2 (data not shown). In this parameter set the original $\boldsymbol{E_2}$ value was 0.0025 and the original $\boldsymbol{K_3}$ value was 0.25. While changing $\boldsymbol{E_2}$ did not alter the response (**Figure 4.5C**), increasing $\boldsymbol{K_3}$ to 2.5 completely eliminated the response in Y_2* and decreasing $\boldsymbol{K_3}$ to 0.025 substantially increased the response in Y_2* to 0.60 (**Figure 4.5D**). These results suggest that when using physiologically realistic parameter values, changing one kinetic parameter or species concentration by an order of magnitude has the capacity to dramatically alter whether a targeted inhibitor induces an off-target effect.

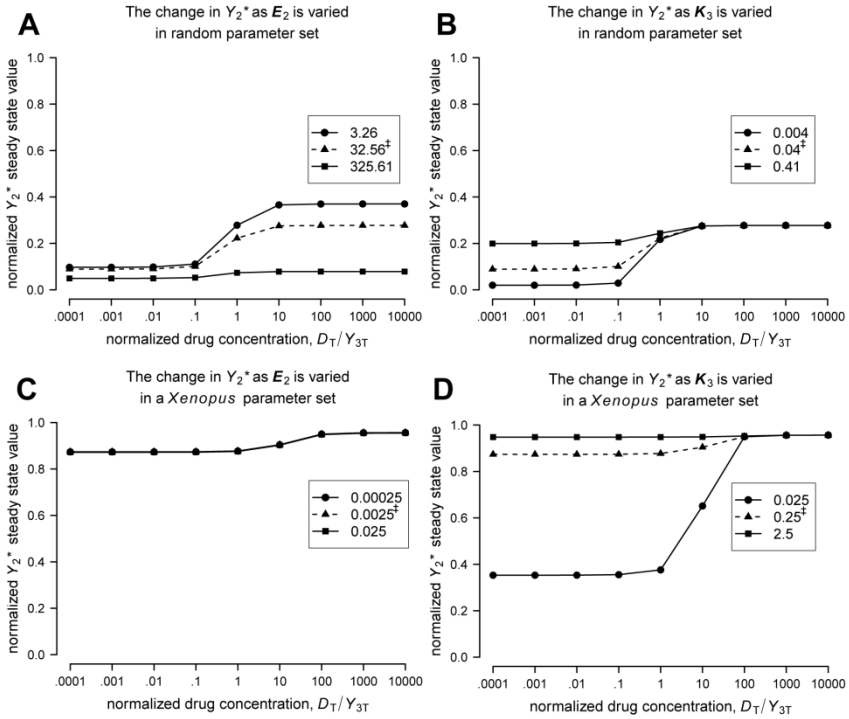


Figure 4.5. Varying a single parameter value can produce a large change in the off-target response. Stimulus response curves were plotted for the $n = 3$ network using a randomly selected parameter set and a parameter set derived from a *Xenopus* model [39] (all parameters values are listed in **Table 4.2**). For each parameter set, $\boldsymbol{E_2}$ and $\boldsymbol{K_3}$ were increased or decreased by 1 order of magnitude and the resulting stimulus response curves were plotted. (**A**-**B**) The random parameter set produced an off-target effect in Y_2 of 0.40 (data not shown) and in Y_2* of 0.19. (**A**) Increasing $\boldsymbol{E_2}$ from 32.56 to 326.61 substantially decreased the off-target effect in Y_2 *, while decreasing $\boldsymbol{E_2}$ to 3.26 increased the off-target effect in Y_2 * to 0.27. (**B**) Increasing $\boldsymbol{K_3}$ from 0.04 to 0.41 reduced the response in Y_2 * below the detection threshold to 0.07, while decreasing $\boldsymbol{K_3}$ to 0.004 increased the off-target response to 0.26. (**C**-**D**) A second parameter set was derived from the literature using MAPK parameters from a *Xenopus* model. This parameter set did not initially produce an off-target effect because the response in both Y_2* and Y_2 was 0.08, which is below the detection threshold. (**C**) Increasing or decreasing $\boldsymbol{E_3}$ to 0.025 or 0.00025, respectively, from 0.0025 had no effect on the response to the targeted inhibitor. (**D**) In contrast, increasing $\boldsymbol{K_3}$ from 0.25 to 2.5 eliminated the original response completely, while decreasing $\boldsymbol{K_3}$ from 0.25 to 0.025 produced a large off-target response of 0.60. Original parameter values prior to variation are indicated by ‡ on the plots (see also **Table 4.2**).

A few of the enzyme to substrate ratios in the *Xenopus* parameter set (E_2 = 0.0025, E'_2 = 0.00025, and E_3 = 0.0025) were outside the limits of parameter ranges allowed in our random parameter space explorations (**Table 4.1** and **Figure 4.3H**), suggesting that off-target effects are possible for a larger range of parameter values than we specifically tested. While we may have been too conservative in the estimation of the ranges defined in **Table 4.1**, this finding supports the position that a targeted inhibitor can naturally induce an off-target effect via retroactivity over a range of physiologically relevant conditions.

Table 4.2. Parameter sets used in stimulus response curves. The two parameters sets used in **Figure 4.5** are summarized in the table. The Random set refers to a randomly selected parameter set and the *Xenopus* set refers to a parameter set derived from a *Xenopus* MAPK model. Bolded values represent the original parameter values varied in **Figure 4.5**.

	Random set	*Xenopus* set
E_1	4.87	0.1
E_2	**32.56**	**0.0025**
E_3	0.28	0.0025
E'_1	0.05	0.1
E'_2	1.26	0.00025
E'_3	0.29	0.1
K_1	5.07	100
K_2	28.18	0.25
K_3	**0.04**	**0.25**
K'_1	66.34	100
K'_2	9.33	0.25
K'_3	0.59	0.25
P_1	0.21	1.00
P_2	3.43	1.00
P_3	0.42	0.025
K_B	0.05	0.0833

Table 4.3. The percentage of off-target effects decreased as the network size increased. The $n = 3$ network's restricted parameter space (depicted in **Figure 4.3I**) produced off-target effects in 73.94% of sampled parameter sets. The $n = 5$ and $n = 7$ vertical networks produced 27.92% and 13.50% off-target effects, respectively, using the same parameter space. In the lateral case, the drop was more dramatic with the $n = 5$ and $n = 7$ networks producing 6% and 0% off-target effects, respectively. Parameter values used in cycles 4, 5, 6 or 7 were homogenous with cycle 3. All percentages are out of 5000 randomly selected parameter sets using the parameter space depicted in **Figure 4.3I**.

	Cycle 2	
n	***% Off Target Effects***	
3	73.9	
	vertical	*lateral*
5	27.9	6.0
7	13.5	0.0

Table 4.4. Off-target effects can amplify downstream of cycle 2. To test for downstream propagation of off-target effects from cycle 2, we created an *extended* $n = 3$ network by adding a 4th cycle activated by Y_2^*. If an off-target effect occurred in cycle 4 and the size of the response in cycle 4 exceeded the size of the response in cycle 2, then an off-target effect with amplification was reported for cycle 4. First, we looked at off-target effects in cycle 4 using the $n = 3$ restricted parameter space (**Figure 4.3I**) for cycles 1 – 3 and the default parameter ranges from **Table 4.1** for cycle 4 (this parameter space is depicted in Appendix C, **Figure C.4H**). Next, we looked at off-target effects in a similar parameter space but with cycles 2 – 4 restricted to ranges found to favor propagation of off-target effects from cycle 2 to cycle 4 (this parameter space is depicted in **Figure C.4I**). Values listed in the table are percentages out of 5000 randomly selected parameter sets.

	Cycle 2	*Cycle 4*
extended n = 3 network with cycle 4 default ranges		
% Off Target Effects	**75.3**	**35.5**
% Off Target Effects with Amplification	--	23.4
% Off Target Effects without Amplification	--	12.1
extended n = 3 network with cycle 4 restricted ranges		
% Off Target Effects	**45.3**	**67.4**
% Off Target Effects with Amplification	--	61.9
% Off Target Effects without Amplification	--	5.5

4.3.3 The percentage of off-target effects decreased as the size of the vertical and lateral networks increased

We next investigated networks with more than 3 cycles by randomly exploring the parameter spaces of the vertical (**Figure 4.2A**) and lateral (**Figure 4.2B**) cases using $n = 5$ and $n = 7$ cycles. As before, we measured the steady state change in cycle 2 as the normalized concentration of the drug that

targeted cycle n was increased. The restricted parameter space depicted in **Figure 4.3I** (from which 73.9% of sampled parameter sets produced off-target effects in cycle 2) was used for this analysis. Networks were analyzed using homogenous parameter values in cycles 4, 5, 6 and 7 that equaled the corresponding parameter values randomly selected for cycle 3 (e.g., in the $n = 5$ case, $\boldsymbol{E_3} = \boldsymbol{E_4} = \boldsymbol{E_5}$). This allowed us to keep the size of the parameter space fixed so that 5000 parameter sets remained a reasonable number to sample from each network's parameter space.

In the vertical case, the percentage of off-target effects in the $n = 5$ and $n = 7$ networks were 27.92% and 13.50%, respectively (**Table 4.3**). The reduced probability of off-target effects as the cascade lengthened suggests that applying a targeted inhibitor near the bottom of a long cascade can produce a detectable off-target response but the signal may attenuate as it travels up the cascade. This conclusion is in agreement with a recent work that investigated retroactivity in long signaling cascades [253] and found retroactive signals are likely to attenuate as they travel up long cascades.

In the lateral case, the drop in the percentage of off-target effects was more dramatic than in the vertical case, with the $n = 5$ and $n = 7$ networks producing 6% and 0% off-target effects, respectively (**Table 4.3**). This result suggests applying a targeted inhibitor to a cycle that is activated by a signaling molecule involved in the simultaneous activation of many other cycles decreases the likelihood of off-target effects. This conclusion is based on a limited exploration of the parameter space (due to the homogenous parameter selection used for cycles 3 and greater) but is in agreement with a model proposed by Kim et al. [252] that showed retroactivity (or what they referred to as subsrate-dependent control) is attenuated by the number of substrates available.

4.3.4 Off-target effects from retroactivity can propagate down a non-targeted cascade

Our results suggest that, under appropriate conditions, it is possible for a downstream perturbation from a targeted inhibitor to transmit up a cascade resulting in a detectable off-target effect near the top of another cascade.

Because signal amplification is an important cellular sensory mechanism [261], we next investigated whether upstream off-target effects from targeted inhibitors are likely to amplify down a non-targeted cascade.

To test for downstream propagation of off-target effects from cycle 2, we created an $extended\ n = 3$ network by adding a 4th cycle activated by Y_2* (**Figure 4.2D**). If a change in the steady state concentration of Y_4 and/or Y_4* occurred that was at least 0.10 of the total Y_4 protein pool, then an off-target effect was considered to have occurred in cycle 4. If an off-target effect occurred in cycle 4 and the size of the response in cycle 4 exceeded the size of the response in cycle 2, then an off-target effect with amplification was considered to have occurred in cycle 4.

When the default parameter ranges defined in **Table 4.1** were used for all cycles in the $extended\ n = 3$ network, the percentage of off-targets in cycle 2 and cycle 4, respectively, was 1.78% and 0.03%. We next tested the $extended\ n = 3$ network using the restricted $n = 3$ parameter space (**Figure 4.3I**) for cycles 1 – 3 and the default parameter ranges from **Table 4.1** for cycle 4. In this partially restricted parameter space (depicted in Appendix C, **Figure C.4H**), the percentage of off-target effects in cycle 2 and cycle 4 increased to 75.3% and 35.5%, respectively, and amplification contributed to cycle 4 off-target effects in 23.3% of the sampled parameter sets (representing more than half of the off-target effects in the sampled parameter sets). The remaining off-target effects in cycle 4 occurred in 12.2% of the sampled parameter and had a response size that was either attenuated relative to cycle 2 or equal to the cycle 2 response (**Table 4.4**).

To identify the parameters that were most important for amplifying an off-target effect from cycle 2 to cycle 4 in the $extended\ n = 3$ network, we performed a numeric perturbation analysis (as previously described) on the partially restricted parameter space depicted in **Appendix C**, **Figure C.4H**. From these results, we generated a new parameter space (Appendix C, **Figure C.4I**) which produced off-target effects of 45.3% and 67.4% in cycle 2 and cycle 4, respectively. Amplification contributed to cycle 4 off-target effects in 61.9% of the sampled

parameter sets. The remaining off-target effects in cycle 4 occurred in 5.5% of the sampled parameter sets and had a response size that was either attenuated relative to cycle 2 or equal to the cycle 2 response (**Table 4.4**).

4.4 Discussion

We developed a computational model to test whether targeted therapies, such as kinase inhibitors, can produce off-target effects in upstream pathways as a consequence of retroactivity alone. Using a numeric perturbation method, we identified specific conditions (**Figure 4.6**) for the promotion of steady state off-target effects via retroactivity when a targeted inhibitor was applied to cycle n in a series of simple signaling networks (**Figure 4.2**).

Our investigation considered only the effect of retroactivity and targeted inhibitors on the individual motifs we studied in the absence of genetic and/or other regulatory relationships. This allowed us to investigate whether such motifs have the capacity to produce off-target effects without regulatory feedback connections. In addition, the present study only considered the steady state response to a targeted therapy. The primary reason we considered only steady state responses was because it provided us with an objective measure that could be used to compare the effect of a targeted inhibitor across many different parameter sets. It is important to note that the dynamics of a retroactive signaling process are likely to induce transient changes in the levels of key signaling molecules. These transient changes, which are not observable at steady state, may lead to important *in vivo* responses.

It is also well known that the dynamics of signal transduction networks can be modulated by important oscillatory behavior, for example, from the P53/MDM2 regulatory feedback loop [45, 137]. Because we have not considered transient dynamics, our approach cannot be assumed to apply to all signaling networks. Nevertheless, we expect conditions that favor the induction of off-target effects at steady state to also favor the induction of detectable transient changes associated with the steady state response. In fact, this is what we observed when we plotted the time-course of the cycle 2 proteins with the parameter sets used in **Figure 4.5** (data not shown).

This work has led to very interesting and somewhat surprising results. A major importance of this work is that it did not investigate off-target effects related to a specific therapeutic intervention and, instead, investigated whether off-target effects from retroactivity are possible in a set of common network motifs, making the work generalizable to a number of signaling pathways. There are, however, examples of targeted inhibitors of great clinical interest that are involved in signaling motifs similar to the network motifs we examined. The drug NSC 74859 [262], for example, is a selective inhibitor that targets STAT3. JAK is an upstream activator of both STAT3 and PI3K [263], thus when NSC 74859 inhibits STAT3, JAK could potentially facilitate the propagation off-target effects due to retroactive signaling from STAT3 to PI3K. Moreover, the inhibitor GSK690693 [264] targets AKT and could potentially give rise to a retroactive signal that propagates upstream to a common activator of either the MAPK or STAT3 cascades, generating off-target effects in these pathways.

The binding affinity of the inhibitor for its target did not play a substantial role in the promotion of off-target effects in our model. Instead, the kinetics of the component cycles in the network were more important for increasing the likelihood of off-target effects (**Figure 4.3, Figure C.3**, and **Figure C.4**). In general, off-target effects were more likely to occur in the networks studied when the targeted cycle n favored the deactivation reaction because the V_{max} of the deactivation reaction was larger than the V_{max} of the activation reaction and/or both enzymatic reactions in cycle n operated in or near the zero order regime. Off-target effects were also more likely when cycle 1 (the source of the shared activator in our models) favored the activation reaction and its kinase reaction operated in or near the zero order regime.

If cycle 2's cascade was extended to include cycle 4 (**Figure 4.2D**), which was activated by Y_2^*, off-target effects were more likely to propagate to cycle 4 when cycle 2 favored deactivation and cycle 4 favored activation. In cycle 2 this meant that the kinetics of the kinase reaction were generally inefficient and operating in or near the linear regime and that the V_{max} of the deactivation reaction was generally larger than the V_{max} of the activation reaction. Thus, off-

target effects were promoted when cycle 2 was “off” and not consuming significant amounts of the shared upstream activator, Y_1^*.

The results also indicate that off-target effects were more likely when the total kinase to substrate and the total phosphatase to substrate ratios in the inhibited cycle ($\boldsymbol{E_n}$ and $\boldsymbol{E'_n}$, respectively) were less than 1. In the $n = 3$ network, this meant that there was less total protein in cycle 1 than in cycle 3 because $\boldsymbol{E_3} < 1$ implies $Y_{1T} < Y_{3T}$. The reason for this is that the smaller the Y_{1T}/Y_{3T} ratio, the stronger the sequestration of Y_1^* will be. The impact of this ratio increases if cycle 3 favors the deactivation reaction such that a large fraction of Y_{3T} is in the inactive Y_3 form, promoting the binding of Y_3 to Y_1^*.

The immediate experimental implications of this result is that, in the absence of kinetic information, the likelihood of off-target effects may potentially be estimated for a network configuration of this type (**Figure 4.2A**,**C**-**D**) based on the ratio of the concentrations of components in the inhibited cycle and the preceding cycle (using, for example, proteomic or gene expression data). While this ratio would not be an absolute predictor, the presence of this condition would suggest an increased probability of off-target effects.

In agreement with the work of other groups [252, 253], we found the probability of off-target effects attenuated when the targeted cycle was near the bottom of a long cascade or when there were many substrates competing for a common upstream activator (**Table 4.3**). Our results also suggest that within physiologically realistic parameter ranges, changing a single kinetic parameter or species concentration by 1 order of magnitude has the capacity to dramatically alter whether an off-target effect occurs as a direct result of targeted inhibition. It is also worth noting that, even though we varied the normalized drug concentration over a very large range, in general, the normalized inhibitor concentration needed to change by only 2 orders of magnitude to induce an off-target effect (see, for example, **Figure 4.5**).

4.4.1 Conclusions

Off-target drug effects *in vivo* are typically attributed to cross-talk arising from a feedback connection in a signaling network or to non-specific interactions

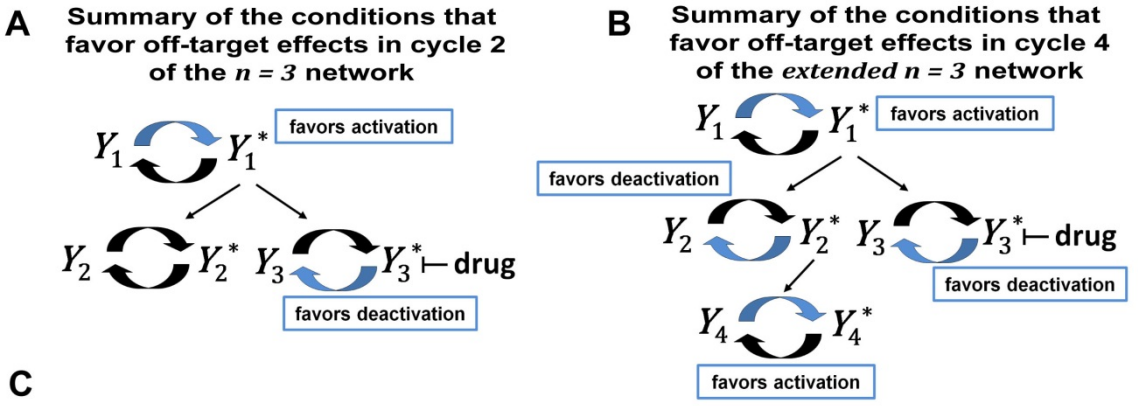
with other proteins. In this work we have demonstrated that off-target drug effects can also arise naturally from retroactivity in a covalently modified signaling network. This view of signaling challenges the widespread notion that information in signaling cascades only flows from the cell surface to the nucleus and, consequently, this work has far reaching implications for targeted cancer therapies.

A crucial finding of this work is that the kinetics governing the covalently modified cycles in a signaling network are likely to be far more important for propagating an off-target effect due to retroactive signaling than the binding affinity of the drug for the targeted protein, which is a commonly optimized property in drug development. Another particularly paramount finding is that an off-target effect due to retroactive signaling is more likely when the first cycle in a non-inhibited cascade is “off” and essentially inactive. This suggests that, in the motifs we studied, a targeted therapy has the capacity to turn “on” an otherwise “off” tributary cascade.

To emphasize, it is entirely probable for a branch of a signaling network that is “off” to become activated or “on” due to the inhibition of another protein in the network based on retroactivity alone, suggesting an inherent opportunity for negative therapeutic effects. Our findings, therefore, have implications for somatic evolution in cancer and the onset of therapeutic resistance, which has been widely reported for many targeted cancer therapeutics [265], most notably for the targeted inhibition of BCR-ABL by imatinib [266]. Moreover, a single mutation could conceivably give rise to a spontaneous off-target effect without the need for any direct regulatory connections between the targeted protein and the effected protein.

While our approach does not definitively establish that the predicted responses will occur *in vivo*, our results demonstrate that off-target effects are indeed possible in the absence of direct regulatory relationships and suggest that additional (and more specific) experimental and theoretical investigations are warranted. A proper characterization of a pathway’s structure is important for identifying the optimal protein to target as well as what concentration of the

targeted therapy is required to modulate the pathway in a safe and effective manner. We believe our results strongly support the position that such characterizations should consider retroactivity as a potential source of off-target effects induced by kinase inhibitors and other targeted therapies. This work has also provided an initial roadmap for how to assess the likelihood of off-target effects in a signaling network.



C

Cycle	Both networks
1	V_{max} kinase > V_{max} phosphatase Kinase reaction in zero order regime Phosphatase reaction in linear regime Phosphatase saturated
3	V_{max} kinase < V_{max} phosphatase Kinase reaction in zero order regime Phosphatase reaction in zero order regime Enzymes saturated

Cycle	*n = 3* network	*extended n = 3* network
2	Kinase reaction in linear regime Enzymes *not* saturated	V_{max} kinase < V_{max} phosphatase Kinase reaction in linear regime Enzymes saturated
4	n/a	V_{max} kinase > V_{max} phosphatase Kinase reaction in zero order regime Phosphatase reaction in linear regime

Figure 4.6. A summary of conditions that favor off-target in the *n = 3* and the *extended n = 3* networks. The conditions that promoted off-target effects in our model are summarized for two network types. Off-target effects in (**A**) the $n = 3$ network and (**B**) the *extended* $n = 3$ network were favored when cycle 3 tended toward the deactivation reaction and cycle 1 tended toward the activation reaction. Off-target effects were favored in cycle 4 of (**B**) the *extended* $n = 3$ network when cycle 2 tended toward the deactivation reaction and cycle 4 tended toward the activation reaction. Blue arrows in a cycle indicate which V_{max} is larger when off-target effects are favored. (**C**) A summary of the specific conditions in each cycle found to favor off-target effects in the *n = 3* network, in the *extended n = 3* network, or in both networks.

Chapter 5
Variations in carbon flux in distinct breast cancer phenotypes

5.1 Background

Otto Warburg was the first to observe that cancer cells, unlike normal quiescent cells, rely on glycolysis for energy production and produce large amounts of lactate relative to surrounding tissue, even when sufficient oxygen is present [3]. This form of metabolism, now known as the Warburg effect or aerobic glycolysis [4-7], generates only 2 molecules of ATP per molecule of glucose on average, compared to approximately 36 molecules of ATP when glucose is fully oxidized via mitochondrial oxidative phosphorylation. The Warburg effect is so pervasive across cancers that it is routinely leveraged in the clinic via fluorodeoxyglucose-positron emission tomography (FDG-PET) scans, where high glucose uptake correlates with tumor cell proliferation and poor prognostic outcome [5, 8]. There is evidence that cancer cells divert some glucose derived carbon to nucleotide and fatty acid synthesis [5], suggesting that the dependence on glycolysis exhibited by cancer may help sustain the high biosynthetic requirements of proliferation. In addition to glucose, glutamine is also an important metabolic substrate consumed by cancer cells (**Figure 5.1**) [5].

The Warburg effect in cancer is similar to proliferative metabolism in normal cells but some molecular and kinetic differences have been discovered. *I hypothesize that there are key drivers of the Warburg effect that can be exploited to target cancer proliferation without substantial negative effects on normal tissue growth*. As a consequence of the highly complex and interconnected nature of

intracellular networks, identifying these drivers requires systems biology approaches. The objective of the work presented in this chapter is to identify some of the system level metabolic differences that occur at distinct stages of breast cancer and to provide a framework for identifying the key drivers of the Warburg effect. The four breast cell lines used in this study represent normal-like breast tissue (MCF-10A), a primary tumor (MCF-7), and metastatic breast cancer (MDA-MB-231 and SUM-149) (**Figure 1.1D-G**).

Understanding which mechanisms give rise to the Warburg effect is an active area of research. Most current approaches focus on the elucidation of gene and protein differences between normal and malignant metabolism. Another increasingly important approach for the study of cancer metabolism is the use of stable isotopes, such as ^{13}C [267]. The primary goal of isotope experiments is to quantify the enrichment in a molecule resulting from the addition of an isotopically labeled tracer. Stable isotopes are not radioactive and differ in the number of neutrons in the nucleus, which alters the mass but not the chemical behavior of the atom. The use of stable isotopes in metabolic research, predates the use of radioactive isotopes [268].

Metabolic fluxes, which are time-dependent reaction rates involved in cellular metabolism, cannot be directly observed but can be inferred from isotope experiments. Relative fluxes between pathways can be inferred from mass isotopomer analysis, while absolute fluxes can be determined using metabolic flux analysis [267, 269]. Differences in molecular weight that occur as a consequence of enrichment from an isotopically labeled tracer are usually represented as a mass isotopomer distribution (MID). A MID of a metabolite, which is measured with mass spectrometry, represents the relative abundance of isotopomers that differ only by one or more atomic mass units [269]. The MID of a 6 carbon metabolite exposed to ^{13}C will have up to 7 mass isotopomers: M0, M1, M2, M3, M4, M5, and M6, where the number following the M indicates the atomic mass shift from the base mass of M0. It is important to note that a mass isotopomer does not provide positional information about which atoms are labeled. For example, the M2 abundance of a metabolite labeled with ^{13}C

reflects only the relative abundance of all isotopomers where ^{13}C is found in any two positions in the carbon backbone.

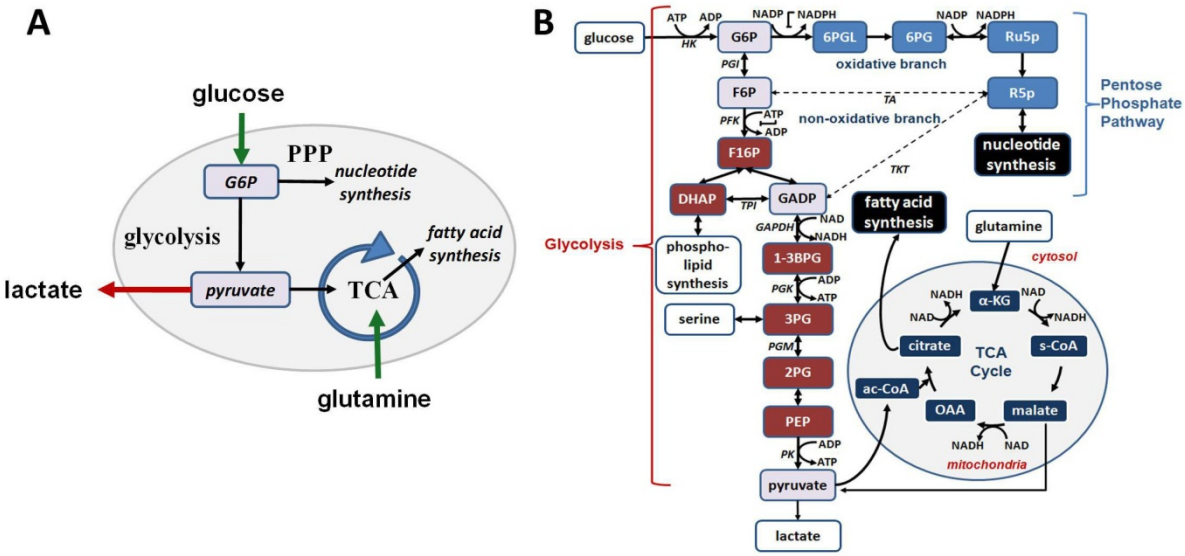


Figure 5.1. Central carbon metabolism. (A) Central carbon metabolism involves glycolysis, the pentose phosphate pathway (PPP), and the TCA cycle, which are linked by shared intermediates and play critical roles in biosynthesis, proliferation, as well as the production of energy and reducing equivalents. Glucose and glutamine are taken up from the extracellular environment and lactate, the primary product of glycolysis, is exported from the cell. **(B)** A schematic summarizing some key enzymatic reactions involved in central carbon metabolism.

5.2 Methods

Isotope Experiments

For isotope labeling experiments, cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing uniformly labeled glutamine. Some cells also received positionally labeled glucose. Two forms of labeled media were prepared and are identified as M* and M** to indicate the number of labeled substrates. The M* medium contained 2.5 mM [U^{13}C] glutamine and 10 mM unlabeled glucose. The M** medium contained 2.5 mM [U^{13}C] glutamine and 10 mM [1,2^{13}C] glucose. Variations of each labeled media were prepared to support the normal growth conditions of each cell line. For MCF10A cells, the media was supplemented with 5% horse serum, 1% insulin and 0.05% hydrocortisone. For MCF7 and MDA-MB-231 cells, the media was supplemented with 5% fetal bovine serum (FBS). For SUM-149 cells, the media was supplemented with 5%

FBS, 0.5% insulin, and 0.1% hydrocortisone. SUM-149 cells were grown in 10% CO_2 and the remaining cells were grown in 5% CO_2. For each cell line specific media formulation, the M* and M** form of the media were chemically identical.

Each cell line was grown to approximately 60 – 80% confluence in twelve 6-cm dishes in unlabeled media. For 1 h, cells were incubated in fresh media chemically identical (except for the tracer) to the labeled media to allow cells to stabilize from effects related to metabolic waste in the spent media. After 1 h, four of the plates from each cell line were changed to 5 mL M* media and four of the plates from each cell line were changed to 5 mL M** media. The remaining four plates from each cell line were saved for total protein analysis. The cells were incubated for 5 h to achieve isotopic steady state. A total of nine 6-cm plates of media only with no cells were simultaneously incubated as controls for 5 h (three for each cell line media type). After 5 h, all media were collected for extracellular flux determination. Cells were washed and quenched with liquid nitrogen and immediately stored at -80 C for subsequent LC-MS analysis.

Measuring Extracellular flux

The concentration of glucose, lactate, glutamine, and glutamate were measured in fresh media and spent media (after 5 h of incubation) with a YSI 2900D. The rate of glucose and glutamine uptake and the rate of lactate production in each cell line were estimated from these data. Metabolite concentrations in the 6 cm dishes that contained only media were used to correct for evaporation as well as degradation of glutamine, both of which were observed in pilot studies.

In the 5 h control experiment, the effect of glutamine degradation was not found to be significant, but an increase in both glucose and glutamine concentration was observed in the control media (**Figure 5.10A-B**). The lactate control media did not show a change between the 0 h and 5 h control concentrations, most likely because the starting lactate concentration was too low to be measurably effected by evaporative loss after only 5 h. It is worth noting that the 0 h starting concentration of both glucose and lactate were measured higher than expected. This difference was confirmed to be a

consequence of glucose and lactate in horse serum and fetal bovine serum (data not shown).

Computational Simulation

For metabolic flux analysis, OpenFlux [270] was used, which is an open source implementation of the elementary metabolite unit framework [271]. Confidence intervals were determined based on the non-linear method developed by Antoniewicz et al. for MFA [272] .

5.3 Results

Each of the four cell lines were grown in ^{13}C labeled media for 5 h to achieve isotopic steady state (**Figure 5.2)**. All cells were grown in one of two forms of labeled media containing 2.5 mM glutamine and 10 mM glucose. The first contained [U^{13}C] glutamine, and the second contained both [U^{13}C] glutamine and ([1,2^{13}C]) glucose. This first medium is represented as M***** to indicate the presence of one labeled substrate, and the second medium is represented as M** to indicate the presence of two labeled substrates. This tracer combination was chosen because previous studies have shown that [U^{13}C] glutamine is best for resolving flux within the TCA cycle and [1,2^{13}C] glucose is the optimal glucose tracer for elucidating flux in glycolysis and the pentose phosphate pathway (PPP) [273].

Most studies of mammalian cells indicate that isotopic steady state of central carbon metabolism occurs within 3 - 6 h [273-275]. A pilot study was performed to identify when isotopic steady state was reached in the cells used in this work. In agreement with other studies, glycolytic intermediates were near isotopic steady state in approximately 15 min, while TCA cycle intermediates took between 4 and 5 h to reach isotopic steady state (**Figure 5.2**).

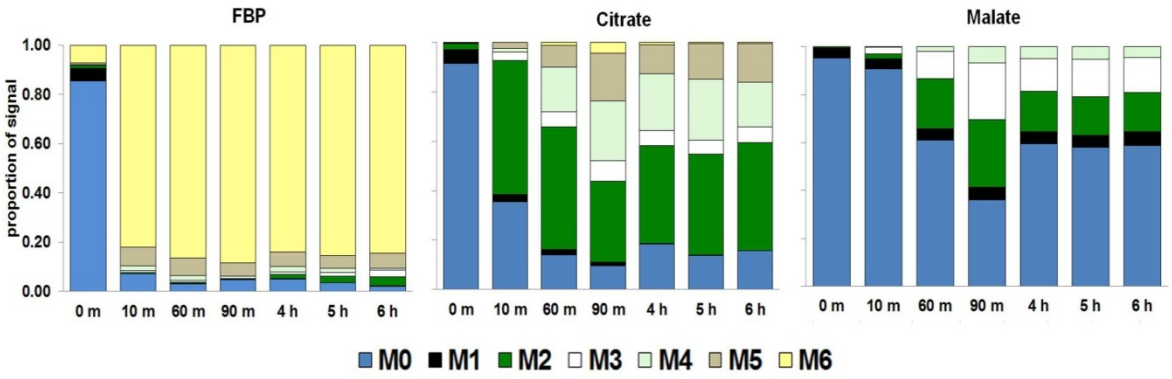


Figure 5.2. Evidence of isotopic steady state. A time course of ^{13}C enrichment after exposing cells to [U^{13}C] glucose for 6 h is presented for 3 metabolites. Bars represent the mass isotopomer distribution (MID) of fructose-1,6-bisphosphate (FBP), citrate and malate measured at each time point. Representative plots from MDA-MB-231 cells are shown. Isotopic steady state assumes a quasi-metabolic steady state where the flux in and out of metabolite pools is balanced, generating unchanging isotope distribution patterns.

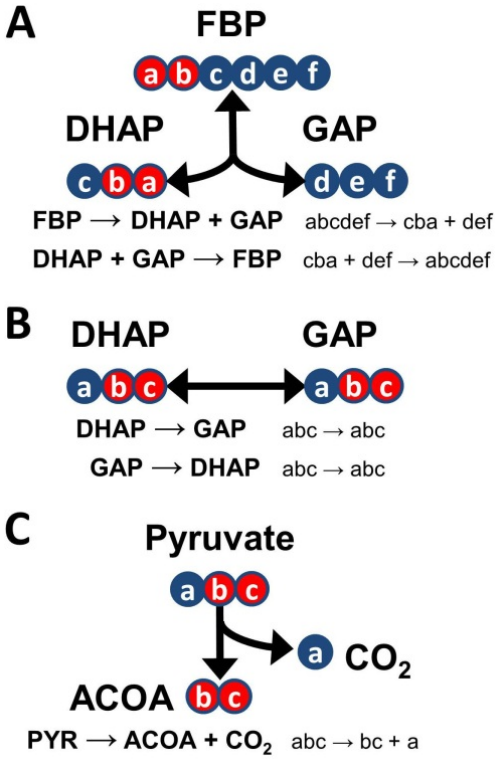


Figure 5.3. Carbon atom transitions. When [1,2^{13}C] glucose is used as a tracer substrate, the conversion of (**A**) FBP to DHAP and GAP results in labeling of the 2nd and 3rd carbon positions of DHAP and no labeling in GAP. (**B**) If the isomerization reaction between DHAP and GAP proceeds, then the 2nd and 3rd carbons of GAP become labeled. Thus, significant enrichment of ^{13}C in GAP is only possible when DHAP isomerizes to GAP. If a significant portion of DHAP is diverted to phospholipid synthesis, however, limited GAP labeling will result. (**B**) GAP labeled on the 2nd and 3rd carbons will generate pyruvate labeled on the 2nd and 3rd carbon and, ultimately, fully labeled Acetyl-CoA (ACOA)**.** The carbons of CoA have been neglected because only the two carbons in the acetyl group of ACOA participate directly in central carbon metabolism. Red circles indicate ^{13}C and blue circles indicates ^{12}C.

5.3.1 Natural abundance of stable isotopes

^{13}C is a stable isotope that naturally occurs at a non-negligible abundance of approximately 1.11% [276]. To accurately quantify enrichment for a ^{13}C based tracer, the MIDs of each metabolite measured after exposure to ^{13}C must be corrected for natural abundance of carbon as well as other atoms. The probability that any given carbon atom position is naturally labeled by ^{13}C follows a binomial distribution. As a result, simply subtracting 1.11% of the mass from each isotopomer in a MID will neglect the non-linear labeling distribution inherent to natural ^{13}C abundance and can lead to an underestimation of the experimental enrichment of ^{13}C at higher masses [276, 277]. While it is especially important to correct MIDs measured by gas chromatography mass spectrometry (GC-MS) because the derivatizing agents generate fragments with very high masses, it is also necessary to correct MIDs measured by liquid chromatography mass spectrometry (LC-MS) before performing additional analysis or applying parameter estimation algorithms. This work relied on the method developed by Fernandez et al. [276] for natural isotope correction to account for the non-linearity inherent in natural isotope abundance. A program written in Matlab was used to generate a set of corrected MIDs that was then used as input to the MFA algorithm.

5.3.2 Mass isotopomer analysis

Acetyl-CoA M2 enrichment

As depicted in **Figure 5.3**, when [1,2^{13}C] glucose is used as a tracer, acetyl-CoA (ACOA) is expected to exhibit ^{13}C enrichment in the M2 isotopomer. When cells were grown in the M* medium lacking labeled glucose, the normal-like MCF10A and the IBC SUM-149 were the only cells to exhibit M2 label in ACOA (**Figure 5.4**). In contrast, when cells were grown in M** medium, all four cell lines exhibited M2 enrichment in ACOA. These data suggest that all ACOA flux is glucose derived in MCF-7 and MDA-MB-231 cells and that some ACOA flux is glutamine derived in MCF10A and SUM-149 cells.

Very little labeled ACOA was found across all the cell lines as a percentage of the total ACOA pool. When [1,2^{13}C] glucose (which was used in the M** media) is the tracer, ACOA is expected to become labeled when DHAP isomerizes to GAP. Thus, it is reasonable to expect approximately half of ACOA to be labeled under these conditions in normal cells with active isomerization from DHAP to GAP. SUM-149 cells exhibited the most ACOA labeling at approximately 20% of the total pool, and most of this label appeared to be glutamine derived. In contrast, in a pilot study performed on MDA-MB-231 cells with [U^{13}C] glucose, approximately 90% of ACOA was labeled within 5 h (data not shown). The most likely explanation for the lack of significant label in the ACOA pool is that a significant flux to GAP (from intermediates of the non-oxidative branch of the PPP) takes place in all the cells.

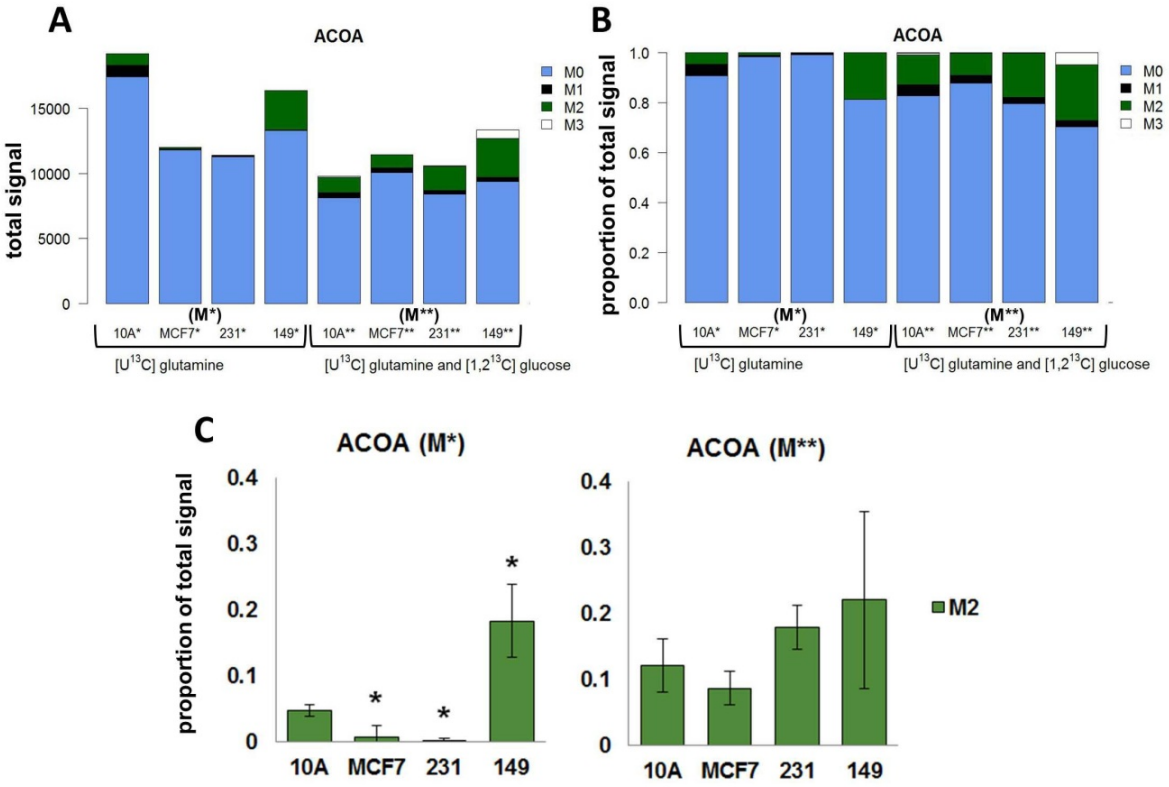


Figure 5.4. Enrichment of ^{13}C in M2 Acetyl-CoA. (**A**) MIDs of Acetyl-CoA (ACOA) produced by labeling with [U^{13}C] glutamine (M*) or [U^{13}C] glutamine and [1,2^{13}C] glucose (M**), normalized by total protein. Data represent the average of 4 experimental repeats. (**B**) The same MIDs were normalized to 1.0 to reveal the proportion of enrichment of each isotopomer. (**C**) Comparison of M2 enrichment in ACOA after M* and M** labeling. Data represent the average +/- the standard deviation of 4 experimental replicates. M2 fractions statistically different from MCF-10A are indicated by ***** on each plot ($p < 0.05$). All data were corrected for natural abundance of stable isotopes.

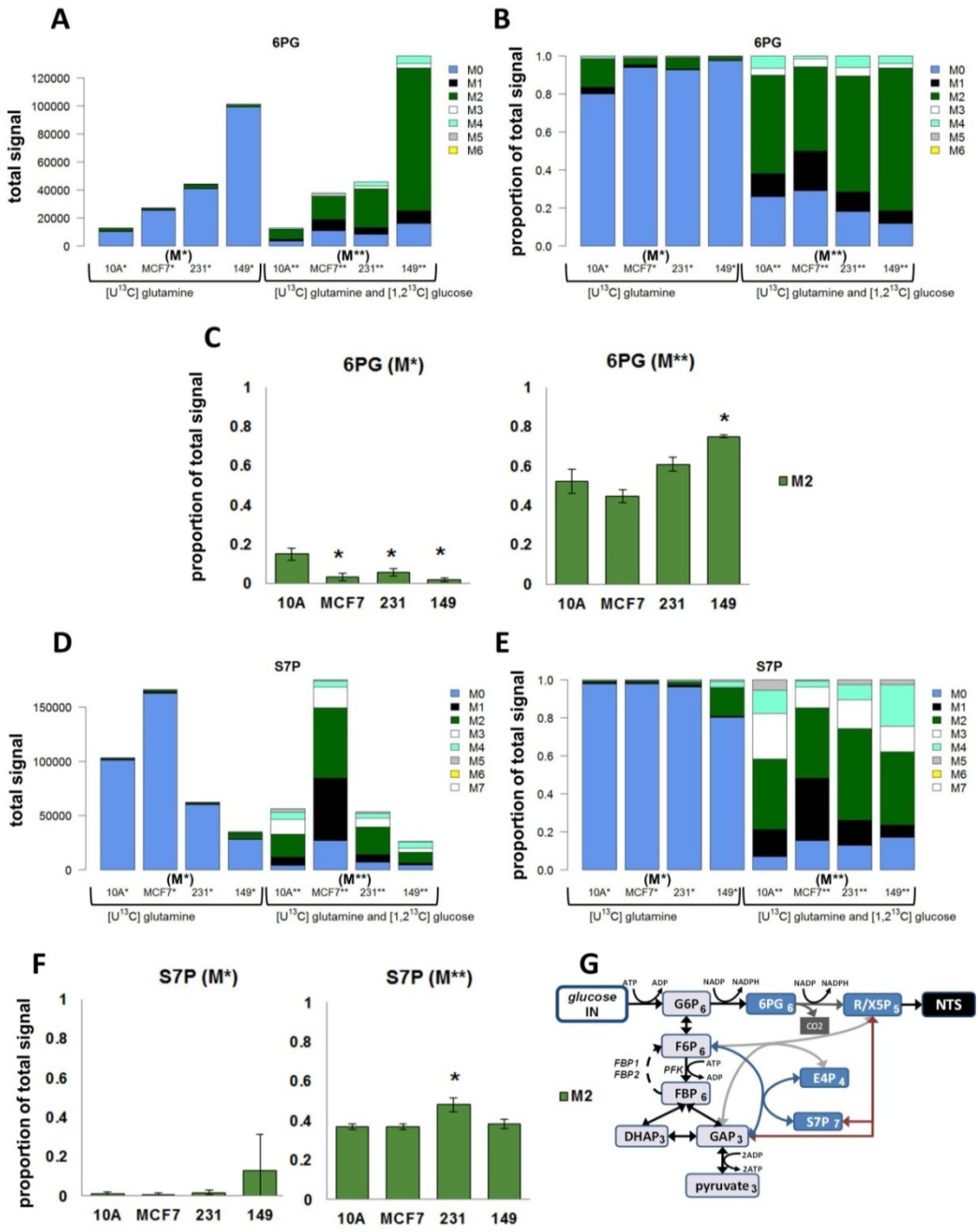


Figure 5.5. Evidence of oxidative and non-oxidative flux in the pentose phosphate pathway. (**A**) MIDs of 6-phosphogluconate (6PG), an intermediate of the oxidative pentose phosphate pathway (PPP) branch, produced by labeling M* or M** media were normalized by total protein. (**B**) The same MIDs were normalized to 1.0 to reveal the proportion of enrichment of each isotopomer. (**C**) Comparison of M2 enrichment in 6PG after M* and M** labeling. (**D**) MIDs of sedoheptulose-7-phosphate (S7P), an intermediate of the non-oxidative PPP branch, produced by labeling with M* or M**, were normalized by total protein. (**E**) The same MIDs were normalized to 1.0 to reveal the proportion of enrichment of each isotopomer. (**F**) Comparison of M4 enrichment in S7P after M* and M** labeling. (**G**) Simplified schematic of the PPP pathway (blue nodes). Ribose generated in the oxidative branch can be used for nucleotide synthesis (NTS). All data represent the average of 4 experimental repeats and error bars reflect the standard deviation. M2 fractions statistically different MCF10A are indicated by * on each plot ($p < 0.05$). All data were corrected for natural abundance of stable isotopes.

Pentose phosphate flux

Cells utilize the pentose phosphate pathway (PPP) to generate reducing power and ribose for nucleotide biosynthesis. The PPP consists of oxidative and non-oxidative branches that operate in distinct modes, depending on the cell's need for reducing power, ATP, and ribose [278]. The oxidative branch irreversibly converts glucose 6-phosphate (G6P) to ribulose 5-phosphate (R5P) and generates 2 molecules of NADPH, while the non-oxidative branch involves the reversible conversion of R5P to fructose 6-phosphate (F6P) or glyceraldehyde 3-phosphate (GAP) (**Figure 5.1B and Figure 5.5G**) When the cell's need for reducing power is very high, reverse glycolytic flux is possible from fructose 1,6-bisphosphate (FBP) to G6P and from G6P back into the oxidative branch of the PPP. The glycolytic reaction that converts F6P to FBP is catalyzed by phosphofructokinase 1 (PFK1) and is effectively irreversible. As a consequence, it is considered the first committed step of glycolysis. The gluconeogenic enzyme fructose 1,6-bisphosphatase can catalyze the reverse reaction of FBP to F6P, however. If a cell's need for reducing power in the form of NADPH exceeds its need for ATP production, glycolytic flux can drive PPP cycling from the non-oxidative back to the oxidative branch of the PPP [278], presumably via this enzyme.

6-phosphogluconate (GPG) is an intermediate of the oxidative branch of the PPP whose 6 carbon backbone comes from G6P (**Figure 5.5G**). In M** labeled samples with [1,2 ^{13}C] glucose, at least two thirds of the 6PG pool was labeled in all four cell lines, suggesting that there is significant oxidative PPP flux in these cells. Surprisingly, when M* samples lacking labeled glucose were analyzed, MCF10A cells exhibited an M2 6PG pool of almost 20%, suggesting that some glutamine derived carbon is utilized via reverse glycolytic flux to generate reducing power in the oxidative branch of the PPP. While some M2 6PG was observed in the cancer cell lines after M* labeling, it represented a significantly smaller proportion of the total 6PG (**Figure 5.5B**), indicating that this behavior is likely to be characteristic of a normal metabolic phenotype.

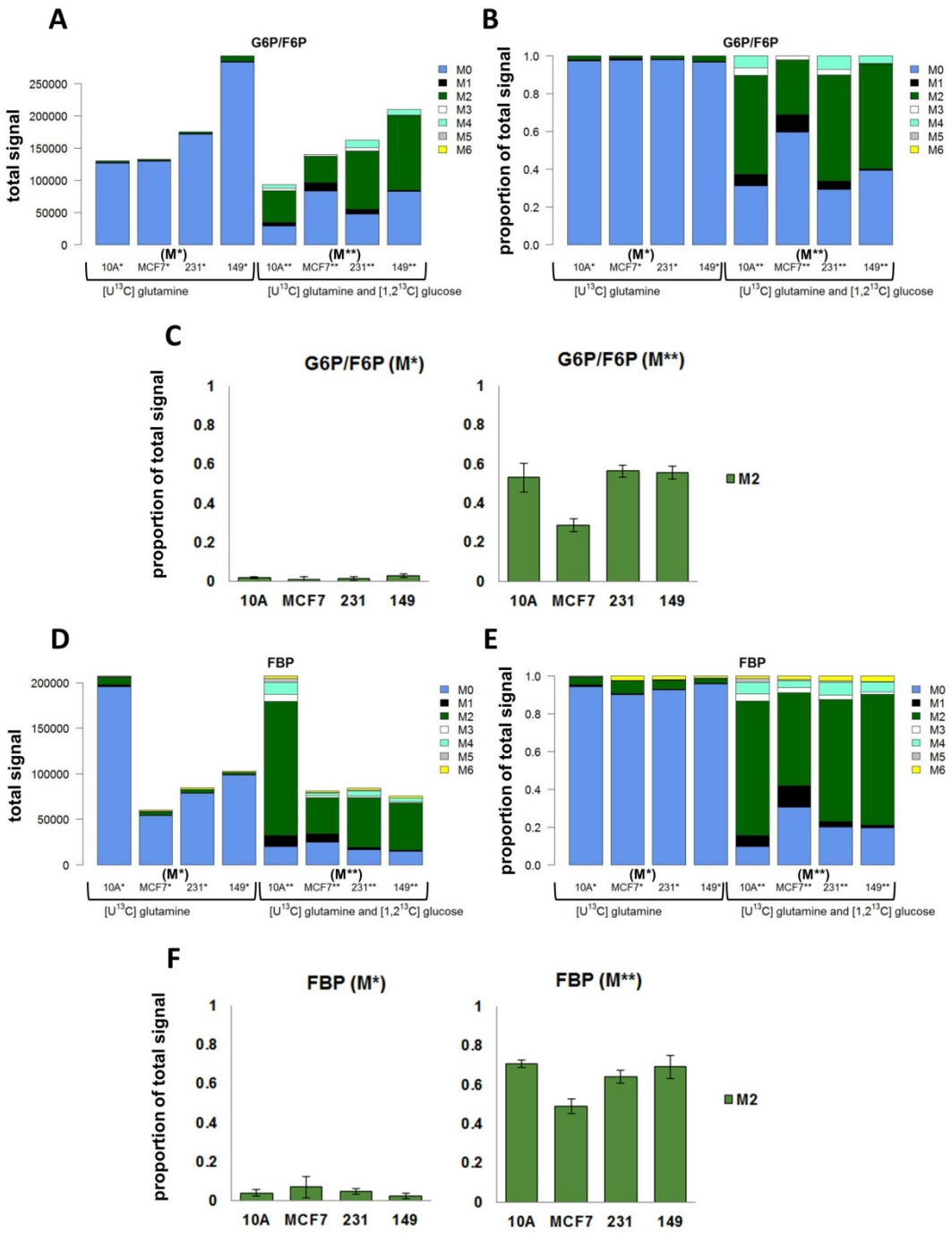


Figure 5.6. Enrichment of glycolytic intermediates. (A) MIDs of the glucose 6-phosphate (G6P) and fructose 6-phosphate (F6P) pools, produced by labeling with M* or M**, were normalized by total protein. The G6P and F6P pools cannot be distinguished via LC-MS. (B) The same MIDs were normalized to 1.0 to reveal the proportion of enrichment of each isotopomer. (C) Comparison of M2 enrichment in 6PG after M* and M** labeling. (D) MIDs of fructose 1,6-bisphosphate (FBP), produced by labeling with M* or M**, were normalized by total protein. (E) The same MIDs were normalized to 1.0 to reveal the proportion of enrichment of each isotopomer. (F) Comparison of M2 enrichment in FBP after M* and M** labeling. All data represent the average of 4 experimental repeats and error bars reflect the standard deviation. M2 fractions statistically different from MCF10A are indicated by * on each plot ($p < 0.05$). All data were corrected for natural abundance of stable isotopes.

Sedoheptulose 7-phosphate (S7P) is an intermediate of the non-oxidative branch of the PPP (**Figure 5.5G**) and participates in highly reversible non-oxidative reactions that involve R5P and GAP or erythrose 4-phosphate (E4P) and F6P. In M** labeled samples, more than 80% of the S7P pool was labeled as M1 or higher in all cell lines, indicating non-oxidative PPP flux exists in all cell lines. In M* labeled samples lacking labeled glucose, no significant labeling was observed in S7P, however, indicating that the non-oxidative flux to S7P is largely glucose derived (**Figure 5.5E**). While an M2 and M4 peak can be observed in SUM-149 cells under M* labeling conditions, the large error bars and lack of statistical difference with MCF10A cells suggests that the SUM-149 M2 isotopomer may be the product of measurement errors (**Figure 5.5F**).

The 6PG MIDs produced from M* labeling suggest that some amount of glutamine derived carbon makes its way to the oxidative branch of the PPP. The MIDs of G6P/F6P (which cannot be separated via LC-MS) and FBP also suggest that a very small portion of their respective pools are in the M2 form when M* media is used (which contains labeled glutamine and no labeled glucose) (**Figure 5.6**). While the data is equivocal, it does suggest that some form of reverse gluconeogenic-like flux may play a role in all of the cells, perhaps in an early transient phase prior to isotopic steady state. In the future, more detailed time course experiments where MIDs of glycolytic and PPP intermediates are measured prior to isotopic steady state could help elucidate the mechanism by which 6PG becomes labeled with glutamine derived carbon. Potential mechanisms include glycolytic oscillations and/or futile cycles, which are important in a number of tissues but have not, to my knowledge, been systematically explored within cancer cells.

Reductive carboxylation in SUM-149 cells

Several recent studies have demonstrated the importance of reductive carboxylation of alpha-ketoglutarate (AKG) in aggressive cancers. In this type of metabolism, the conversion of glutamine derived AKG to isocitrate and CO_2 is catalyzed by isocitrate dehydrogenase (IDH). Isocitrate isomerizes to citrate, which ultimately generates ACOA for cytosolic fatty acid synthesis [279-282].

This type of reductive metabolism appears to be especially important in cancer cells with mitochondrial impairment [281].

Cells use citrate to maintain both anaplerotic reactions in the TCA cycle and macromolecular synthesis in the cytoplasm [279]. Under sufficient oxygen, normal cells primarily generate citrate from glucose derived ACOA. Under hypoxic conditions, however, glutamine has been shown to be a primary cellular source of citrate via reductive metabolism. A variety of mutations in isoforms of IDH have been identified in cancer cells that drive reductive carboxylation of AKG [279-281], underscoring the importance that this metabolic process may have in cancer.

When [U^{13}C] glutamine is used as a tracer in cells undergoing reductive carboxylation of AKG, M5 citrate is expected to form [279, 280]. In **Figure 5.7**, citrate MIDs produced from M* and M** labeling are presented. SUM-149 cells produce approximately 5 times as much M5 citrate compared to the other cell lines (**Figure 5.7C**), suggesting that the SUM-149 cells are actively metabolizing AKG via reductive carboxylation.

When M5 citrate is formed from reductive carboxylation of AKG, the M5 citrate can be metabolized in the cytoplasm to form M2 ACOA and M3 oxaloacetate (OAA) via a reaction catalyzed by ATP citrate lyase. M3 OAA can then form M3 aspartate (via transamination) or M3 malate. In contrast, in the mitochondrial oxidative reactions of the TCA cycle, M5 AKG will generate M4 succinate, followed by M4 fumurate, M4 malate, and M4 OAA [279, 280].

The MID profiles of aspartate (**Figure 5.8A-B**) and malate (**Figure 5.8D-E**) after labeling with M* and M** are very similar. In the TCA cycle and in the cytoplasm, malate is converted to OAA, which equilibrates to form aspartate via a transaminase reaction. Therefore, even though the OAA pool cannot be observed by LC-MS, its presence and labeling patterns can be inferred from malate and aspartate. SUM-149 cells exhibit substantially more M3 aspartate (**Figure 5.8C**) and M3 malate (**Figure 5.8F**) than the other cell lines. Together these data are in agreement with other studies that have identified reductive carboxylation as essential to glutamine dependent fatty acid synthesis under

hypoxic conditions. Evidence suggests that SUM-149 cells are adapted to hypoxia [283].

Another possible metabolic pathway for M2 ACOA and M3 aspartate generated from reductive carboxylation of AKG in SUM-149 cells is the formation of the 5 carbon N-acetylaspartate (NAA), a free amino acid found in very high concentrations in the brain that is associated with a number of neuronal disorders. There is some evidence to suggest NAA may also be a marker of neuronal mitochondrial dysfunction [284]. High concentrations of NAA were recently reported in ovarian tumors [285, 286]. In addition, NAA levels, along with the related compound N-acetyl-aspartylglutamate (NAAG), were found to strongly correlate with primary epithelial ovarian cancer and metastatic ovarian tumor samples in a large scale metabolomics analysis of 100 ovarian tumor biopsy samples [287]. In the brain, NAA is thought to control osmotic stress as well as play a role in fatty acid and myelin synthesis [284, 287]. It is possible that NAA also plays some role in the fatty acid synthesis of aggressive breast and ovarian cancer cells.

The MIDs of NAA after M* and M** labeling are presented in **Figure 5.9**. All cell lines except for MCF7 exhibit some amount of ^{13}C labeling in NAA, though the majority of the NAA pool in all cell lines in unlabeled. In the MDA-MB-231 and SUM-149 metastatic cell lines, the MIDs suggest that NAA labeling is almost exclusively glutamine derived (given that there is no observable difference in labeling between the M* and M**). In MCF10A, however, the NAA label appears to derive from both glucose and glutamine. It is also interesting that, while MCF7 and SUM-149 cells appear to exhibit similarly large NAA pool sizes in the MIDs normalized by total protein (**Figure 5.9A**), virtually no labeled NAA is found in the MCF7 primary tumor cells. These data suggest the glutamine derived NAA may be important to the metastatic metabolic phenotype but not to the non-invasive phenotype.

The functional role of NAA in breast and ovarian cancer is unknown. The large NAA pool size in SUM-149 cells (approximately 20% of which is labeled) suggest that NAA could play a metabolic role in these cells. Given the large

unlabeled pool size in MCF7, the data presented here are equivocal. It is possible that the NAA in SUM-149 cells represents a sink for access M2 ACOA and M3 aspartate generated from reductive carboxylation reactions. Given the aggressiveness of IBC, additional investigation of the role of NAA in SUM-149 and other IBC cell lines is warranted.

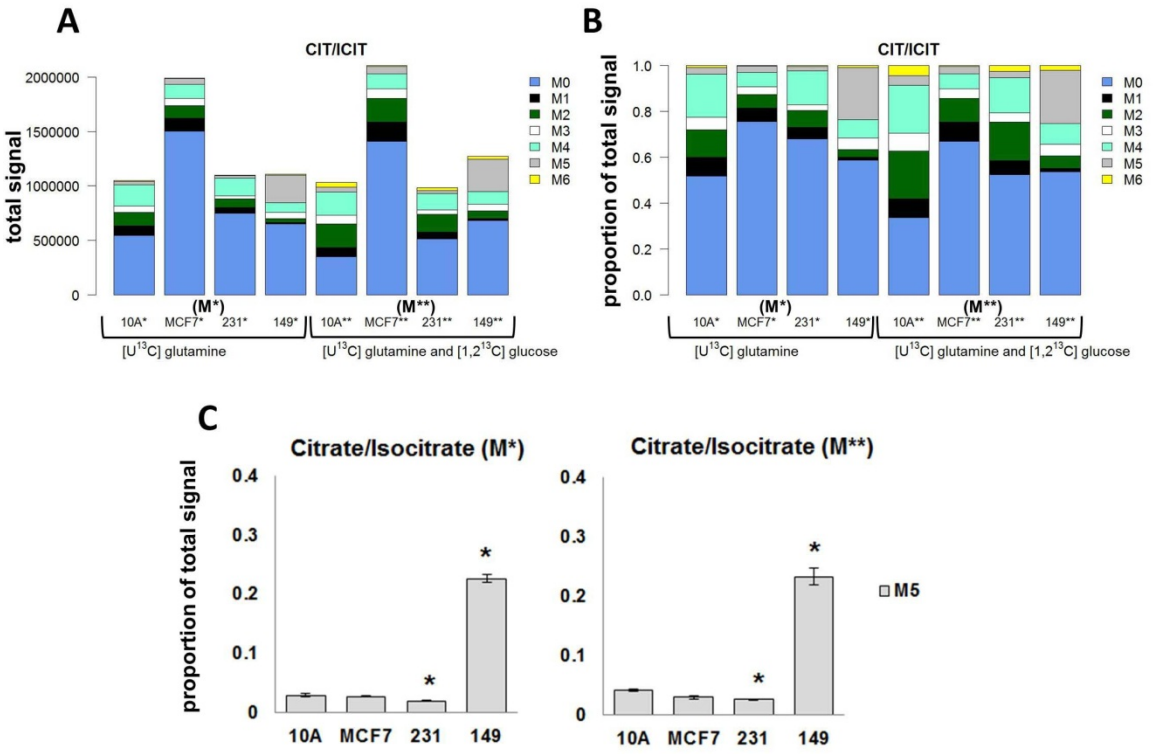


Figure 5.7. Enrichment of ^{13}C in M5 Citrate in SUM-149 cells. (**A**) MIDs of citrate produced by labeling with M* or M**, normalized by total protein. (**B**) The same MIDs were normalized to 1.0 to reveal the proportion of enrichment of each isotopomer. (B) Comparison of M5 enrichment in citrate after M* and M** labeling. Data represent the average +/- the standard deviation of 4 experimental repeats. M5 fractions statistically different from MCF10A are indicated by * on each plot ($p < 0.05$). All data were corrected for natural abundance of stable isotopes.

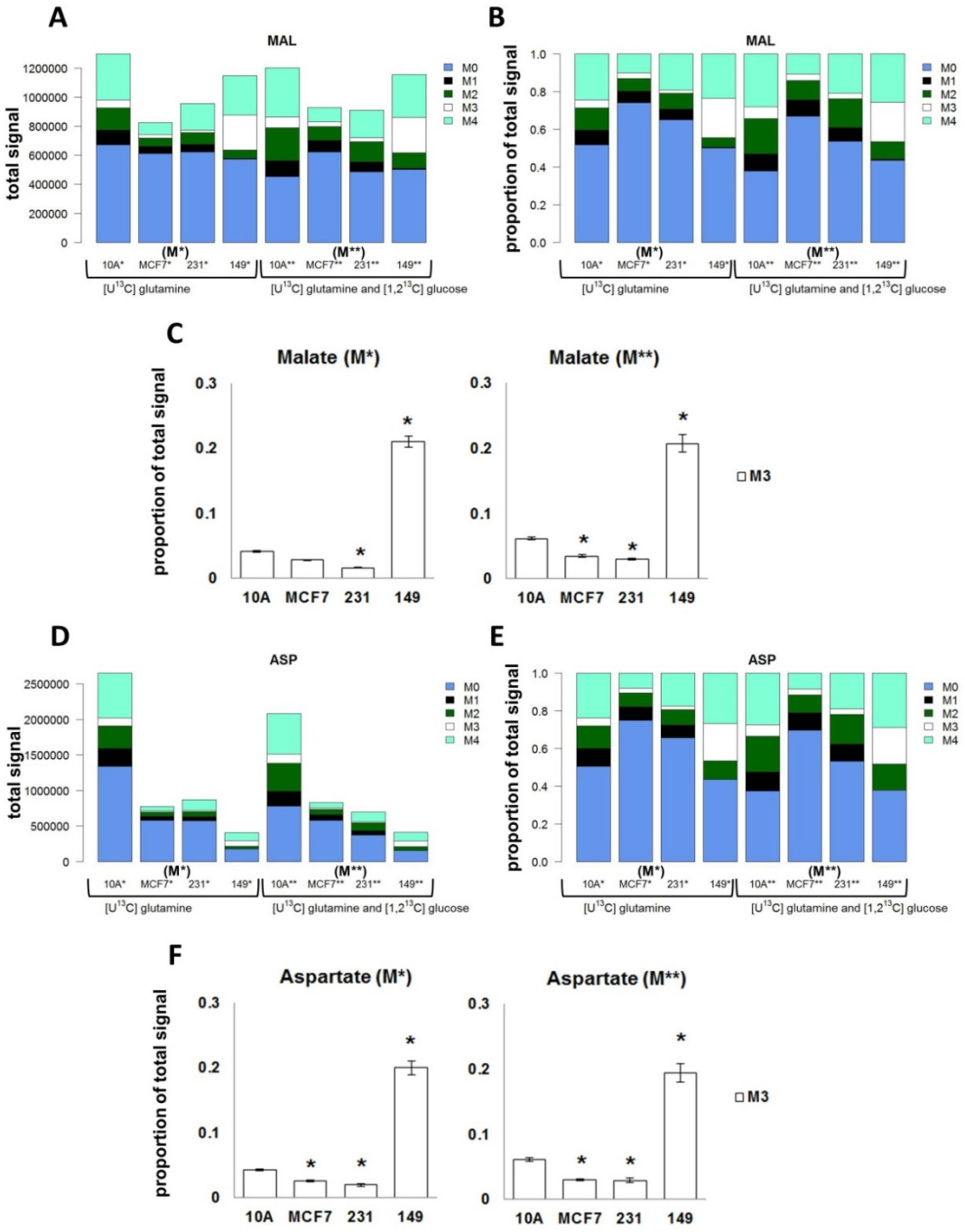


Figure 5.8. Enrichment of ^{13}C in M3 Malate and M3 Aspartate in SUM-149 cells. (**A**) MIDs of malate produced by labeling with M* and M**, normalized by total protein. (**B**) The same MIDs were normalized to 1.0 to reveal the proportion of enrichment of each isotopomer. (**C**) Comparison of M3 enrichment in malate after M* and M** labeling. (**D**) MIDs of aspartate produced by labeling with M* or M**, normalized by total protein. (**E**) The same MIDs were normalized to 1.0 to reveal the proportion of enrichment of each isotopomer. (**F**) Comparison of M3 enrichment in aspartate after M* and M** labeling. All data represent the average of 4 experimental repeats and error bars reflect the standard deviation. M3 fractions statistically different from MCF10A are indicated by * on each plot ($p < 0.05$). All data were corrected for natural abundance of stable isotopes

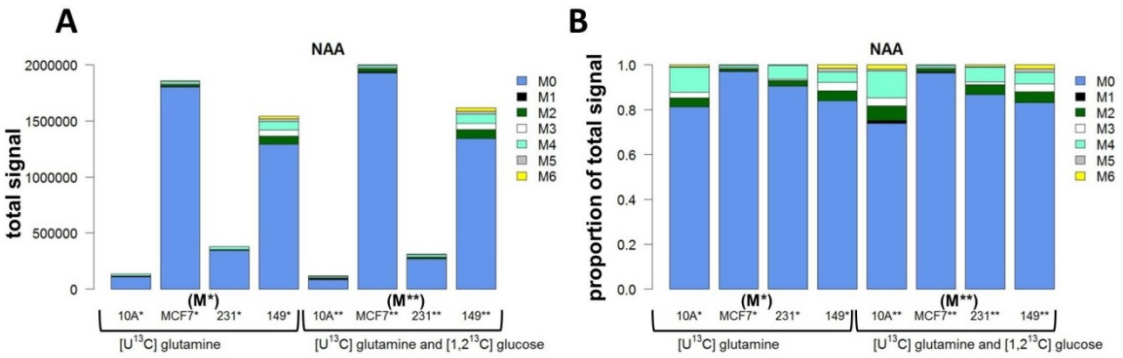


Figure 5.9. Evidence of glutamine derived ^{13}C enrichment in N-acetylasparate in SUM-149 cells. (**A**) MIDs of N-acetylasparate (NAA) produced by labeling with M* or M**, normalized by total protein. (**B**) The same MIDs were normalized to 1.0 to reveal the proportion of enrichment of each isotopomer. All data were corrected for natural abundance of stable isotopes.

5.3.3 Estimation of extracellular flux

The extracellular flux of glucose, lactate, and glutamine was estimated by measuring the change in the concentration of these metabolites in each cell line's media after 5 h of incubation in labeled media (**Figure 5.10A**). In pilot studies, the concentration of glucose was observed to increase significantly in incubated media when no cells were present. This change is assumed to be caused by evaporative water loss. In addition, glutamine concentration has also been observed to decrease in incubated media when no cells were present (data not shown), which is a known phenomenon [239]. To control for changes in extracellular metabolite concentration not related to cellular uptake or secretion, control media were also incubated for 5 h without cells. The concentrations in the control media after 5 h were used to calculate the rate of extracellular flux in cell culture media (see **Methods** for more details).

The estimated flux in mM/h of glucose uptake, lactate production, and glutamine uptake in each cell line's media is presented in **Figure 5.10B**. The net concentration change in mM of each metabolite after 5 h is also shown as a function of total protein (**Figure 5.10C**). From these data, it is apparent that SUM-149 cells are producing a large of amount of lactate relative to the other cell lines. Of interest is the observation that the ratio of glucose to lactate in SUM-149 cells is approximately 1 to 2, which is the stoichiometric ratio expected when

all or most glucose is converted to lactate via glycolysis. This finding suggests that very little glucose is oxidized in the mitochondria of SUM-149 cells.

Extracellular and intracellular flux values are typically represented as nmol/min/mg protein. The estimated extracellular flux values in **Figure 5.10B** were converted to nmol/min/mg units for comparison with other flux studies (**Figure 5.10D**). For MCF10A, the rate of glucose uptake was estimated to be 8.9 nmol/min/mg, which compares favorably to the glucose uptake rate of 13.3 nmol/min/mg reported for MCF10A cells in a study by Richardson et al. [288]. In a study using the A549 lung cancer cell line, Metallo et al. [273] reported extracellular rates of 45, 75, and 10 nmol/min/mg for glucose uptake, lactate production, and glutamine uptake, respectively. The order of magnitude of these rates is similar to the rates estimated for the cancer cell lines in this study (**Figure 5.10D**).

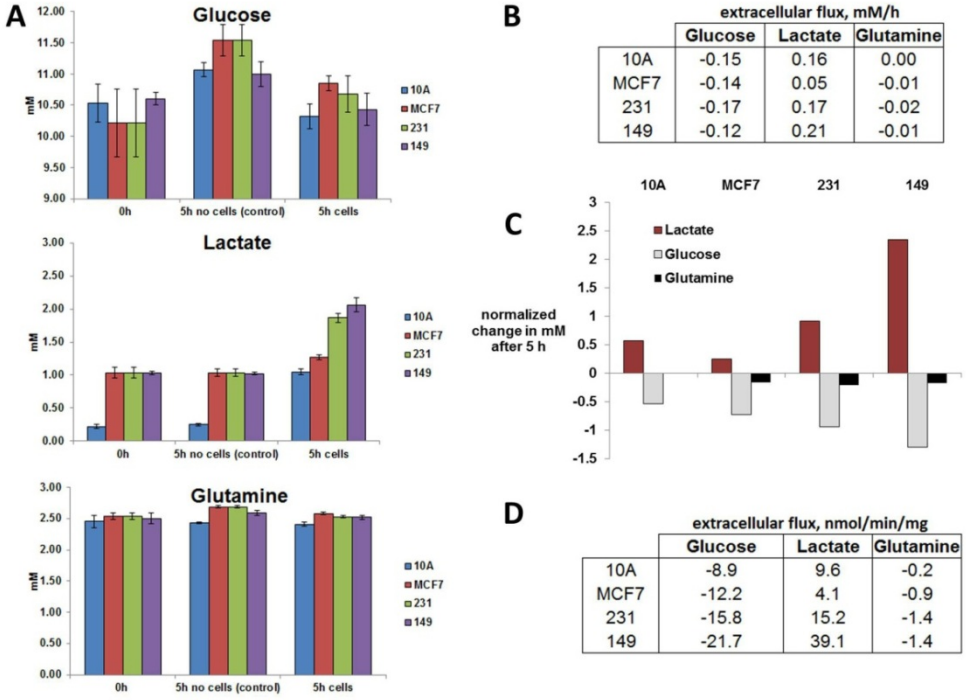


Figure 5.10. Estimation of extracellular flux. (**A**) The concentration of glucose, lactate, and glutamine were measured in fresh media (0 h), media incubated without cells for 5 h (control), and media incubated with cells for 5 h. For the 0 h and 5 h control samples, bars represent the average of 3 experimental repeats +/- the standard deviation. For the 5 h samples with cells, bars represent the average of 8 experimental repeats +/- the standard deviation for each cell line. Metabolite measurements were obtained using a YSI 2900D analyzer. (**B**) The rate of change (mM/h) of each metabolite in each cell line's media after correction and (**C**) the change after 5 h of each metabolite after correction and normalization by total protein. (**D**) The extracellular flux rates converted to nnmol/min/mg. All cells were cultured in 5 mL of media.

5.3.4 Estimation of intracellular TCA cycle flux

Metabolic Flux Analysis

Metabolic flux analysis (MFA) is a technique developed in the metabolic engineering field [271, 272, 289, 290] for the absolute estimation of the steady state flux in cells. MFA requires knowledge of carbon atom transitions involved in glucose metabolism (**Figure 5.3**), measurements of MIDs after the introduction of a tracer, and a system at isotopic steady state (**Figure 5.2**). If all metabolic fluxes are known, a unique mathematical solution exists for the isotope distribution generated for any tracer combination at metabolic steady state [272, 291]. The objective of MFA is to solve the more difficult inverse problem of identifying unknown fluxes from a small subset of measured MIDs. In steady state ^{13}C MFA, the set of reactions involved in the metabolic network under study are mathematically represented as balance equations describing the system at steady state.

The system can be expressed as $\mathbf{S}\boldsymbol{v} = 0$, where **S** is a stoichiometric matrix of the metabolic reactions and $\boldsymbol{v}$ is the vector of unknown fluxes [272, 291-293]. The problem is made more complicated by the fact that each possible positional isotopomer of a metabolite must be accounted for in the balance equations. For a metabolite containing n carbon atoms, there are 2^n positional isotopomers [271] and this large number of positional isotopomers generates an extremely large system of equations. While the mathematics of MFA are well understood [272, 291, 293-297], a number of different computational approaches for solving the underdetermined metabolic system in MFA have been developed [271, 291, 296, 298]. Importantly, each of these methods have been shown to converge to the same solution [271]. The most common computational approach presently used is the elementary metabolite unit (EMU) method [271]. The EMU approach eliminates unnecessary computational overhead by only including those isotopomers in the balance equations that can theoretically be obtained from the tracer used and the atom transitions involved in the metabolic network. In this work, an open-source implementation of the EMU framework called OpenFlux [299] was used.

Regardless of the underlying computational approach used to generate balance equations, ultimately MFA must solve a constrained least squares optimization problem where the difference between experimentally observed MIDS and simulated MIDs is minimized [271, 300]. In addition to the assumption of isotopic steady state, ^{13}C MFA requires explicit knowledge of all carbon atom transitions involved in the network under study (**Figure 5.3**). Many of the reactions involved in central carbon metabolism are well characterized and this information can be found in the literature [270, 273] and in many biochemistry text books [301].

MFA has traditionally relied on MIDs determined by gas chromatography GC-MS [267]. As a consequence of the fairly high turn-over and relatively low concentrations of metabolic intermediates, methods based on GC-MS and nuclear magnetic resonance (NMR) typically must infer ^{13}C labeling patterns in metabolic intermediates. This inference is usually accomplished by obtaining MIDs of biomass components involved in central carbon metabolism, such as protein-bound amino acids [267, 302]. Recently, however, LC-MS has also been used for MFA. In contrast to GC-MS, LC-MS allows for the direct measurement of several metabolic intermediates involved in central carbon metabolism [302].

TCA Cycle model

MFA was used to estimate the intracellular flux in a simplified model of the TCA cycle in all four cell lines. The reactions and associated carbon atom transitions used in the model are summarized in **Figure 5.11** and **Table 5.1**. Measured MIDs of aspartate, citrate/isocitrate (which cannot be distinguished by LC-MS), glutamate, lactate, malate, and ACOA after labeling with M** media were used as model input. The fitted fluxes generated by the model and their 95% confidence intervals are listed in **Table 5.2**.

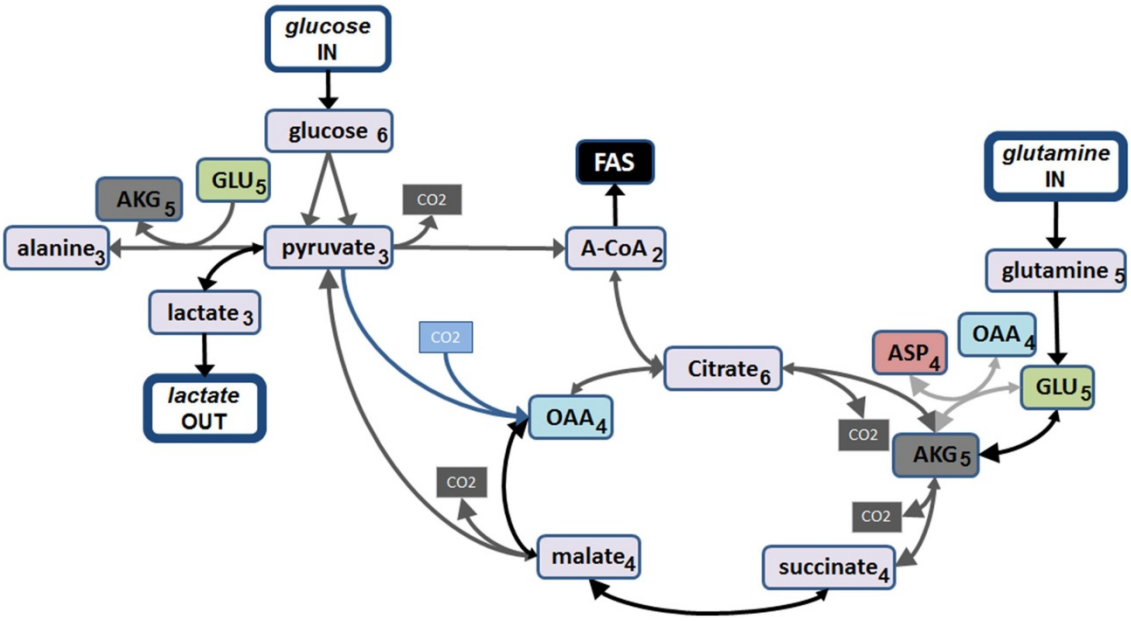


Figure 5.11. Metabolic model of TCA cycle used in metabolic flux analysis. Numbers to the bottom right of a metabolite indicate the number of carbons. Glycolytic and pentose phosphate reactions have been neglected. For bidirectional reactions, the direction treated as the forward direction is represented by a larger arrow head. Metabolites in colors other than light grey indicate the metabolite appears in more than one reaction. For clarity, colored arrow pairs indicate all reactants involved in an overall reaction. Abbreviations: AKG – alpha-ketoglutorate, FAS – fatty acid synthesis, OAA – oxaloacetate.

Extracellular flux rates can be used as constraints for solving the underdetermined system in MFA. In these simulations, the glutamine uptake was set to 1 nmol/min/mg and the final fluxes were scaled by the measured glucose uptake flux (**Figure 5.10D**). The simulations did not explicitly constrain the glucose and lactate fluxes in the fitting procedure because it is informative to see how closely the measured extracellular fluxes are to the fitted extracellular flux when they are allowed to fit freely. For MDA-MB-231 and SUM-149 cells, the glucose uptake and lactate production rates estimated by MFA are very close to the measured values (**Figure 5.10D** and **Table 5.2)**. For MCF7, however, MFA estimated the lactate production to be 14.7 nmol/min/mg, which is larger than the measured value of 4.1 nmol/min/mg. For MCF10A, MFA estimated the lactate production rate to be 18.0 nmol/min/mg, which is also larger than the measured value of 9.6 nmol/min/mg.

Acetyl-CoA and citrate flux

In the pyruvate dehydrogenase (PDH) reaction that converts pyruvate to ACOA, the fitted flux values for MCF10A and SUM-149 cells were very low (0.2 and 0.4 nmol/min/mg, respectively). In contrast, the fitted fluxes in MCF7 and MDA-MB-231 reactions were much higher (10.1 and 15.8 nmol/min/mg, respectively). This difference can be attributed to the fitted lactate production fluxes in MCF10A and SUM-149 cells. In these cells the lactate production rate was fitted at almost exactly two times the glucose uptake rate, which effectively means the model simulated that virtually all glucose in the system was eventually converted to lactate, leaving little excess pyruvate for the PDH reaction.

The ACOA and OAA condensation reaction that forms citrate had a net negative fitted flux in all the cell lines. For simplicity, the model did not assume any compartmentalization and treated all ACOA as a single pool. In reality, the reactions involving ACOA are very different in the mitochondria and the cytoplasm. In the mitochondria, citrate synthase catalyzes the reaction between ACOA and OAA to form citrate. In the cytoplasm, ATP citrate lyase cleaves citrate to make OAA and ACOA and requires the input of ATP. The availability of cytoplasmic ACOA is important for biosynthetic processes, such as fatty acid synthesis. As previously discussed, this reaction is also very important in reductive carboxylation reactions in some cancer cells.

While the MCF7 and MDA-231 cells had negative net fitted flux toward citrate of -0.7 and -9.0 nmol/min/mg, they both had relatively large forward reaction fluxes of 94.3 and 13.1 nmol/min/mg, respectively. This result suggests that, while the ATP citrate layase reaction fluxes were larger in these cells, the mitochondrial citrate synthase reaction was still very active. In contrast, both MCF10A and SUM-149 cells had fitted flux toward citrate of 0.0 nmol/min/mg, indicating that the mitochondrial citrate synthase reaction was not active in these cells in the simulations. The lack of citrate synthase flux is likely a function of the very low fitted flux to ACOA in these cells. While there is some evidence from work in the Merajver lab that citrate synthase may be less active in SUM-149 cells (data not shown), it is unlikely that there would be no flux to mitochondrial

citrate in the normal-like MCF10A cells. Additional experimental validation of this result is warranted.

Treating ACOA as two distinct pools in the model, rather than one, may improve the predictive power of the model. At least one other MFA study of cancer cells treated ACOA as two distinct pools [273]. That study used GC-MS, however, and ACOA MIDs were not used as model input as they were in this study. With LC-MS ACOA can be directly measured. The inclusion of direct measurement of ACOA labeling is expected to be important for discerning flux to the TCA cycle. If ACOA is treated as two spatially separate pools in the model, then ACOA MIDs should not be used as input to the model. This is because the model would assume both pools have identical labeling, which is not likely to be the case. A future iteration of the model will explore the effect of treating ACOA as a mitochondrial pool and a cytoplasmic pool.

Pyruvate flux

The fitted flux estimates suggest that the pyruvate carboxylase reaction, which converts pyruvate directly to OAA, was not very active in the MCF10A and MCF7 cells but was very active in the MDA-MB-231 and SUM-149 cells (with fitted flux values of 15.8 and 17.1 nmol/min/mg, respectively). MFA also estimated that all cells had a positive net flux in the malic enzyme reaction that converts malate to pyruvate. Malic enzyme has been suggested as important in some cancers [6, 303], possibly because the reaction can be used to produce additional NADPH [6]. Separate studies of MCF10A cells [288] and MCF7 cells [304] concluded that malic enzyme is not active in these cells. The latter study looked at intra-mitochondrial malic enzyme, however. It is possible that cytosolic malic enzyme is important in breast cancer [303]. Additional isotope experiments and kinetic studies will be necessary to elucidate the nature of pyruvate flux and the role, if any, of malic enzyme in aggressive breast cancers.

Table 5.1. Carbon atom transitions used in metabolic flux analysis model of the TCA cycle. Reactions assumed to be non-reversible in the model are indicated by NR. The forward (F) and reverse (R) directions of a reversible reaction in the model are listed sequentially and shaded the same color. Letters are used to identify carbon scrambling in a reaction.

Reaction	Carbon Transitions	Enzyme	Type
glucose ⟶ PYR + PYR	abcdef ⟶ cba + def	glycolytic enzymes	NR
PYR + CO2 ⟶ OAA	abc + d ⟶ abcd	citrate synthase	NR
MAL ⟶ PYR + CO2	abcd ⟶ abc + d	malic enzyme	F
PYR + CO2 ⟶ MAL	abc + d ⟶ abcd		R
PYR ⟶ ACOA + CO2	abc ⟶ bc + a	pyruvate dehydrogenase complex	NR
ACOA + OAA ⟶ CIT	ab + cdef ⟶ fedbac	citrate synthase	F
CIT ⟶ ACOA + OAA	fedbac ⟶ ab + cdef	ATP citrate lyase	R
CIT ⟶ AKG + CO2	abcdef ⟶ abcde + f	isocitrate dehydrogenase complex	F
AKG + CO2 → CIT	abcde + f ⟶ abcdef		R
AKG ⟶ SUC + CO2	abcde ⟶ bcde + a	oxoglutarate dehydrogenase complex (not generally reversible)	F
SUC + CO2 ⟶ AKG	bcde + a ⟶ abcde		R
SUC ⟶ 0.5 MAL + 0.5 MAL	abcd ⟶ 0.5 abcd + 0.5 dcba	succinate dehydrogenase, fumarate hydratase	F
MAL ⟶ 0.5 SUC + 0.5 SUC	abcd ⟶ 0.5 abcd + 0.5 dcba		R
OAA ⟶ MAL	abcd ⟶ abcd	malate dehydrogenase	F
MAL ⟶ OAA	abcd ⟶ abcd		R
glutamine ⟶ GLU	abcde ⟶ abcde	glutanimase	F
GLU ⟶ glutamine	abcde ⟶ abcde	glutamine synthetase	R
GLU ⟶ AKG	abcde ⟶ abcde	glutamate dehydrogenas	F
AKG ⟶ GLU	abcde ⟶ abcde		R
PYR + GLU ⟶ ALA + AKG	abc + defgh ⟶ abc + defgh	transaminase	NR
OAA + GLU ⟶ ASP + AKG	abcd + efghi ⟶ abcd + efghi	transaminase	F
ASP + AKG ⟶ OAA + GLU	abcd + efghi ⟶ abcd + efghi		R
PYR ⟶ lactate	abc ⟶ abc	lactate dehydrogenase	F
lactate ⟶ PYR	abc ⟶ abc		R
ACOA ⟶ FA	ab ⟶ ab	acetyl-CoA carboxylase, others	NR

Table 5.2. Estimated TCA cycle flux. Foreword reactions are on the left side of the reaction. Net flux in nmol/min/mg is displayed with the 95% confidence interval. For all reversible reactions, identified by a double arrow, the forward flux is shown in parenthesis.

Reaction	10A flux	95% CI	MCF7 flux	95% CI	231 flux	95% CI	149 flux	95% CI
glucose IN ⟶ glucose	8.9	[8.9, 8.9]	12.2	[11.4, 13.1]	15.8	[15.8, 15.8]	21.7	[21.7, 21.7]
glucose ⟶ PYR + PYR	8.9	[8.9, 8.9]	12.2	[11.4, 13.1]	15.8	[15.8, 15.8]	21.7	[21.7, 21.7]
PYR + CO2 ⟶ OAA	0.0	[0.0, 8.0]	3.0	[0.0, 14.9]	15.8	[13.3, 15.8]	17.1	[17.1, 21.7]
MAL ⟷ PYR + CO2	0.6 (118.9)	[0.6, 8.6]	4.0 (10.1)	[1.0, 15.9]	16.9 (17.9)	[16.2, 16.9]	18.5 (306.9)	[1.4, 23.2]
PYR ⟶ ACOA + CO2	0.2	[0.2, 0.3]	10.1	[10.1, 14.9]	15.8	[15.8, 15.8]	0.4	[0.1, 0.4]
ACOA + OAA ⟷ CIT	-0.7 (0.0)	[-0.7, -0.7]	-6.1 (94.3)	[-8.5, -6.0]	-9.0 (13.1)	[-9.0, -8.9]	-1.6 (0.0)	[-4.9, -1.5]
CIT ⟷ AKG + CO2	-0.7 (0.0)	[-0.7, -0.7]	-6.1 (9.0)	[-8.5, -6.1]	-9.0 (0.0)	[-9.0, -8.9]	-1.6 (47.6)	[-1.6, -1.5]
AKG ⟷ SUC + CO2	-0.1 (9.0)	[-0.1, -0.1]	-5.1 (0.0)	[-7.5, -5.0]	-7.9 (0.0)	[-7.9, -7.9]	-0.2 (0.0)	[-2.9, -0.1]
SUC ⟷ MAL	-0.1 (118.2)	[-0.1, -0.1]	-5.1 (0.0)	[-7.5, -5.1]	-7.9 (46.6)	[-7.9, -7.9]	-0.2 (288.2)	[-0.2, -0.2]
OAA ⟷ MAL	0.7 (4.9)	[0.7, 8.7]	9.0 (207.2)	[8.3, 23.4]	24.8 (234.5)	[24.1, 24.8]	18.7 (54.1)	[18.7, 23.5]
glutamine IN ⟷ glutamine	0.6	[0.6, 0.6]	1.0	[1.0, 1.0]	1.1	[1.1, 1.1]	1.4	[1.4, 1.4]
Glutamine ⟷ GLU	0.6 (1.8)	[0.6, 0.6]	1.0 (50.2)	[1.0, 1.0]	1.1 (1.4)	[1.1, 1.1]	1.4 (18.8)	[1.4, 1.4]
GLU ⟷ AKG	0.4 (0.8)	[-0.6, 0.6]	0.5 (0.8)	[0.4, 1.0]	-0.2 (0.9)	[-1.6, 1.1]	-0.7 (287.7)	[-20.3, 1.4]
PYR + GLU ⟶ ALA + AKG	0.2	[0.0, 1.2]	0.5	[0.0, 0.6]	1.2	[0.0, 2.7]	2.1	[0.0, 2.1]
OAA + GLU ⟷ ASP + AKG	0.0 (0.2)	[0.0, 0.0]	0.0 (49.2)	[0.0, 0.0]	0.0 (0.5)	[0.0, 0.0]	0.0 (0.0)	[0.0, 0.0]
PYR ⟷ LAC	18.0 (47.4)	[16.6, 18.2]	0.0 (0.0)	[0.0, 20.6]	15.6 (225.3)	[14.1, 16.9]	42.4 (188.7)	[38.2, 44.7]
lactate ⟶ lactate OUT	18.0	[16.6, 18.2]	14.7	[0.9, 20.1]	15.6	[14.1, 16.9]	42.4	[18.0, 44.3]
ACOA ⟶ FA	0.9	[0.9, 1.0]	12.3	[11.4, 13.3]	24.8	[24.7, 24.8]	2.0	[1.6, 2.0]

Succinate flux

Succinate is formed in the TCA cycle from AKG in a reaction catalyzed by enzymes associated with the oxoglutarate dehydrogenase complex. This reaction is considered one of the non-reversible steps in the TCA cycle. I was not able to achieve an acceptable convergence of the flux solution with MFA without making the AKG to succinate reaction reversible, however. The fitted net fluxes reported in **Table 5.2** for AKG to succinate are negative in all cell lines. Given the low probability that all four cell lines have a mutation that renders this reaction reversible, it is likely that the model is structurally incomplete with respect to succinate. Succinate cannot be measured by LC-MS but is frequently measured in GC-MS [273, 280] and NMR [288] flux studies. At least a few studies have reported a significant dilution flux in succinate [273, 280], the source of which was not identified. The predictive power of the model presented may be improved by including a dilution flux of an unknown source. Additional clarification of the pathways involved in succinate formation and metabolism is needed.

5.4 Discussion

The experimental data and computational fitted flux values presented in this chapter suggest that some observable network differences exist in the metabolism of the four cell lines studied. It is likely that metabolic transformation occurs as a cancer continues to adapt and become more aggressive over time. Metabolic changes that are representative of distinct stages of breast cancer may suggest dependence by those cells on certain metabolic processes that are not relevant to normal cells (for example, a dependence on reductive carboxylation for proliferation). Importantly, the identification of such differences may help identify therapeutic targets that are non-lethal to normal cells.

The fitted flux values of the TCA cycle suggest some interesting metabolic differences across the cell lines, including variations in ACOA, citrate, pyruvate, and AKG flux. In addition, there were significant differences in the fitted flux values from ACOA to fatty acid synthesis (**Table 5.2**). It is important to note, however, that the fitted flux values presented here are based on a model that

considered the TCA cycle reactions only (**Figure 5.11**). The next iteration of this model is under construction (**Figure 5.12**) and will include the PPP and glycolytic reactions. It is expected that the larger model will be more predictive and able to further elucidate differences in flux related to the malic enzyme, citrate synthase, ATP citrate lyase, and fatty acid synthesis reactions as well as differences in oxidative and non-oxidative flux in the PPP. The source of ACOA flux will be particularly important to resolve in all the cell lines. MCF10A and SUM-149 cells exhibited MIDs that suggest ACOA flux is derived from both glucose and glutamine. In contrast, MCF7 and MDA-MB-231 MIDs suggested that ACOA flux in these cells is only derived from glucose (**Figure 5.4**).

Glutamine is clearly a very important substrate in cancer [4, 7, 280, 282, 305] and significant glutamine derived ^{13}C labeling was observed in many metabolites in this study (**Figure 5.4**, **Figure 5.7**, **Figure 5.8**, and **Figure 5.9**). Despite the importance of glutamine, the glutamine uptake rate is significantly lower than the glucose uptake rate making it more sensitive to measurement errors. Using more accurate measurements of the extracellular flux of glutamine in cell culture media to constrain the MFA model may improve the predictive power of the MFA simulations. While total protein was used to normalize for cell density differences across the cell lines at the time of labeling, a few of the cell lines exhibited confluence around 60% rather than the 80% of the other cell lines. Lower confluence may have been a source of measurement error in the glutamine flux because fewer cells will consume fewer moles of glutamine, affecting the detectable concentration changes in solution. Another possibility for improving the accuracy of glutamine concentration measurements in media is to culture cells for 5 h in 3 mL of media rather than 5 mL. For example, if we assume the glutamine uptake rate is 1.4 nmol/h, then after 5 h the concentration change expected in 5 mL is 0.084 mM, while the concentration change expected in 3 mL is 0.14 mM. The latter net change should be more resistant to experimental and technical errors.

The isotopomer data strongly supports a role for reductive carboxylation of AKG in the highly aggressive SUM-149 cells. This process provides cancer cells

with an additional source of NADPH as well as precursors for cytosolic fatty acid synthesis. Evidence suggests that a reliance on this type of metabolism is induced by hypoxia [279]. Silvera et al. referred to SUM-149 cells and other IBC cells as “constitutively adapted to hypoxia” because of their upregulation of a number of genes and proteins associated with a hypoxic response [283]. These workers suggested that the hypoxic phenotype may confer the ability to produce proteins that promote the proliferation and survival of the highly angiogenic emboli associated with IBC. It is also possible that a prolonged state of hypoxic like conditions confers certain metabolic advantages, perhaps related to reductive carboxylation. Studying the metabolism of other IBC cells to identify if reductive carboxylation is a universal IBC phenotype or if it is specific to some subtypes will be important for future studies.

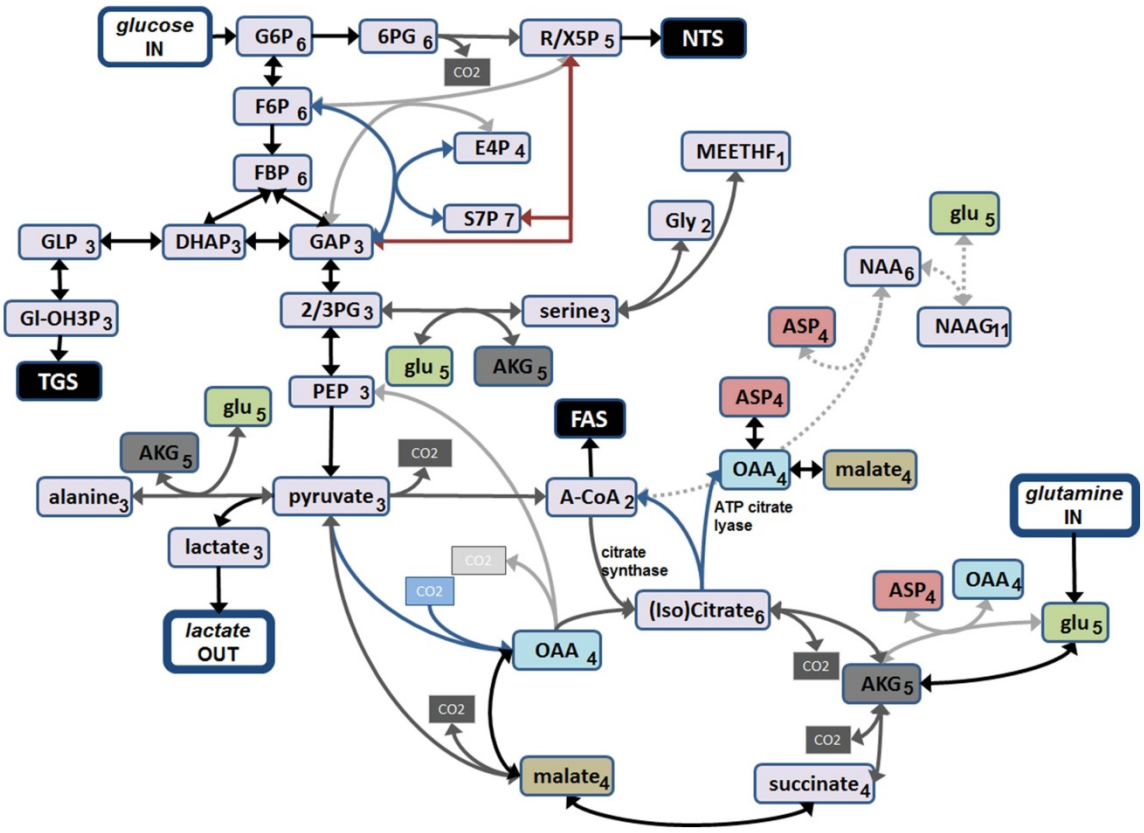


Figure 5.12. Larger model of central carbon metabolism. Numbers to the bottom right of a metabolite indicate the number of carbons. Glycolytic and pentose phosphate reactions have been neglected. For bidirectional reactions, the direction treated as the forward direction is represented by a larger arrow head. Metabolites in colors other than light grey indicate the metabolite appears in more than one reaction. Refer to list of abbreviations for metabolite names.

Chapter 6
Conclusion and Future Directions

6.1 Summary

New approaches for treating aggressive breast cancer are needed. Targeted molecular inhibitors frequently induce off-target effects as well as drug resistant clones capable of evading therapy. The complexity of intracellular networks makes the prediction of off-target effects difficult. Developing new basic science approaches that integrate experimental and theoretical methods to study breast cancer as a complex and adaptive system should generate predictive models that can be used to make rational predictions about candidate drug targets as well as the likelihood of potential off-target effects.

This thesis includes a set of models that represent a step toward that goal. **Chapter 2** introduced the theory of parameter-free logic-based network models for studying the dynamics of signaling networks, while **Chapter 3** presented an application of this theory for studying the PI3K signaling network in progressively tumorigenic breast cancer cells. **Chapter 4** introduced the concept of retroactivity and presented a differential equation based model designed to investigate the likelihood that retroactivity may be a cause of off-target effects from targeted kinase inhibitors. **Chapter 5** relied on experimental and theoretical applications of ^{13}C based isotope tracers to help elucidate differences in intracellular reaction rates in progressively tumorigenic breast cancer cells.

An objective of systems biology is the elucidation of systems level properties associated with a complex phenomenon that cannot be readily observed via experimental methods alone. A critical component of successful systems biology is the iterative development of theoretical models from

experimental data. The models in **Chapter 3** and **Chapter 5**, which relied directly on experimental data collected from breast cancer cells, should be refined, as discussed in each chapter, to improve each model's predictive capacity.

6.2 Future work

Cancer exhibits a number of emergent and adaptive behaviors. A typical solid tumor contains 10^9 cells and each cell contains a set of complex intracellular networks that control gene regulation, signal transduction, cell cycle progression, and metabolism. Cancer cells seem remarkably well equipped to survive in a variety of inhospitable conditions. Changes in a tumor's macro environment, such as the presence of anti-cancer therapeutics, increased acidification, or hypoxia, select for cells that will survive and thrive in these conditions, giving rise to new, perhaps more virulent, sub-clones within the tumor. To design new and effective strategies for treating aggressive cancers, we must consider the underling complexity inherent to a tumor as well as a single cancer cell. We must build models that can more accurately predict the outcome of introducing a perturbation within a single cell as well as within a population of cells. To do so requires the development of models that span the varied spatial scales within a tumor and incorporate the process of clonal evolution.

The importance of metabolic regulation

Warburg was the first to observe the high rate of glucose uptake and lactate production associated with tumors. FDG-PET scans are a routine cancer diagnostic tool because they provide evidence of the glucose uptake rate in a tumor, which correlates with proliferation and aggressiveness. The *in vitro* model of breast cancer progression used in this thesis recapitulated this phenotype, as shown in **Figure 5.10**. The near universality of the Warburg effect across very different cancers in very different tissues (compared to the extreme heterogeneity of genetic lesions and protein deregulation in cancer), suggests that cancers are most likely exploiting an existing (rather than evolving a new) metabolic program. If tumors were actively selecting for novel metabolic programs via somatic

evolution, more metabolic variation than has been reported thus far would be expected. It appears, instead, that cancer cells rely on an evolutionarily conserved form of proliferative metabolism. As cancers continue to progress they appear to acquire mutations that help sustain this metabolic program. Future predictive models of cancer will ideally incorporate metabolism and its regulation.

Developing a hybrid model of glucose metabolism and PI3K signaling

I plan to further develop the model of PI3K signaling presented in **Chapter 3** by completing a detailed sensitivity analysis to evaluate the relative importance of each feedback regulation included in the network and by validating successive iterations of the model with experimental data. Once satisfied that the normal like-model is representative of normal-like breast epithelial PI3K signaling, it will become the normal-like baseline for future models.

Likewise, once the next iteration of the metabolic flux model presented in **Chapter 5** has been refined, the estimated intracellular flux values for each reaction can be used to estimate kinetic parameters for a normal-like ODE model of central carbon metabolism as well as for models representative of distinct stages of breast cancer progression. The continuous ODE model of metabolism and the discrete logic model of PI3K signaling (with links to glucose metabolism and the MAPK pathway) can then be combined into a hybrid discrete continuum model as discussed in the **Future directions** section of **Chapter 3.** Because most models and control theories of metabolism assume constant enzyme concentrations, a primary objective of this model will be the inclusion of enzyme synthesis and regulation.

Developing a multiscale model of breast cancer

Tumors are highly complex, consisting of many layers of spatial complexity that dynamically interact across a wide range of timescales, spanning seconds for post-translational modification to years for cancer progression and metastatic growth. Some of the models presented in this thesis can be used to develop a multiscale model of breast cancer. Multiscale models incorporate the effects of multiple non-linear processes into coherent structures that can then be

used for hypothesis testing or the generation of experimentally testable predictions.

In addition to the intracellular models presented in this thesis, I have also developed a discrete agent based model (ABM) of neural crest cell migration [306]. ABMs are simulation techniques that rely on computing power to investigate complex phenomena that cannot readily be explored by experimental methods. ABMs consist of autonomous decision-making entities called agents that execute behaviors according to a set of rules defined from empirical or experimental observations.

A long term goal is to extend the ABM framework already developed to support a model of mammary cell proliferation capable of simulating malignant transformation in mammary epithelial tissue. The planned ABM simulation domain will consist of a circular layer of luminal epithelial cells arranged to mimic a normal hollow mammary duct (**Figure 1.1A**). Each luminal epithelial cell will be represented as an autonomous agent whose behavior is determined by a set of pre-defined rules based on interactions with other agents and scales in the model. Ultimately, the full model will link several additional spatial and temporal scales of ductal epithelial growth, including a large genetic network, a cell cycle clock, and tissue level dynamics (such as oxygen consumption and lactate production rates, which will be outputs of the metabolic scale of the model). The planned hybrid model of glucose metabolism and PI3K signaling will be an integral part of this model.

A primary output of this future model will be the extent of ductal proliferation, oxygen consumption, and tissue acidification rates under a variety of initial conditions. The ultimate goal will be the generation of predictions related to system level differences between normal and malignant mammary cell proliferation, which can be used to elucidate details of metabolic control within cancer cells.

Although the integration of intracellular molecular networks as a means of understanding targeted therapy resistance and off-target effects is often mentioned as a goal in the literature, tools to accomplish this have not been

developed for broad application. This thesis provides a framework for building more powerful and predictive models of cancer in the future.

Appendix A
Additional Analyses for Chapter 2

Network node reductions are often essential in complex logic networks. An example of node reduction from in a logic network is presented in Figure A1. In this example, AKT inhibits BAD in the interaction network (**Figure A.1A**). Both BAD and P53 are positive regulators of BAX, which is a direct regulator of apoptosis in this network. It is assumed in this logic network that both the activation of BAD (by AKT inhibition) and the activation of P53 are required to induce apoptosis: the activation of only one or the other will not induce apoptosis. Given the logic network in (**Figure A.1B**) and the linear relationship between AKT and BAD, the logic network can be further simplified by removing the BAX node and making AKT a direct inhibitor of apoptosis (**Figure A.1C**).

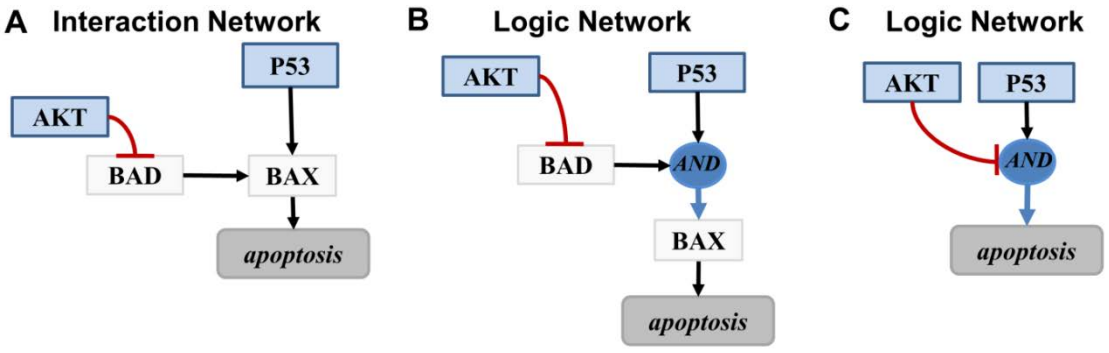


Figure A.1. Network reduction example with linear regulations. (**A**) Interaction network with AKT inhibiting BAD. **(B)** Logic network assuming that both P53 and BAD are both needed activate BAX. (**C**) Equivalent logic network after BAX is removed

Importantly, node network reductions should generate the same qualitative behavior in a network. In **Figure A.2**, the comparison of output from the full and reduced hypothetical network examples with 12 nodes and 4 nodes, respectively, is presented. As can be seen in the synchronous plots at the top of **Figure A.2A** and **Figure A.2B**, the same qualitative output is produced in both

forms of the network model. The most striking difference in this example is the length of the oscillations produced by the two models. Depending on the initial conditions chosen (this example was randomized), the length of the oscillations may be shorter than that observed in **Figure A.2A** (data not shown).

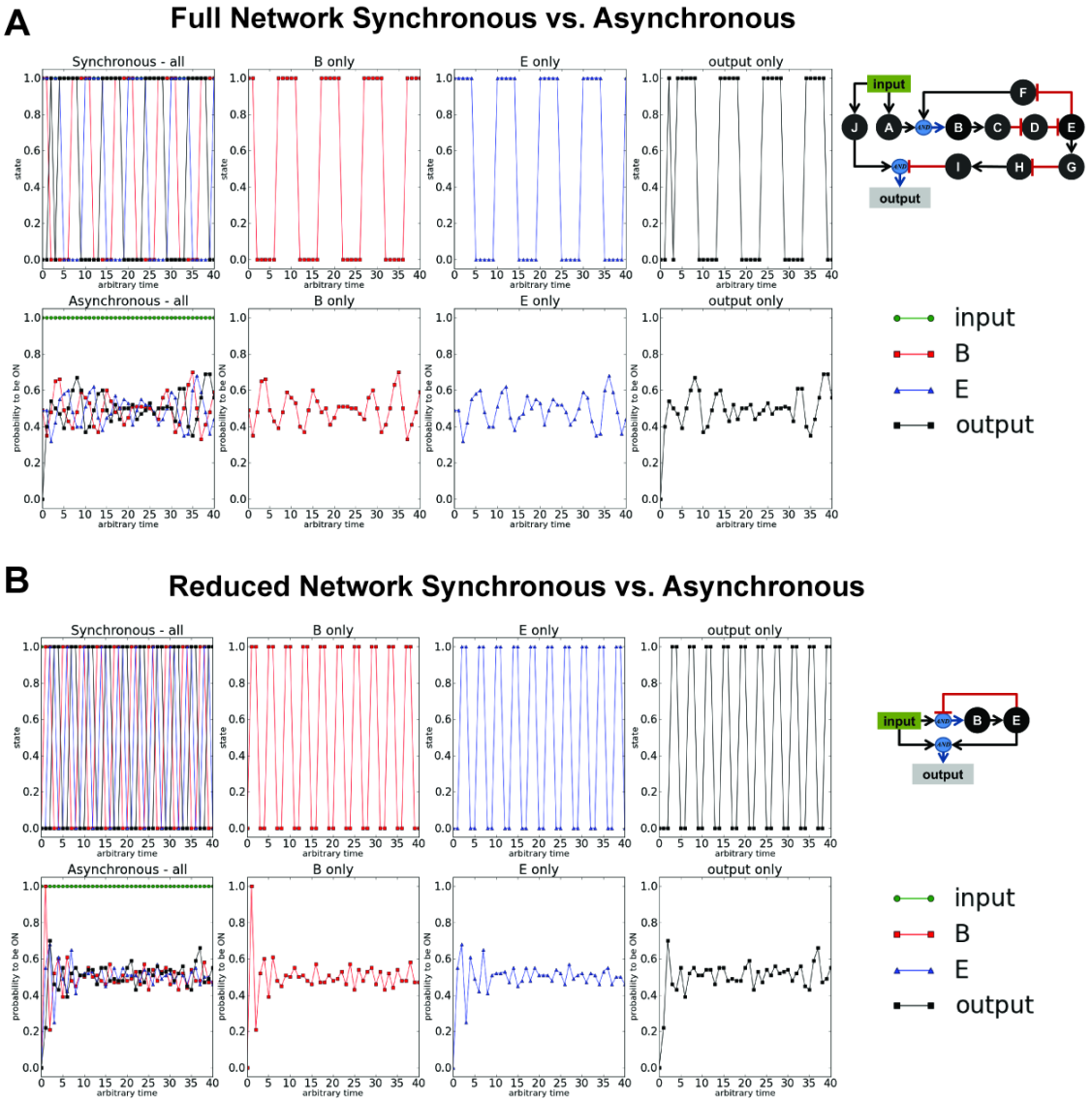


Figure A.2. Qualitative comparison of output from the full and reduced hypothetical network examples with 12 nodes and 4 nodes. (**A**) Synchronous (top) and asynchronous (bottom) simulations for the 12 node network presented in **Figure 2.2** of the main text. (**B**) Synchronous (top) and asynchronous (bottom) simulations for the reduced 4 node network presented in **Figure 2.3** of the main text. The reduced logic network was derived by lumping linear regulations into a single regulation, as described in the main text. Input conditions used for the simulations: **input** node ***ON*** in all cases; all other node values randomly selected. Asynchronous simulations included 100 repeated simulations to get a probability that a node was ***ON*** in the attractor.

Appendix B
Additional Analyses for Chapter 3

Table B.1 summarizes simulation results for the network model in **Figure 3.5** when the ***AND*** logic function is used to regulate AKT instead of the ***OR*** logic function (**Table 3.4**). Under ***AND*** regulation of AKT, the MCF10A and MDA-MB-231 cells compared most favorably to conditions where PTEN is ***OFF***. The experimental data indicate that PTEN is present and activate in these cell lines, however. In contrast, both cell lines compared more favorably to network simulations where PTEN is ***ON*** in the version of the network model that regulated AKT with an ***OR*** logic function, which is the expected outcome.

Table B.1. Oncogenic perturbations to the PI3K/MAPK cross-talk model with AKT regulated by *AND*. The table summarizes the difference in *in silico* simulated results and the experimental *in vitro* results by displaying the total AKT, total ERK and total combined cost penalty for each cell line and each condition. The conditions that generated the lowest total cost penalty are displayed in green and the conditions that generated the highest total cost penalty are displayed in red for each cell line. A penalty of 0.00 indicates that the simulation and experimental results were qualitatively identical for all 6 experimental conditions.

Perturbations													
Inputs			Constitutive		10A Penalties			231 Penalities			149 Penalities		
PTEN	Stress/P53	IR	MTOR	RAS	AKT	ERK	Total	AKT	ERK	Total	AKT	ERK	Total
RAND	*OFF*	*OFF*	--	--	1.60	1.40	3.00	2.10	2.10	4.20	3.70	2.30	6.00
ON	*RAND*	*OFF*	*ON*	--	1.50	1.60	3.10	2.00	2.10	4.10	4.00	2.40	6.40
ON	*RAND/OFF*	*OFF*	--	--	1.50	1.60	3.10	2.00	2.20	4.20	4.00	2.30	6.30
ON	*RAND/OFF*	*RAND*	--	--	1.50	1.60	3.10	2.00	2.20	4.20	4.00	2.30	6.30
OFF	*RAND*	*OFF*	*ON*	--	1.75	1.50	3.25	2.25	2.80	5.05	0.75	3.20	3.95
OFF	*RAND/OFF*	*RAND/OFF*	--	--	0.10	1.40	1.50	0.60	2.10	2.70	3.30	2.30	5.60
OFF	*RAND*	*OFF*	*ON*	*ON*	1.75	1.25	3.00	2.25	3.00	5.25	0.75	2.00	2.75
OFF	*RAND/OFF*	*RAND/OFF*	--	*ON*	1.00	1.25	2.25	1.50	3.00	4.50	1.00	2.00	3.00
ON	*RAND*	*OFF*	*ON*	*ON*	1.50	1.25	2.75	2.00	1.00	3.00	4.00	0.75	4.75
ON	*RAND*	*RAND/OFF*	--	*ON*	1.50	1.25	2.75	2.00	1.00	3.00	4.00	0.75	4.75
ON	*OFF*	*RAND/OFF*	--	*ON*	1.50	1.25	2.75	2.00	1.00	3.00	4.00	0.75	4.75

Appendix C
Additional Analyses and Model Details for Chapter 4

C.1 Non-dimensionalization of the n = 3 network

In order to reduce the complexity of the networks studied, model parameters were non-dimensionalized. The following explains the non-dimensionalization of the $n = 3$ network. The dimensionless parameters of the $n = 5$ and $n = 7$ vertical and lateral networks' were obtained in a similar manner. The ODEs and conservation laws governing the $n = 3$ network (**Figure 4.2C**) at steady state are:

$$\frac{d[Y_1^*]}{dt} = k_1[C_1] - a_1'[Y_i^*]\left[E_{p_1}\right] + d_1'[C_1'] - a_2[Y_2][Y_1^*] + (d_2 + k_2)[C_2] - a_3[Y_3][Y_1^*] + (d_3 + k_3)[C_3] = 0$$

$$\frac{d[C_1]}{dt} = a_1[Y_1]\left[E_{k_1}\right] - (d_1 + k_1)[C_1] = 0$$

$$\frac{d[C_1']}{dt} = a_1'[Y_i^*]\left[E_{p_1}\right] - (d_1' + k_1')[C_1'] = 0$$

$$\frac{d[Y_2^*]}{dt} = -a_2'[Y_2^*]\left[E_{p_2}\right] + d_2'[C_2'] + k_2[C_2] = 0$$

$$\frac{d[C_2]}{dt} = a_2[Y_2][Y_1^*] - (d_2 + k_2)[C_2] = 0$$

$$\frac{d[C_2']}{dt} = a_2'[Y_2^*]\left[E_{p_2}\right] - (d_2' + k_2')[C_2'] = 0$$

$$\frac{d[Y_3^*]}{dt} = -a_3'[Y_3^*]\left[E_{p_3}\right] + d_3'[C_3'] + k_3[C_3] - k_{on}[Y_3^*][D] + k_{off}[C_D] = 0$$

$$\frac{d[C_3]}{dt} = a_3[Y_3][Y_1^*] - (d_3 + k_3)[C_3] = 0$$

$$\frac{d[C_3']}{dt} = \ a_3'[Y_3^*]\left[E_{p_3}\right] - (d_3' + \ k_3')[C_3'] = 0$$

$$\frac{d[C_D]}{dt} = \ k_{on}[Y_3^*][D] - k_{off}[C_D] = 0$$

$$Y_{1T} = \ [Y_1] + [Y_1^*] + [C_1] + [C_1'] + [C_2] + [C_3]$$

$$Y_{2T} = \ [Y_2] + [Y_2^*] + [C_2] + [C_2']$$

$$Y_{3T} = \ [Y_3] + [Y_3^*] + [C_3] + [C_3'] + [C_D]$$

$$E_{k_{1\mathrm{T}}} = [\ E_{k_1}] + [C_1]$$

$$E_{p_{1\mathrm{T}}} = [\ E_{p_1}] + [C_1']$$

$$E_{p_{2\mathrm{T}}} = [E_{p_2}] + [C_2']$$

$$E_{p_{3\mathrm{T}}} = [E_{p_3}] + [C_3']$$

$$D_T = [D] + [C_D]$$

Dimensionless Parameters

$$\boldsymbol{P_i} = \frac{k_i E_{k_{\mathrm{iT}}}}{k_i' E_{p_{\ i\mathrm{T}}}} = \frac{V_{max_{k_i}}}{V_{max_{p_i}}}$$

$$\boldsymbol{E_i} = \ \frac{E_{k_{\mathrm{iT}}}}{Y_{iT}} \quad \boldsymbol{E'_i} = \ \frac{E_{p_{\ i\mathrm{T}}}}{Y_{iT}}$$

$$\boldsymbol{K_i} = \frac{d_i + k_i}{a_i\ Y_{iT}} \ = \frac{K_{m_{k_i}}}{Y_{iT}}$$

$$\boldsymbol{K'_i} = \frac{d_i' + k_i'}{a_i'\ Y_{iT}} = \frac{K_{m_{p_i}}}{Y_{iT}}$$

$$\boldsymbol{K_B} \ = \frac{k_{off}/k_{on}}{Y_{3T}} \text{ (for } n = 3)$$

$$\boldsymbol{K_B} \ = \frac{k_{off}/k_{on}}{Y_{nT}} \text{ (for all } n)$$

$$\boldsymbol{I} = \frac{D_T}{Y_{3T}} \text{(for } n = 3) \quad \boldsymbol{I} = \frac{D_T}{Y_{nT}} \text{(for all } n)$$

$$\text{where } \ E_{k_iT} = \begin{cases} i = 1, \ E_{k_{1\mathrm{T}}} \\ i = 2, i = 3, \ Y_{1T} \\ i > 3 \text{ (vertical)}, \ Y_{(i-1)T} \\ i > 3 \text{ (lateral)}, \ Y_{1T} \end{cases}$$

Dimensionless Variables

$$y_i = \ \frac{[Y_i]}{Y_{iT}} \quad y_i^* = \frac{[Y_i^*]}{Y_{iT}}$$

Finally, algebraic rearrangement and substitution yield the following dimensionless steady state equations:

$$(1)\ \boldsymbol{P_1}\frac{y_1}{y_1+\boldsymbol{K_1}}-\frac{y_1^*}{y_1^*+\boldsymbol{K'_1}}=0$$

$$(2)\ -1+y_1+y_1^*\left(1+\frac{y_2}{\boldsymbol{K_2}}+\frac{y_3}{\boldsymbol{K_3}}\right)+\boldsymbol{E_1}\frac{y_1}{y_1+\boldsymbol{K_1}}+\boldsymbol{E'_1}\frac{y_1^*}{y_1^*+\boldsymbol{K'_1}}=0$$

$$(3)\ \boldsymbol{P_2}\frac{y_2y_1^*}{\boldsymbol{K_2}}-\frac{y_2^*}{y_2^*+\boldsymbol{K'_2}}=0$$

$$(4)\ -1+y_2\left(1+\boldsymbol{E_2}\frac{y_1^*}{\boldsymbol{K_2}}\right)+y_2^*+\boldsymbol{E'_2}\frac{y_2^*}{y_2^*+\boldsymbol{K'_2}}=0$$

$$(5)\ \boldsymbol{P_3}\frac{y_3y_1^*}{\boldsymbol{K_3}}-\frac{y_3^*}{y_3^*+\boldsymbol{K'_3}}=0$$

$$(6)\ -1+y_3\left(1+\boldsymbol{E_3}\frac{y_1^*}{\boldsymbol{K_3}}\right)+y_3^*+\boldsymbol{E'_3}\frac{y_3^*}{y_3^*+\boldsymbol{K'_3}}+\boldsymbol{I}\frac{y_3^*}{y_3^*+\boldsymbol{K_B}}=0$$

C.2 Mapping of dimensionless parameter values to dimensional parameter values for numeric simulation

It was numerically more efficient (and in some cases more accurate) to solve the system of ODEs for a long period to ensure steady state was obtained rather than directly solving the set of non-linear steady state equations for each network. Prior to solving the ODE system, a set of dimensionless parameter values was mapped to dimensional parameter values and initial conditions. The example below applies to the $n = 3$ network (all dimensionless parameters are in bold).

Initial conditions

$[Y_1] = 1 \qquad Y_{1T} = [Y_1] \qquad [Y_1^*] = 0$

$[Y_2] = Y_{1T}/\boldsymbol{E_2} \qquad Y_{2T} = [Y_2] \qquad [Y_2^*] = 0$

$[Y_3] = Y_{1T}/\boldsymbol{E_3} \qquad Y_{3T} = [Y_3] \qquad [Y_3^*] = 0$

$E_{k_{1T}} = \boldsymbol{E_1}\, Y_{1T} \quad E_{p_{1T}} = \boldsymbol{E'_1}\, Y_{1T}$

$E_{p_{2T}} = \boldsymbol{E'_2}\, Y_{2T}$

$E_{p_{3T}} = \boldsymbol{E'_3}\, Y_{3T}$

$[C_1] = [C'_1] = [C_2] = [C'_2] = [C_3] = [C'_3] = [C_D] = 0$

Kinetic parameters

$k'_1 = k'_2 = k'_3 = 1$

$k_1 = (\boldsymbol{P_1}\, E_{p_{1T}}\, k'_1)/E_{k_{1T}}$

$k_2 = (\boldsymbol{P_2}\, E_{p_{2T}}\, k'_2)/Y_{1T}$

$k_3 = (\boldsymbol{P_3}\, E_{p_{3T}}\, k'_3)/Y_{1T}$

$d_1 = d'_1 = d_2 = d'_2 = d_3 = d'_3 = k_{on} = 1$

$a_1 = (d_1 + k_1)/(\boldsymbol{K_1}\, Y_{1T}) \quad a'_1 = (d'_1 + k'_1)/(\boldsymbol{K'_1}\, Y_{1T})$

$a_2 = (d_2 + k_2)/(\boldsymbol{K_2}\, Y_{2T}) \quad a'_2 = (d'_2 + k'_2)/(\boldsymbol{K'_2}\, Y_{2T})$

$a_3 = (d_3 + k_3)/(\boldsymbol{K_3}\, Y_{3T}) \quad a'_3 = (d'_3 + k'_3)/(\boldsymbol{K'_3}\, Y_{3T})$

$k_{off} = \boldsymbol{K_B}\, Y_{3T}\, k_{on}$

C.3 Randomly selecting 50,000 parameter sets instead of 5000 produced the same percentage of off-target effects

As described in the Methods section of Chapter 4 and in **Figure C.1**, 5,000 randomly selected parameter sets were used to determine a probability of off-target effects in a given parameter space. When the default parameter ranges defined in **Table 4.1** were used, the probability of off-target effects (i.e., the percentage of off-target effects in the randomly selected parameter sets) was 1.6%. We also selected 50,000 parameter sets using the same ranges and found a probability of 1.5%. Histograms of all parameter values in the 50,000 sampled parameter sets indicate a uniform distribution across all ranges sampled as expected based on the LHS selection method used (**Figure C.2A**). Histograms of the 745 parameter sets (out of these 50,000) that produced off-target effects reveal specific ranges that favor an off-target effect in cycle 2

(**Figure C.2B**). These ranges are in general agreement with the ranges identified by the perturbation method in **Figure 4.3**.

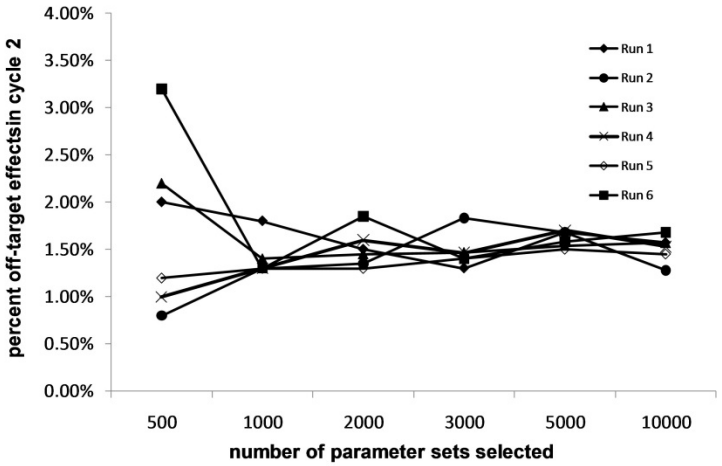


Figure C.1. Determination of the number of parameter sets to sample. The percentage of off-targets in the $n = 3$ network (**Figure4.2C**) was measured in q randomly selected parameter sets using the full parameter ranges defined in **Table 4.1**. The values of q tested were 500, 1000, 2000, 3000, 5000, and 10,000. The test was repeated six times. The variation in the percentage off-target effects in the parameter space appeared to stabilize for repeats greater than or equal to 5000. We, therefore, selected 5000 as the number of parameter sets to sample. A similar test was performed with the *extended* $n = 3$ network (**Figure4.2D**). No significant differences was found with the *extended* $n = 3$ network when values of q tested were 5000 and 50,000 (data not shown).

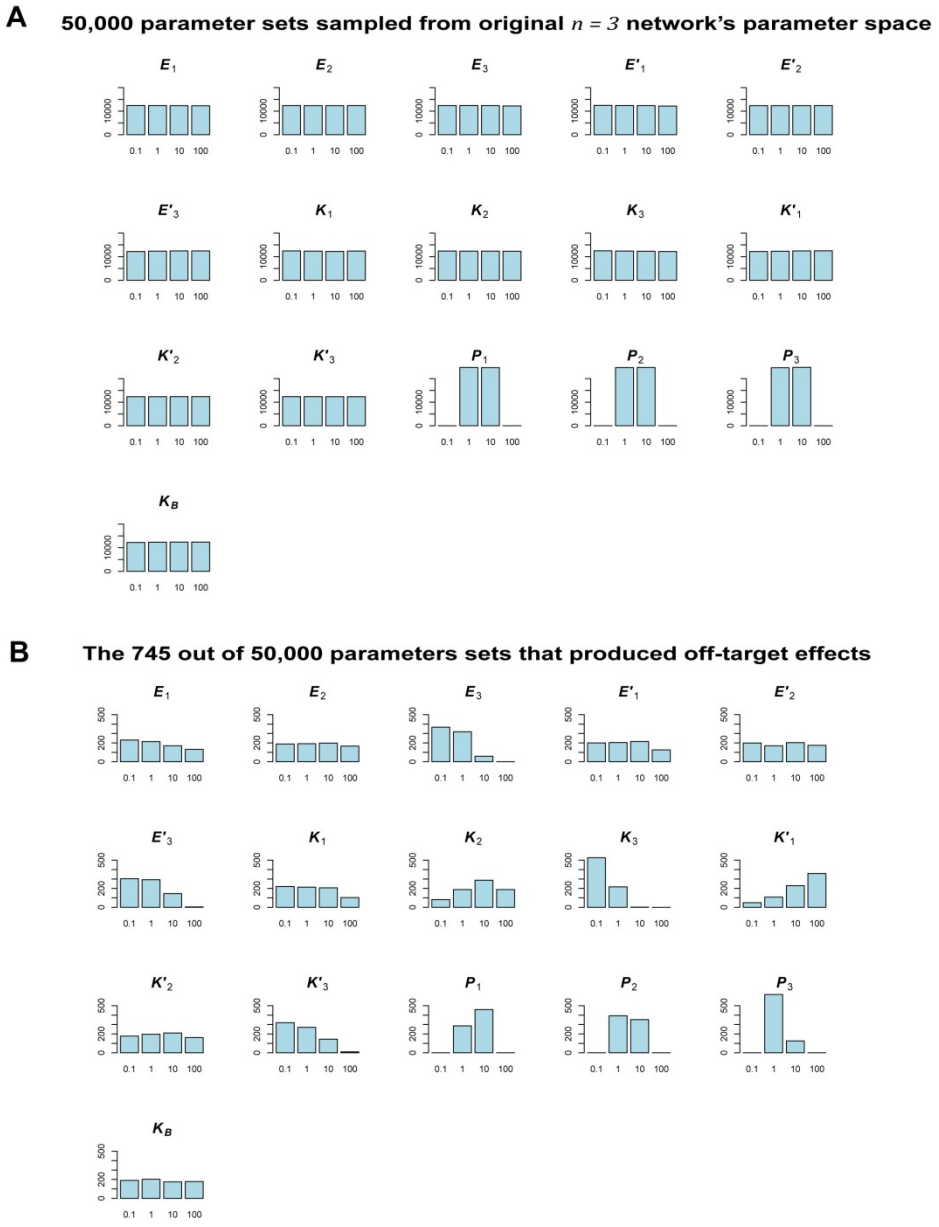


Figure C.2 Histograms of parameter values in 50,000 randomly sampled parameter sets in the n = 3 network. When 50,000 parameter sets were sample from the original *n = 3* parameter space (depicted in **Figure 4.3H** and defined in **Table 4.1**). 1.5% of the sampled parameter sets produced off-target effects in cycle 2. (**A**) Histograms of all parameter values in the 50,000 parameter sets sampled indicate uniform sampling distributions. (**B**) Histograms of the 745 out of the 50,000 parameter sets sampled which produced off-target effects suggest parameter ranges that favor an off-target effect in cycle 2 (compare to **Figure 4.3**).

C.4 Application of Xenopus MAPK Model Parameters

The following values were extracted from the Huang and Ferrell [260] *Xenopus* model:

$$E_{k_{1\mathrm{T}}} = 0.0003\mu\mathrm{M}$$

$Y_{1\mathrm{T}} = 0.003\mu\mathrm{M}$	$Y_{2\mathrm{T}} = 1.2\mu\mathrm{M}$	$Y_{3\mathrm{T}} = 1.2\mu\mathrm{M}$
$E_{p_{1\mathrm{T}}} = 0.0003\mu\mathrm{M}$	$E_{p_{2\mathrm{T}}} = 0.0003\mu\mathrm{M}$	$E_{p_{3\mathrm{T}}} = 0.12\mu\mathrm{M}$
$\mathrm{K}_{\mathrm{m}_{k_1}} = 0.3\mu\mathrm{M}$	$\mathrm{K}_{\mathrm{m}_{k_2}} = 0.3\mu\mathrm{M}$	$\mathrm{K}_{\mathrm{m}_{k_3}} = 0.3\mu\mathrm{M}$
$\mathrm{K}_{\mathrm{m}_{p_1}} = 0.3\mu\mathrm{M}$	$\mathrm{K}_{\mathrm{m}_{p_2}} = 0.3\mu\mathrm{M}$	$\mathrm{K}_{\mathrm{m}_{p_3}} = 0.3\mu\mathrm{M}$
$k_1 = 150\ \mathrm{min}^{-1}$	$k_2 = 150\ \mathrm{min}^{-1}$	$k_3 = 150\ \mathrm{min}^{-1}$
$k_1' = 150\ \mathrm{min}^{-1}$	$k_2' = 150\ \mathrm{min}^{-1}$	$k_3' = 150\ \mathrm{min}^{-1}$

$\mathrm{K_d}$ (the disassociation constant of the inhibitor binding to Y_3*) was not part of the original model and was arbitrarily set to .10 $\mu\mathrm{M}$. Substituting these values into the dimensionless parameter formulas found in **Section C.1** gives the values listed for the *Xenopus* parameter set in **Table 4.2.** The concentrations and first order rate constants used were based on published values in Huang and Ferrell [260], the representation of their MAPK model in the Biomodels repository [307] (http://www.ebi.ac.uk/biomodels-main/BIOMD0000000009), and a MAPK model published by Blüthgen and Herzel [308] which also used the Huang and Ferrell parameters. The Huang and Ferrell *Xenopus* model included double phosphorylation of the MAPK cascade, while our model only considered single phosphorylation, however.

C.5 Additional analysis of the extended n = 3 network with amplification

A numeric perturbation analysis (as described in the Methods) was performed on the *extended* $n = 3$ network parameter space depicted in **Figure C.4H**. In this analysis, off-target effects in cycle 4 which were amplified from cycle 2 were considered. The results indicated that a generally inefficient kinase reaction in cycle 2 ($\boldsymbol{K_2} >> 1$) favored amplification of off-target effects in cycle 4 (**Figure C.3D**). This result is similar to the behavior of $\boldsymbol{K_2}$ in the restricted $n = 3$

parameter space (**Figure C.3D**). In contrast, a more efficient kinase reaction in cycle 4 that operated near the zero-order regime ($K_4 < 1$) favored amplification of off-targets in cycle 4 (**Figure C.4D**). Amplification of off-target effects in cycle 4 were more likely for smaller values of E_2 or E'_2 (**Figure C.4A-B**). Because E_2 is the total enzyme to substrate ratio of the kinase reaction in cycle 2 (i.e., Y_{1T}/Y_{2T}), these results suggest that amplification of off-target effects are more likely to propagate downstream from cycle 2 if the Y_2 protein pool is larger than the Y_1 protein pool. In contrast, off-target effects in cycle 2 were favored in the $n = 3$ network when E_2 or E'_2 was greater than 1 (i.e., the Y_2 protein pool was smaller than the Y_1 protein pool) (**Figure C.3A-B**).

The results also indicate that in the *extended* $n = 3$ network, analogous parameters in cycle 2 and cycle 4 work against each other. For example, large K_2 and small K_4 values contributed to amplification of off-target effects in cycle 4 (**Figure C.4E**). Similarly, large P_4 and small P_2 values favored amplification of off-target effects in cycle 4 (**Figure C.4C**).

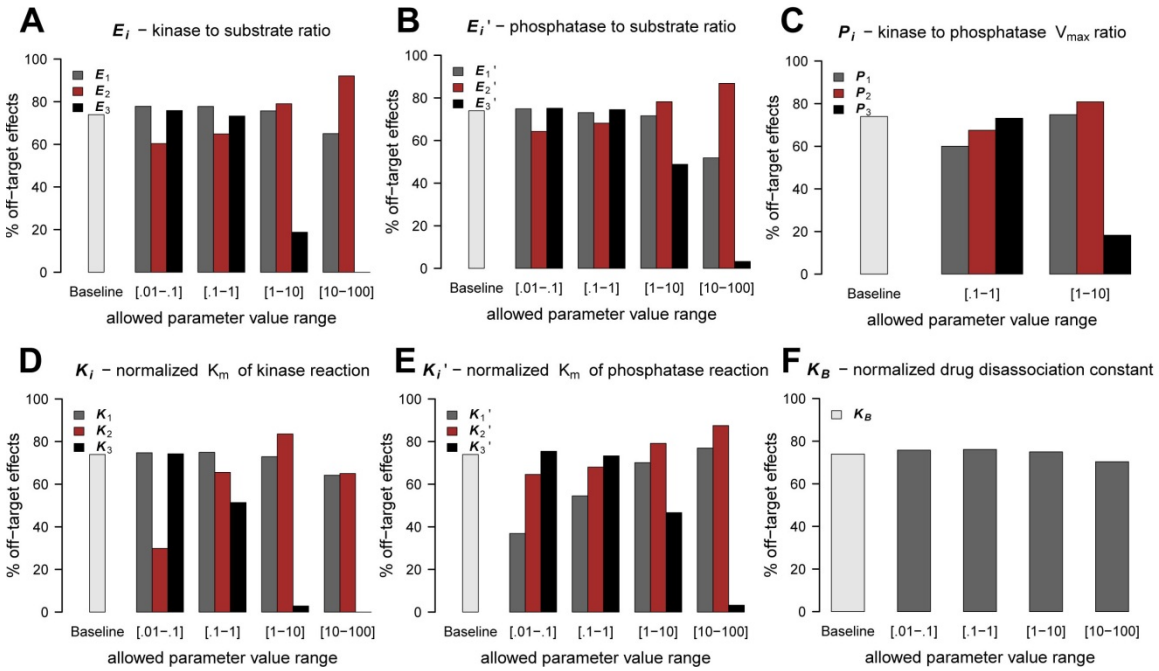


Figure C.3. A numerical perturbation analysis revealed parameter value ranges that promote off-target effects when the $n = 3$ network's parameter space was restricted. A perturbation analysis of the restricted parameter space (depicted in **Figure 43I**) of the $n = 3$ network was performed. The baseline in each plot is 74% (the percentage of off-target effects when 5000 parameter sets were randomly sampled from this parameter space). All other bars reflect the results of systematically perturbing each parameter (one at a time) using the given sub-ranges **(A-F)**.

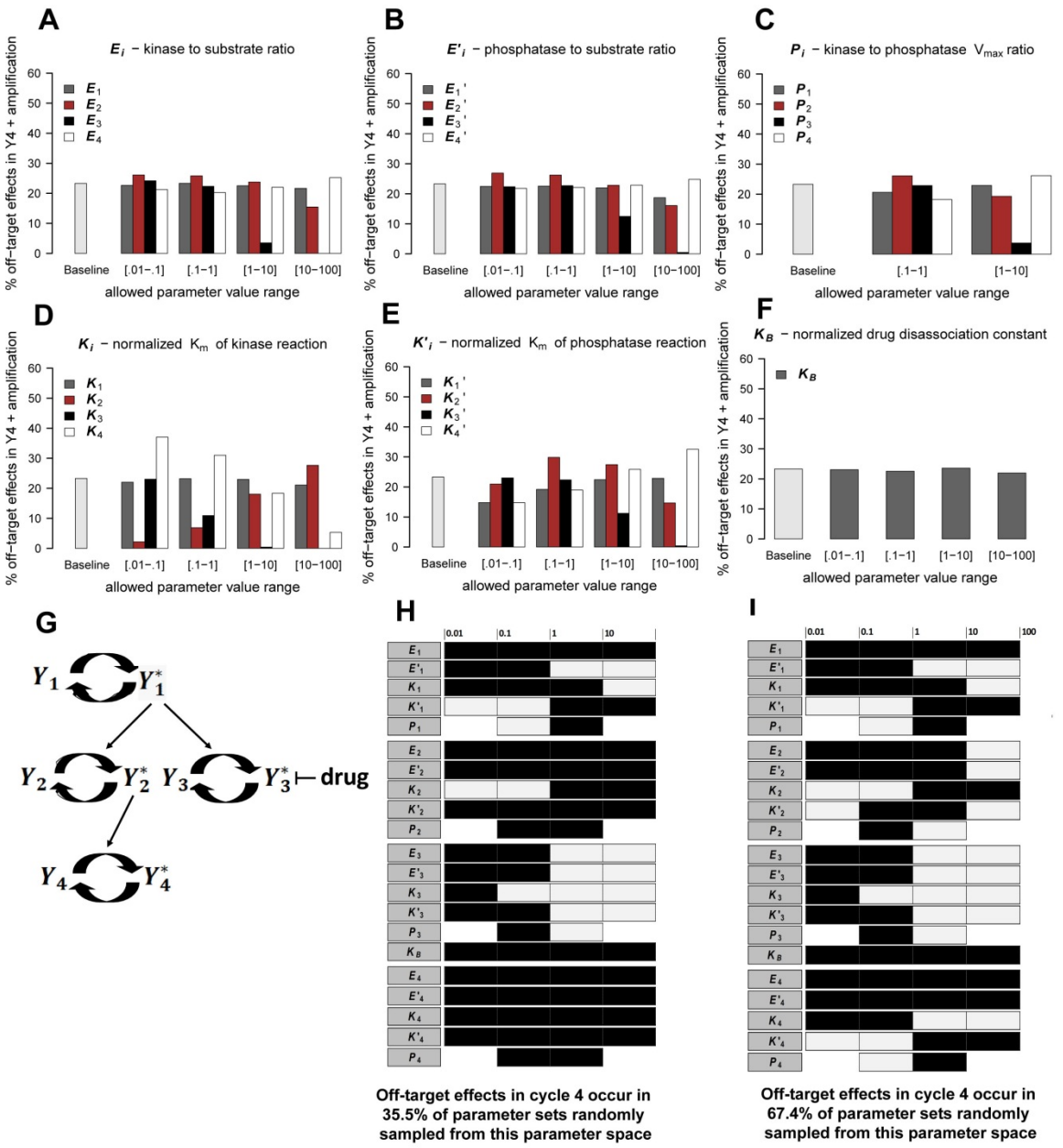


Figure C.4. A numerical perturbation analysis revealed parameter value ranges that promote off-target effects with amplification in *extended* $n = 3$ network. A perturbation analysis of the partially restricted parameter space (depicted in **H**) of the *extended* $n = 3$ network (**G**) was performed. The baseline in each plot is 23.4% (the percentage of off-target effects which amplified from cycle 2 to cycle 4 when 5000 parameter sets were randomly sampled from this parameter space). All other bars reflect the results of systematically perturbing each parameter (one at a time) using the given sub-ranges **(A-F)**. Next, cycle 2 and cycle 4 parameter ranges were further restricted based on the results in **(A-F)** to produce a new restricted parameter space for the *extended* $n = 3$ network. **(I)** This parameter space produced off-target effects in 67.4% of sampled parameter sets and produced off-target effects with amplification from cycle 2 to cycle 4 in 61.9% of sample parameter sets.

Bibliography

1. Wynn, M.L., S.D. Merajver, and S. Schnell, *Unraveling the complex regulatory relationships between metabolism and signal transduction in cancer.* Advances in experimental medicine and biology, 2012. **736**: p. 179-89.
2. Vogelstein, B. and K.W. Kinzler, *Cancer genes and the pathways they control.* Nature medicine, 2004. **10**(8): p. 789-99.
3. Warburg, O., *On the origin of cancer cells.* Science, 1956. **123**(3191): p. 309-14.
4. DeBerardinis, R.J., et al., *Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(49): p. 19345-50.
5. DeBerardinis, R.J., et al., *Brick by brick: metabolism and tumor cell growth.* Current Opinion in Genetics & Development, 2008. **18**(1): p. 54-61.
6. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science, 2009. **324**(5930): p. 1029-33.
7. Wise, D.R. and C.B. Thompson, *Glutamine addiction: a new therapeutic target in cancer.* Trends in biochemical sciences, 2010. **35**(8): p. 427-33.
8. Nakajo, M., et al., *FDG PET Findings of Chronic Myeloid Leukemia in the Chronic Phase Before and After Treatment.* 2007. **32**(10): p. 775-778.
9. Zu, X.L. and M. Guppy, *Cancer metabolism: facts, fantasy, and fiction.* Biochemical and biophysical research communications, 2004. **313**(3): p. 459-65.
10. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
11. Hsu, P.P. and D.M. Sabatini, *Cancer cell metabolism: Warburg and beyond.* Cell, 2008. **134**(5): p. 703-7.
12. Gillies, R.J., I. Robey, and R.A. Gatenby, *Causes and consequences of increased glucose metabolism of cancers.* Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 2008. **49 Suppl 2**: p. 24S-42S.
13. McCormick, F., *Mutant onco-proteins as drug targets: successes, failures, and future prospects.* Curr Opin Genet Dev, 2011. **21**(1): p. 29-33.
14. Hellerstein, M.K., *A critique of the molecular target-based drug discovery paradigm based on principles of metabolic control: advantages of pathway-based discovery.* Metabolic engineering, 2008. **10**(1): p. 1-9.
15. Vivanco, I. and C.L. Sawyers, *The phosphatidylinositol 3-Kinase AKT pathway in human cancer.* Nat Rev Cancer, 2002. **2**(7): p. 489-501.

16. Buzzai, M., et al., *The glucose dependence of Akt-transformed cells can be reversed by pharmacologic activation of fatty acid beta-oxidation.* Oncogene, 2005. **24**(26): p. 4165-73.
17. Elstrom, R.L., et al., *Akt stimulates aerobic glycolysis in cancer cells.* Cancer Res, 2004. **64**(11): p. 3892-9.
18. Courtney, K.D., R.B. Corcoran, and J.A. Engelman, *The PI3K pathway as drug target in human cancer.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2010. **28**(6): p. 1075-83.
19. Engelman, J.A., *Targeting PI3K signalling in cancer: opportunities, challenges and limitations.* Nature reviews. Cancer, 2009. **9**(8): p. 550-62.
20. Shaywitz, A.J., et al., *PI3K enters beta-testing.* Cell metabolism, 2008. **8**(3): p. 179-81.
21. Wynn, M.L., et al., *Kinase inhibitors can produce off-target effects and activate linked pathways by retroactivity.* BMC systems biology, 2011. **5**: p. 156.
22. Christofk, H.R., et al., *The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth.* Nature, 2008. **452**(7184): p. 230-233.
23. Christofk, H.R., et al., *Pyruvate kinase M2 is a phosphotyrosine-binding protein.* Nature, 2008. **452**(7184): p. 181-6.
24. Hitosugi, T., et al., *Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth.* Sci Signal, 2009. **2**(97): p. ra73.
25. Mazurek, S., et al., *Pyruvate kinase type M2 and its role in tumor growth and spreading.* Seminars in Cancer Biology, 2005. **15**(4): p. 300-308.
26. Shaw, R.J. and L.C. Cantley, *Ras, PI(3)K and mTOR signalling controls tumour cell growth.* Nature, 2006. **441**(7092): p. 424-430.
27. Menon, R., et al., *Identification of novel alternative splice isoforms of circulating proteins in a mouse model of human pancreatic cancer.* Cancer Res, 2009. **69**(1): p. 300-9.
28. Jones, S., et al., *Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses.* 2008. **321**(5897): p. 1801-1806.
29. Costello, L.C. and R.B. Franklin, *Tumor cell metabolism: the marriage of molecular genetics and proteomics with cellular intermediary metabolism; proceed with caution!* Mol Cancer, 2006. **5**: p. 59.
30. Aldridge, B.B., et al., *Physicochemical modelling of cell signalling pathways.* Nat Cell Biol, 2006. **8**(11): p. 1195-203.
31. Turner, T.E., S. Schnell, and K. Burrage, *Stochastic approaches for modelling in vivo reactions.* Comput Biol Chem, 2004. **28**(3): p. 165-78.
32. Cornish-Bowden, A., *Fundamentals of enzyme kinetics*. 3rd ed2004, London: Portland Press. xvi, 422 p.
33. Goldbeter, A. and D.E. Koshland, Jr., *An amplified sensitivity arising from covalent modification in biological systems.* Proc Natl Acad Sci U S A, 1981. **78**(11): p. 6840-4.
34. Kholodenko, B.N., *Cell-signalling dynamics in time and space.* Nat Rev Mol Cell Biol, 2006. **7**(3): p. 165-176.
35. Flach, E.H. and S. Schnell, *Use and abuse of the quasi-steady-state approximation.* Systems biology, 2006. **153**(4): p. 187-91.
36. Segel, L.A., *Modeling dynamic phenomena in molecular and cellular biology*1984, Cambridge ; New York: Cambridge University Press. xx, 300 p.
37. Gombert, A.K. and J. Nielsen, *Mathematical modelling of metabolism.* Current opinion in biotechnology, 2000. **11**(2): p. 180-6.
38. Heinrich, R. and S. Schuster, *The regulation of cellular systems*1996, New York: Chapman & Hall. xix, 372 p.

39. Huang, C.Y. and J.E. Ferrell, Jr., *Ultrasensitivity in the mitogen-activated protein kinase cascade.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(19): p. 10078-83.
40. Ventura, A.C., et al., *Signaling properties of a covalent modification cycle are altered by a downstream target.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(22): p. 10032-7.
41. Ventura, A.C., J.A. Sepulchre, and S.D. Merajver, *A hidden feedback in signaling cascades is revealed.* PLoS computational biology, 2008. **4**(3): p. e1000041.
42. Novak, B. and J.J. Tyson, *A model for restriction point control of the mammalian cell cycle.* Journal of theoretical biology, 2004. **230**(4): p. 563-79.
43. Roussel, M.R., *The use of delay differential equations in chemical kinetics.* Journal of Physical Chemistry, 1996. **100**(20): p. 8323-8330.
44. Borisov, N., et al., *Systems-level interactions between insulin-EGF networks amplify mitogenic signaling.* Molecular systems biology, 2009. **5**: p. 256.
45. Ciliberto, A., B. Novak, and J.J. Tyson, *Steady states and oscillations in the p53/Mdm2 network.* Cell Cycle, 2005. **4**(3): p. 488-93.
46. Hu, H., et al., *Feedforward and feedback regulation of the MAPK and PI3K oscillatory circuit in breast cancer.* Cellular signalling, 2013. **25**(1): p. 26-32.
47. Aldridge, B.B., et al., *Fuzzy Logic Analysis of Kinase Pathway Crosstalk in TNF/EGF/Insulin-Induced Signaling.* PLoS Comput Biol, 2009. **5**(4): p. e1000340.
48. Albert, R. and H.G. Othmer, *The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster.* J Theor Biol, 2003. **223**(1): p. 1-18.
49. Albert, I., et al., *Boolean network simulations for life scientists.* Source Code Biol Med, 2008. **3**: p. 16.
50. Li, S., S.M. Assmann, and R. Albert, *Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling.* PLoS biology, 2006. **4**(10): p. e312.
51. Zhang, R., et al., *Network model of survival signaling in large granular lymphocyte leukemia.* Proc Natl Acad Sci U S A, 2008. **105**(42): p. 16308-13.
52. Goldbeter, A. and R. Lefever, *Dissipative structures for an allosteric model. Application to glycolytic oscillations.* Biophys J, 1972. **12**(10): p. 1302-15.
53. Heinrich, R., S.M. Rapoport, and T.A. Rapoport, *Metabolic regulation and mathematical models.* Prog Biophys Mol Biol, 1977. **32**(1): p. 1-82.
54. Sel'kov, E.E., *Self-oscillations in glycolysis. 1. A simple kinetic model.* Eur J Biochem, 1968. **4**(1): p. 79-86.
55. Fersht, A., *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*1999, New York: W. H. Freeman and Company.
56. Cook, P.F. and W.W. Cleland, *Enzyme kinetics and mechanism*2007, London: Garland Science.
57. Schnell, S. and P.K. Maini, *A century of enzyme kinetics: Reliability of the K_M and v_{max} estimates.* Comments on Theoretical Biology, 2003. **8**(2-3): p. 169-187.
58. Savageau, M.A., *Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation.* Journal of Theoretical Biology, 1969. **25**(3): p. 370-379.
59. Savageau, M.A., *Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions.* Journal of Theoretical Biology, 1969. **25**(3): p. 365-369.

60. Savageau, M.A., *Biochemical systems analysis. III. Dynamic solutions using a power-law approximation.* Journal of Theoretical Biology, 1970. **26**(2): p. 215-226.
61. Heinrich, R. and T.A. Rapoport, *A linear steady state treatment of enzymatic chains. Critique of the crossover theorem and a general procedure to identify interaction sites with an effector.* European Journal of Biochemistry, 1974. **42**(1): p. 97-105.
62. Heinrich, R. and T.A. Rapoport, *A linear steady state treatment of enzymatic chains: general properties, control and effector strength.* European Journal of Biochemistry, 1974. **42**(1): p. 89-95.
63. Kacser, H. and J.A. Burns, *The control of flux.* Symposia of the Society for Experimental Biology, 1973. **27**: p. 65-104.
64. Crabtree, B. and E.A. Newsholme, *Sensitivity of a near-equilibrium reaction in a metabolic pathway to changes in substrate concentration.* European Journal of Biochemistry, 1978. **89**(1): p. 19-22.
65. Crabtree, B. and E.A. Newsholme, *A quantitative approach to metabolic control.* Current topics in cellular regulation, 1985. **25**: p. 21-76.
66. Crabtree, B. and E.A. Newsholme, *The derivation and interpretation of control coefficients.* Biochemical Journal, 1987. **247**(1): p. 113-120.
67. Varma, A., M. Morbidelli, and H. Wu, *Parametric sensitivity in chemical systems*1999, New York: USA: Cambridge University Press.
68. Fell, D., *Understanding the control of metabolism*. Frontiers in Metabolism1997, London: Portland Press.
69. Voit, E.O., *Computational Analysis of Biochemical Systems: A practical guide for biochemists and molecular biologists*2000, New York: USA: Cambridge University Press.
70. Cornish-Bowden, A. and M.L. Cárdenas, *Systems biology may work when we learn to understand the parts in terms of the whole.* Biochemical Society Transactions, 2005. **33**(3): p. 516-519.
71. Society, A.C., *Breast Cancer Facts & Figures 2009-2010*, 2009, American Cancer Society, Inc.: Atlanta.
72. Lanigan, F., et al., *Molecular links between mammary gland development and breast cancer.* Cell Mol Life Sci, 2007. **64**(24): p. 3159-84.
73. van Golen, K.L., et al., *RhoC GTPase, a novel transforming oncogene for human mammary epithelial cells that partially recapitulates the inflammatory breast cancer phenotype.* Cancer research, 2000. **60**(20): p. 5832-8.
74. Wu, M. and S.D. Merajver, *Molecular biology of inflammatory breast cancer: applications to diagnosis, prognosis, and therapy.* Breast disease, 2005. **22**: p. 25-34.
75. Lehman, H.L., et al., *Regulation of inflammatory breast cancer cell invasion through Akt1/PKBalpha phosphorylation of RhoC GTPase.* Molecular cancer research : MCR, 2012. **10**(10): p. 1306-18.
76. Alarcon, T., H.M. Byrne, and P.K. Maini, *Towards whole-organ modelling of tumour growth.* Progress in biophysics and molecular biology, 2004. **85**(2-3): p. 451-72.
77. Anderson, A.R. and V. Quaranta, *Integrative mathematical oncology.* Nature reviews. Cancer, 2008. **8**(3): p. 227-34.
78. Gatenby, R.A. and P.K. Maini, *Mathematical oncology: cancer summed up.* Nature, 2003. **421**(6921): p. 321.
79. Huang, S., I. Ernberg, and S. Kauffman, *Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective.* Seminars in cell & developmental biology, 2009. **20**(7): p. 869-76.

80. Kitano, H., *Cancer as a robust system: implications for anticancer therapy.* Nature reviews. Cancer, 2004. **4**(3): p. 227-35.
81. Klinke, D.J., 2nd, *A multiscale systems perspective on cancer, immunotherapy, and Interleukin-12.* Molecular cancer, 2010. **9**: p. 242.
82. Zhang, L., et al., *Multiscale agent-based cancer modeling.* Journal of mathematical biology, 2009. **58**(4-5): p. 545-59.
83. Gatenby, R.A., *A change of strategy in the war on cancer.* Nature, 2009. **459**(7246): p. 508-9.
84. Gatenby, R.A., J. Brown, and T. Vincent, *Lessons from applied ecology: cancer control using an evolutionary double bind.* Cancer research, 2009. **69**(19): p. 7499-502.
85. Alarcon, T., H.M. Byrne, and P.K. Maini, *A mathematical model of the effects of hypoxia on the cell-cycle of normal and cancer cells.* Journal of theoretical biology, 2004. **229**(3): p. 395-411.
86. Billy, F., et al., *A pharmacologically based multiscale mathematical model of angiogenesis and its use in investigating the efficacy of a new cancer treatment strategy.* Journal of theoretical biology, 2009. **260**(4): p. 545-62.
87. Ribba, B., T. Colin, and S. Schnell, *A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies.* Theor Biol Med Model, 2006. **3**: p. 7.
88. Wang, Z., et al., *Identifying therapeutic targets in a combined EGFR-TGFbetaR signalling cascade using a multiscale agent-based cancer model.* Mathematical medicine and biology : a journal of the IMA, 2012. **29**(1): p. 95-108.
89. Forbes, N.S., et al., *Estradiol stimulates the biosynthetic pathways of breast cancer cells: Detection by metabolic flux analysis.* Metabolic Engineering, 2006. **8**(6): p. 639-652.
90. Richardson, A.D., et al., *Central carbon metabolism in the progression of mammary carcinoma.* Breast Cancer Res Treat, 2008. **110**(2): p. 297-307.
91. Wynn, M.L., et al., *Logic-based models in systems biology: a predictive and parameter-free network analysis method.* Integrative biology : quantitative biosciences from nano to macro, 2012. **4**(11): p. 1323-37.
92. Thomas, R., *Boolean formalization of genetic control circuits.* Journal of theoretical biology, 1973. **42**(3): p. 563-85.
93. Thomas, R. and R. D'Ari, *Biological feedback*1990, Boca Raton, Fla.: CRC Press. 316 p.
94. Vidal, M., M.E. Cusick, and A.L. Barabasi, *Interactome networks and human disease.* Cell, 2011. **144**(6): p. 986-98.
95. Albert, R. and A.L. Barabasi, *Statistical mechanics of complex networks.* Reviews of Modern Physics, 2002. **74**(1): p. 47-97.
96. Reguly, T., et al., *Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae.* Journal of biology, 2006. **5**(4): p. 11.
97. Morris, M.K., et al., *Logic-based models for the analysis of cell signaling networks.* Biochemistry, 2010. **49**(15): p. 3216-24.
98. Beard, D.A., *Simulation of cellular biochemical system kinetics.* Wiley interdisciplinary reviews. Systems biology and medicine, 2011. **3**(2): p. 136-46.
99. Crampin, E.J., S. Schnell, and P.E. McSharry, *Mathematical and computational techniques to deduce complex biochemical reaction mechanisms.* Prog Biophys Mol Biol, 2004. **86**(1): p. 77-112.
100. Kauffman, S.A., *Metabolic stability and epigenesis in randomly constructed genetic nets.* Journal of theoretical biology, 1969. **22**(3): p. 437-67.
101. Kauffman, S.A., *The origins of order : self-organization and selection in evolution*1993, New York: Oxford University Press. xviii, 709 p.

102. Glass, L. and S.A. Kauffman, *The logical analysis of continuous, non-linear biochemical control networks.* Journal of theoretical biology, 1973. **39**(1): p. 103-29.
103. Thomas, R., *Circular causality.* Systems biology, 2006. **153**(4): p. 140-53.
104. Faure, A., et al., *Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle.* Bioinformatics, 2006. **22**(14): p. e124-31.
105. von Dassow, G., et al., *The segment polarity network is a robust developmental module.* Nature, 2000. **406**(6792): p. 188-192.
106. Akman, O.E., et al., *Digital clocks: simple Boolean models can quantitatively describe circadian systems.* J R Soc Interface, 2012.
107. Pandey, S., et al., *Boolean modeling of transcriptome data reveals novel modes of heterotrimeric G-protein action.* Molecular systems biology, 2010. **6**: p. 372.
108. Achcar, F., J.M. Camadro, and D. Mestivier, *A Boolean probabilistic model of metabolic adaptation to oxygen in relation to iron homeostasis and oxidative stress.* BMC systems biology, 2011. **5**: p. 51.
109. Schlatter, R., et al., *ON/OFF and beyond--a boolean model of apoptosis.* PLoS computational biology, 2009. **5**(12): p. e1000595.
110. Davidson, E.H., et al., *A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo.* Dev Biol, 2002. **246**(1): p. 162-90.
111. de Silva, A.P. and S. Uchiyama, *Molecular logic and computing.* Nature nanotechnology, 2007. **2**(7): p. 399-410.
112. de Silva, A.P., et al., *Signaling Recognition Events with Fluorescent Sensors and Switches.* Chemical reviews, 1997. **97**(5): p. 1515-1566.
113. de Silva, P.A., N.H.Q. Gunaratne, and C.P. McCoy, *A molecular photoionic AND gate based on fluorescent signalling.* Nature, 1993. **364**(6432): p. 42-44.
114. Baron, R., et al., *Two coupled enzymes perform in parallel the 'AND' and 'InhibAND' logic gate operations.* Organic & biomolecular chemistry, 2006. **4**(6): p. 989-91.
115. Baron, R., et al., *Elementary arithmetic operations by enzymes: a model for metabolic pathway based computing.* Angewandte Chemie, 2006. **45**(10): p. 1572-6.
116. Davidson, E.H., et al., *A genomic regulatory network for development.* Science, 2002. **295**(5560): p. 1669-78.
117. Li, F., et al., *The yeast cell-cycle network is robustly designed.* Proc Natl Acad Sci U S A, 2004. **101**(14): p. 4781-4786.
118. Oliveri, P., D.M. Carrick, and E.H. Davidson, *A regulatory gene network that directs micromere specification in the sea urchin embryo.* Dev Biol, 2002. **246**(1): p. 209-28.
119. Samaga, R., et al., *The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data.* PLoS Comput Biol, 2009. **5**(8): p. e1000438.
120. Davidich, M.I. and S. Bornholdt, *Boolean network model predicts cell cycle sequence of fission yeast.* PLoS One, 2008. **3**(2): p. e1672.
121. Brown, C.T., et al., *New computational approaches for analysis of cis-regulatory networks.* Developmental biology, 2002. **246**(1): p. 86-102.
122. Chaves, M., R. Albert, and E.D. Sontag, *Robustness and fragility of Boolean models for genetic regulatory networks.* J Theor Biol, 2005. **235**(3): p. 431-49.
123. Garg, A., et al., *Synchronous versus asynchronous modeling of gene regulatory networks.* Bioinformatics, 2008. **24**(17): p. 1917-25.
124. Aris, R., *Re, k and π : a conversation on some aspects of mathematical modelling.* APPL MATH MODEL, 1977. **1**(7): p. 386-394.
125. Aho, A.V., M.R. Garey, and J.D. Ullman, *The Transitive Reduction of a Directed Graph.* SIAM J Comput, 1973. **1**(2): p. 131-137.

126. Kachalo, S., et al., *NET-SYNTHESIS: a software for synthesis, inference and simplification of signal transduction networks.* Bioinformatics, 2008. **24**(2): p. 293-5.
127. Kohn, K.W. and M.I. Aladjem, *Circuit diagrams for biological networks.* Molecular systems biology, 2006. **2**: p. 2006 0002.
128. Saadatpour, A., I. Albert, and R. Albert, *Attractor analysis of asynchronous Boolean models of signal transduction networks.* Journal of theoretical biology, 2010. **266**(4): p. 641-56.
129. Zhao, Q., *A remark on "scalar equations for synchronous Boolean networks with biological applications" by C. Farrow, J. Heidel, J. Maloney, and J. Rogers.* IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council, 2005. **16**(6): p. 1715-6.
130. Aldana, M. and P. Cluzel, *A natural class of robust networks.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(15): p. 8710-4.
131. Kauffman, S., et al., *Random Boolean network models and the yeast transcriptional network.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(25): p. 14796-9.
132. Willadsen, K. and J. Wiles, *Robustness and state-space structure of Boolean gene regulatory models.* Journal of theoretical biology, 2007. **249**(4): p. 749-65.
133. Chaves, M., E.D. Sontag, and R. Albert, *Methods of robustness analysis for Boolean models of gene control networks.* Systems biology, 2006. **153**(4): p. 154-67.
134. Wu, M., X. Yang, and C. Chan, *A dynamic analysis of IRS-PKR signaling in liver cells: a discrete modeling approach.* PloS one, 2009. **4**(12): p. e8040.
135. Turner, T.E., S. Schnell, and K. Burrage, *Stochastic approaches for modelling in vivo reactions.* Computational Biology and Chemistry, 2004. **28**(3): p. 165-178.
136. Grima, R. and S. Schnell, *Modelling reaction kinetics inside cells.* Essays in Biochemistry: Systems Biology, Vol 45, 2008. **45**: p. 41-56.
137. Lahav, G., et al., *Dynamics of the p53-Mdm2 feedback loop in individual cells.* Nat Genet, 2004. **36**(2): p. 147-50.
138. Ferrell, J.E., Jr., T.Y. Tsai, and Q. Yang, *Modeling the cell cycle: why do certain circuits oscillate?* Cell, 2011. **144**(6): p. 874-85.
139. Epstein, I.R. and J.A. Pojman, *An introduction to nonlinear chemical dynamics : oscillations, waves, patterns, and chaos*. Topics in physical chemistry1998, New York: Oxford University Press. xiv, 392 p.
140. Murray, J.D., *Mathematical biology*. 3rd ed. Interdisciplinary applied mathematics2001, New York: Springer.
141. Epstein, I.R., *Delay Effects and Differential Delay Equations in Chemical-Kinetics.* International Reviews in Physical Chemistry, 1992. **11**(1): p. 135-160.
142. Lane, D.P., *Cancer. p53, guardian of the genome.* Nature, 1992. **358**(6381): p. 15-6.
143. Michael, D. and M. Oren, *The p53-Mdm2 module and the ubiquitin system.* Seminars in cancer biology, 2003. **13**(1): p. 49-58.
144. Wee, K.B., U. Surana, and B.D. Aguda, *Oscillations of the p53-Akt network: implications on cell survival and death.* PloS one, 2009. **4**(2): p. e4407.
145. Yagi, K., et al., *c-myc is a downstream target of the Smad pathway.* The Journal of biological chemistry, 2002. **277**(1): p. 854-61.
146. Feng, X.H., et al., *Direct interaction of c-Myc with Smad2 and Smad3 to inhibit TGF-beta-mediated induction of the CDK inhibitor p15(Ink4B).* Molecular cell, 2002. **9**(1): p. 133-43.

147. Su, Y., et al., *APC is essential for targeting phosphorylated beta-catenin to the SCFbeta-TrCP ubiquitin ligase.* Molecular cell, 2008. **32**(5): p. 652-61.
148. Lewis, T.S., P.S. Shapiro, and N.G. Ahn, *Signal transduction through MAP kinase cascades.* Advances in cancer research, 1998. **74**: p. 49-139.
149. Yang, W., et al., *Repression of transcription of the p27(Kip1) cyclin-dependent kinase inhibitor gene by c-Myc.* Oncogene, 2001. **20**(14): p. 1688-702.
150. Baldin, V., et al., *Cyclin D1 is a nuclear protein required for cell cycle progression in G1.* Genes & development, 1993. **7**(5): p. 812-21.
151. Coqueret, O., *New roles for p21 and p27 cell-cycle inhibitors: a function for each cell compartment?* Trends in cell biology, 2003. **13**(2): p. 65-70.
152. Basu, A. and S. Haldar, *The relationship between BcI2, Bax and p53: consequences for cell cycle progression and cell death.* Molecular human reproduction, 1998. **4**(12): p. 1099-109.
153. Hay, N., *The Akt-mTOR tango and its relevance to cancer.* Cancer cell, 2005. **8**(3): p. 179-83.
154. Drazin, P.G., *Nonlinear systems*. Cambridge texts in applied mathematics1992, Cambridge [England] ; New York, NY, USA: Cambridge University Press. xiii, 317 p.
155. Jiang, P., et al., *Load-induced modulation of signal transduction networks.* Science signaling, 2011. **4**(194): p. ra67.
156. Tyson, J.J., K.C. Chen, and B. Novak, *Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell.* Current opinion in cell biology, 2003. **15**(2): p. 221-31.
157. Batchelor, E., et al., *Stimulus-dependent dynamics of p53 in single cells.* Molecular systems biology, 2011. **7**: p. 488.
158. Klamt, S., et al., *A methodology for the structural and functional analysis of signaling and regulatory networks.* BMC bioinformatics, 2006. **7**: p. 56.
159. Kitano, H., *Biological robustness.* Nature reviews. Genetics, 2004. **5**(11): p. 826-37.
160. Bornholdt, S., *Boolean network models of cellular regulation: prospects and limitations.* Journal of the Royal Society, Interface / the Royal Society, 2008. **5 Suppl 1**: p. S85-94.
161. Gonzalez, A.G., et al., *GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks.* Bio Systems, 2006. **84**(2): p. 91-100.
162. Jiang, Y., et al., *A multiscale model for avascular tumor growth.* Biophysical journal, 2005. **89**(6): p. 3884-94.
163. Plas, D.R. and C.B. Thompson, *Akt-dependent transformation: there is more to growth than just surviving.* Oncogene, 2005. **24**(50): p. 7435-42.
164. Datta, S.R., et al., *Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery.* Cell, 1997. **91**(2): p. 231-41.
165. Kim, W.Y., et al., *GSK-3 is a master regulator of neural progenitor homeostasis.* Nature neuroscience, 2009. **12**(11): p. 1390-7.
166. Robey, R.B. and N. Hay, *Is Akt the "Warburg kinase"?-Akt-energy metabolism interactions and oncogenesis.* Seminars in cancer biology, 2009. **19**(1): p. 25-31.
167. Berwick, D.C., et al., *The identification of ATP-citrate lyase as a protein kinase B (Akt) substrate in primary adipocytes.* The Journal of biological chemistry, 2002. **277**(37): p. 33895-900.
168. Whiteman, E.L., H. Cho, and M.J. Birnbaum, *Role of Akt/protein kinase B in metabolism.* Trends in endocrinology and metabolism: TEM, 2002. **13**(10): p. 444-51.
169. Xie, X. and K.L. Guan, *The ribosome and TORC2: collaborators for cell growth.* Cell, 2011. **144**(5): p. 640-2.

170. Zinzalla, V., et al., *Activation of mTORC2 by association with the ribosome.* Cell, 2011. **144**(5): p. 757-68.
171. Sabatini, D.M., *mTOR and cancer: insights into a complex relationship.* Nature reviews. Cancer, 2006. **6**(9): p. 729-34.
172. Sarbassov, D.D., et al., *Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex.* Science, 2005. **307**(5712): p. 1098-101.
173. Ibrahim, Y.H., et al., *PI3K Inhibition Impairs BRCA1/2 Expression and Sensitizes BRCA-Proficient Triple-Negative Breast Cancer to PARP Inhibition.* Cancer discovery, 2012. **2**(11): p. 1036-47.
174. Juvekar, A., et al., *Combining a PI3K Inhibitor with a PARP Inhibitor Provides an Effective Therapy for BRCA1-Related Breast Cancer.* Cancer discovery, 2012. **2**(11): p. 1048-63.
175. Rhee, S.H., et al., *MEK is a key modulator for TLR5-induced interleukin-8 and MIP3alpha gene expression in non-transformed human colonic epithelial cells.* The Journal of biological chemistry, 2004. **279**(24): p. 25179-88.
176. Qiang, Y.W., et al., *Insulin-like growth factor I induces migration and invasion of human multiple myeloma cells.* Blood, 2004. **103**(1): p. 301-8.
177. Jiang, B.H. and L.Z. Liu, *Role of mTOR in anticancer drug resistance: perspectives for improved drug treatment.* Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy, 2008. **11**(3): p. 63-76.
178. Hahn-Windgassen, A., et al., *Akt activates the mammalian target of rapamycin by regulating cellular ATP level and AMPK activity.* The Journal of biological chemistry, 2005. **280**(37): p. 32081-9.
179. Ogawara, Y., et al., *Akt enhances Mdm2-mediated ubiquitination and degradation of p53.* The Journal of biological chemistry, 2002. **277**(24): p. 21843-50.
180. Majewski, N., et al., *Hexokinase-mitochondria interaction mediated by Akt is required to inhibit apoptosis in the presence or absence of Bax and Bak.* Molecular cell, 2004. **16**(5): p. 819-30.
181. Rommel, C., et al., *Differentiation stage-specific inhibition of the Raf-MEK-ERK pathway by Akt.* Science, 1999. **286**(5445): p. 1738-41.
182. Zimmermann, S. and K. Moelling, *Phosphorylation and regulation of Raf by Akt (protein kinase B).* Science, 1999. **286**(5445): p. 1741-4.
183. Haas-Kogan, D., et al., *Protein kinase B (PKB/Akt) activity is elevated in glioblastoma cells due to mutation of the tumor suppressor PTEN/MMAC.* Current biology : CB, 1998. **8**(21): p. 1195-8.
184. Inoki, K., T. Zhu, and K.L. Guan, *TSC2 mediates cellular energy response to control cell growth and survival.* Cell, 2003. **115**(5): p. 577-90.
185. Bensaad, K., et al., *TIGAR, a p53-inducible regulator of glycolysis and apoptosis.* Cell, 2006. **126**(1): p. 107-20.
186. DeBerardinis, R.J., et al., *The biology of cancer: metabolic reprogramming fuels cell growth and proliferation.* Cell metabolism, 2008. **7**(1): p. 11-20.
187. Dang, C.V., *c-Myc target genes involved in cell growth, apoptosis, and metabolism.* Molecular and cellular biology, 1999. **19**(1): p. 1-11.
188. Montagner, A., et al., *A novel role for Gab1 and SHP2 in epidermal growth factor-induced Ras activation.* The Journal of biological chemistry, 2005. **280**(7): p. 5350-60.
189. Rodrigues, G.A., et al., *A novel positive feedback loop mediated by the docking protein Gab1 and phosphatidylinositol 3-kinase in epidermal growth factor receptor signaling.* Molecular and cellular biology, 2000. **20**(4): p. 1448-59.

190. Maroun, C.R., et al., *The Gab1 PH domain is required for localization of Gab1 at sites of cell-cell contact and epithelial morphogenesis downstream from the met receptor tyrosine kinase.* Molecular and cellular biology, 1999. **19**(3): p. 1784-99.
191. McCubrey, J.A., et al., *Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance.* Advances in enzyme regulation, 2006. **46**: p. 249-79.
192. Song, M.S., L. Salmena, and P.P. Pandolfi, *The functions and regulation of the PTEN tumour suppressor.* Nature reviews. Molecular cell biology, 2012. **13**(5): p. 283-96.
193. Roux, P.P., et al., *Tumor-promoting phorbol esters and activated Ras inactivate the tuberous sclerosis tumor suppressor complex via p90 ribosomal S6 kinase.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(37): p. 13489-94.
194. Lehr, S., et al., *Identification of major ERK-related phosphorylation sites in Gab1.* Biochemistry, 2004. **43**(38): p. 12133-40.
195. Langlois, W.J., et al., *Negative feedback regulation and desensitization of insulin- and epidermal growth factor-stimulated p21ras activation.* The Journal of biological chemistry, 1995. **270**(43): p. 25320-3.
196. Waters, S.B., et al., *Desensitization of Ras activation by a feedback disassociation of the SOS-Grb2 complex.* The Journal of biological chemistry, 1995. **270**(36): p. 20883-6.
197. Kamioka, Y., et al., *Multiple decisive phosphorylation sites for the negative feedback regulation of SOS1 via ERK.* The Journal of biological chemistry, 2010. **285**(43): p. 33540-8.
198. Nelson, D.L., A.L. Lehninger, and M.M. Cox, *Lehninger principles of biochemistry*. 5th ed2008, New York: W.H. Freeman. 1 v. (various pagings).
199. Gu, H. and B.G. Neel, *The "Gab" in signal transduction.* Trends in cell biology, 2003. **13**(3): p. 122-30.
200. Vojtek, A.B. and C.J. Der, *Increasing complexity of the Ras signaling pathway.* The Journal of biological chemistry, 1998. **273**(32): p. 19925-8.
201. Ceresa, B.P. and J.E. Pessin, *Insulin regulation of the Ras activation/inactivation cycle.* Molecular and cellular biochemistry, 1998. **182**(1-2): p. 23-9.
202. Skolnik, E.Y., et al., *The SH2/SH3 domain-containing protein GRB2 interacts with tyrosine-phosphorylated IRS1 and Shc: implications for insulin control of ras signalling.* The EMBO journal, 1993. **12**(5): p. 1929-36.
203. Myers, M.G., Jr., et al., *The COOH-terminal tyrosine phosphorylation sites on IRS-1 bind SHP-2 and negatively regulate insulin signaling.* The Journal of biological chemistry, 1998. **273**(41): p. 26908-14.
204. Pronk, G.J., et al., *Insulin-induced phosphorylation of the 46- and 52-kDa Shc proteins.* The Journal of biological chemistry, 1993. **268**(8): p. 5748-53.
205. Noguchi, T., et al., *Role of SH-PTP2, a protein-tyrosine phosphatase with Src homology 2 domains, in insulin-stimulated Ras activation.* Molecular and cellular biology, 1994. **14**(10): p. 6674-82.
206. Facchinetti, V., et al., *The mammalian target of rapamycin complex 2 controls folding and stability of Akt and protein kinase C.* The EMBO journal, 2008. **27**(14): p. 1932-43.
207. Ikenoue, T., et al., *Essential function of TORC2 in PKC and Akt turn motif phosphorylation, maturation and signalling.* The EMBO journal, 2008. **27**(14): p. 1919-31.

208. Zhang, S.Q., et al., *Receptor-specific regulation of phosphatidylinositol 3'-kinase activation by the protein tyrosine phosphatase Shp2.* Molecular and cellular biology, 2002. **22**(12): p. 4062-72.
209. Hollestelle, A., et al., *Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines.* Molecular cancer research : MCR, 2007. **5**(2): p. 195-201.
210. Shirangi, T.R., A. Zaika, and U.M. Moll, *Nuclear degradation of p53 occurs during down-regulation of the p53 response after DNA damage.* FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2002. **16**(3): p. 420-2.
211. Wasielewski, M., et al., *Thirteen new p53 gene mutants identified among 41 human breast cancer cell lines.* Breast cancer research and treatment, 2006. **99**(1): p. 97-101.
212. Ponchel, F. and J. Milner, *Temperature sensitivity of human wild-type and mutant p53 proteins expressed in vivo.* British journal of cancer, 1998. **77**(10): p. 1555-61.
213. Qiu, W., et al., *PIK3CA mutations in head and neck squamous cell carcinoma.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2006. **12**(5): p. 1441-6.
214. Schonleben, F., et al., *PIK3CA mutations in intraductal papillary mucinous neoplasm/carcinoma of the pancreas.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2006. **12**(12): p. 3851-5.
215. Saal, L.H., et al., *Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair.* Nature genetics, 2008. **40**(1): p. 102-7.
216. Cross, D.A., et al., *Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B.* Nature, 1995. **378**(6559): p. 785-9.
217. Johne, A., I. Roots, and J. Brockmoller, *A single nucleotide polymorphism in the human H-ras proto-oncogene determines the risk of urinary bladder cancer.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2003. **12**(1): p. 68-70.
218. Kimchi-Sarfaty, C., et al., *A "silent" polymorphism in the MDR1 gene changes substrate specificity.* Science, 2007. **315**(5811): p. 525-8.
219. Sauna, Z.E., et al., *Silent polymorphisms speak: how they affect pharmacogenomics and the treatment of cancer.* Cancer research, 2007. **67**(20): p. 9609-12.
220. Irie, H.Y., et al., *Distinct roles of Akt1 and Akt2 in regulating cell migration and epithelial-mesenchymal transition.* The Journal of cell biology, 2005. **171**(6): p. 1023-34.
221. Yart, A., et al., *A critical role for phosphoinositide 3-kinase upstream of Gab1 and SHP2 in the activation of ras and mitogen-activated protein kinases by epidermal growth factor.* The Journal of biological chemistry, 2001. **276**(12): p. 8856-64.
222. Turke, A.B., et al., *MEK inhibition leads to PI3K/AKT activation by relieving a negative feedback on ERBB receptors.* Cancer research, 2012. **72**(13): p. 3228-37.
223. Won, J.K., et al., *The crossregulation between ERK and PI3K signaling pathways determines the tumoricidal efficacy of MEK inhibitor.* Journal of molecular cell biology, 2012. **4**(3): p. 153-63.
224. Serra, V., et al., *PI3K inhibition results in enhanced HER signaling and acquired ERK dependency in HER2-overexpressing breast cancer.* Oncogene, 2011. **30**(22): p. 2547-57.
225. Choi, W.S. and C.K. Sung, *Inhibition of phosphatidylinositol-3-kinase enhances insulin stimulation of insulin receptor substrate 1 tyrosine phosphorylation and extracellular signal-regulated kinases in mouse R- fibroblasts.* Journal of receptor and signal transduction research, 2004. **24**(1-2): p. 67-83.

226. Saez-Rodriguez, J., et al., *A logical model provides insights into T cell receptor signaling.* PLoS computational biology, 2007. **3**(8): p. e163.
227. Hitosugi, T., et al., *Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth.* Science signaling, 2009. **2**(97): p. ra73.
228. Ward, P.S. and C.B. Thompson, *Signaling in control of cell growth and metabolism.* Cold Spring Harbor perspectives in biology, 2012. **4**(7): p. a006783.
229. Cantley, L.C. and B.G. Neel, *New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(8): p. 4240-5.
230. Duncia, J.V., et al., *MEK inhibitors: the chemistry and biological activity of U0126, its analogs, and cyclization products.* Bioorganic & medicinal chemistry letters, 1998. **8**(20): p. 2839-44.
231. Freeman, M.R., et al., *A metabolic perturbation by U0126 identifies a role for glutamine in resveratrol-induced cell death.* Cancer biology & therapy, 2011. **12**(11): p. 966-77.
232. Breuleux, M., et al., *Increased AKT S473 phosphorylation after mTORC1 inhibition is rictor dependent and does not predict tumor cell response to PI3K/mTOR inhibition.* Molecular cancer therapeutics, 2009. **8**(4): p. 742-53.
233. Anderson, A.R. and V. Quaranta, *Integrative mathematical oncology.* Nat Rev Cancer, 2008. **8**(3): p. 227-34.
234. Alarcon, T., H.M. Byrne, and P.K. Maini, *A multiple scale model for tumor growth.* Multiscale Model. Sim., 2004. **3**(2): p. 440-467.
235. Ribba, B., et al., *A multiscale mathematical model of avascular tumor growth to investigate the therapeutic benefit of anti-invasive agents.* J Theor Biol, 2006. **243**(4): p. 532-41.
236. Frieboes, H.B., et al., *Physical oncology: a bench-to-bedside quantitative and predictive approach.* Cancer Res, 2011. **71**(2): p. 298-302.
237. Singhania, R., et al., *A hybrid model of mammalian cell cycle regulation.* PLoS Comput Biol, 2011. **7**(2): p. e1001077.
238. Feng, Z., et al., *The regulation of AMPK beta1, TSC2, and PTEN expression by p53: stress, cell and tissue specificity, and the role of these gene products in modulating the IGF-1-AKT-mTOR pathways.* Cancer research, 2007. **67**(7): p. 3043-53.
239. Khan, K. and M. Elia, *Factors affecting the stability of L-glutamine in solution.* Clinical nutrition, 1991. **10**(4): p. 186-92.
240. Li, H. and G. Jogl, *Structural and biochemical studies of TIGAR (TP53-induced glycolysis and apoptosis regulator).* The Journal of biological chemistry, 2009. **284**(3): p. 1748-54.
241. Peyssonnaux, C. and A. Eychene, *The Raf/MEK/ERK pathway: new concepts of activation.* Biology of the cell / under the auspices of the European Cell Biology Organization, 2001. **93**(1-2): p. 53-62.
242. Passonneau, J.V. and O.H. Lowry, *The role of phosphofructokinase in metabolic regulation.* Advances in enzyme regulation, 1964. **2**: p. 265-74.
243. Evans, P.R., G.W. Farrants, and P.J. Hudson, *Phosphofructokinase: structure and control.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 1981. **293**(1063): p. 53-62.
244. Downward, J., *Targeting RAS signalling pathways in cancer therapy.* Nature reviews. Cancer, 2003. **3**(1): p. 11-22.

245. Thimmaiah, K.N., et al., *Insulin-like growth factor I-mediated protection from rapamycin-induced apoptosis is independent of Ras-Erk1-Erk2 and phosphatidylinositol 3'-kinase-Akt signaling pathways.* Cancer research, 2003. **63**(2): p. 364-74.
246. Sawyers, C., *Targeted cancer therapy.* Nature, 2004. **432**(7015): p. 294-7.
247. Kumar, N., et al., *Multipathway model enables prediction of kinase inhibitor cross-talk effects on migration of Her2-overexpressing mammary epithelial cells.* Mol Pharmacol, 2008. **73**(6): p. 1668-78.
248. Ventura, A.C., J.A. Sepulchre, and S.D. Merajver, *A hidden feedback in signaling cascades is revealed.* PLoS Comput Biol, 2008. **4**(3): p. e1000041.
249. Del Vecchio, D., A.J. Ninfa, and E.D. Sontag, *Modular cell biology: retroactivity and insulation.* Mol Syst Biol, 2008. **4**: p. 161.
250. Ventura, A.C., T.L. Jackson, and S.D. Merajver, *On the role of cell signaling models in cancer research.* Cancer Res, 2009. **69**(2): p. 400-2.
251. Ventura, A.C., et al., *Signaling properties of a covalent modification cycle are altered by a downstream target.* Proc Natl Acad Sci U S A, 2010. **107**(22): p. 10032-7.
252. Kim, Y., et al., *Substrate-dependent control of MAPK phosphorylation in vivo.* Mol Syst Biol, 2011. **7**: p. 467.
253. Ossareh, H.R., et al., *Long signaling cascades tend to attenuate retroactivity.* Biophys J, 2011. **100**(7): p. 1617-26.
254. Komarova, N.L., et al., *A theoretical framework for specificity in cell signaling.* Mol Syst Biol, 2005. **1**: p. 2005 0023.
255. Wagner, E.F. and A.R. Nebreda, *Signal integration by JNK and p38 MAPK pathways in cancer development.* Nat Rev Cancer, 2009. **9**(8): p. 537-549.
256. Marino, S., et al., *A methodology for performing global uncertainty and sensitivity analysis in systems biology.* J Theor Biol, 2008. **254**(1): p. 178-96.
257. McKay, M.D., R.J. Beckman, and W.J. Conover, *A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code.* Technometrics, 2000. **42**(1): p. 55-61.
258. Tsai, T.Y., et al., *Robust, tunable biological oscillations from interlinked positive and negative feedback loops.* Science, 2008. **321**(5885): p. 126-9.
259. Varma, A., M. Morbidelli, and H. Wu, *Parametric sensitivity in chemical systems*. Cambridge series in chemical engineering2005, Cambridge ; New York: Cambridge University Press. xvi, 342 p.
260. Huang, C.Y. and J.E. Ferrell, Jr., *Ultrasensitivity in the mitogen-activated protein kinase cascade.* Proc Natl Acad Sci U S A, 1996. **93**(19): p. 10078-83.
261. Koshland, D.E., Jr., A. Goldbeter, and J.B. Stock, *Amplification and adaptation in regulatory and sensory systems.* Science, 1982. **217**(4556): p. 220-5.
262. Siddiquee, K., et al., *Selective chemical probe inhibitor of Stat3, identified through structure-based virtual screening, induces antitumor activity.* Proc Natl Acad Sci U S A, 2007. **104**(18): p. 7391-6.
263. Rawlings, J.S., K.M. Rosler, and D.A. Harrison, *The JAK/STAT signaling pathway.* J Cell Sci, 2004. **117**(Pt 8): p. 1281-3.
264. Rhodes, N., et al., *Characterization of an Akt kinase inhibitor with potent pharmacodynamic and antitumor activity.* Cancer Res, 2008. **68**(7): p. 2366-74.
265. Astsaturov, I., et al., *Synthetic lethal screen of an EGFR-centered network to improve targeted therapies.* Sci Signal, 2010. **3**(140): p. ra67.

266. Azam, M., R.R. Latek, and G.Q. Daley, *Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL.* Cell, 2003. **112**(6): p. 831-43.
267. Sauer, U., *Metabolic networks in motion: 13C-based flux analysis.* Molecular systems biology, 2006. **2**: p. 62.
268. Wolfe, R.R. and D.L. Chinkes, *Isotope tracers in metabolic research : principles and practice of kinetic analysis*. 2nd ed2005, Hoboken, N.J.: Wiley-Liss. vii, 474 p.
269. Brunengraber, H., J.K. Kelleher, and C. Des Rosiers, *Applications of mass isotopomer analysis to nutrition research.* Annual review of nutrition, 1997. **17**: p. 559-96.
270. Quek, L.E., et al., *OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis.* Microbial cell factories, 2009. **8**: p. 25.
271. Antoniewicz, M.R., J.K. Kelleher, and G. Stephanopoulos, *Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions.* Metab Eng, 2007. **9**(1): p. 68-86.
272. Antoniewicz, M.R., J.K. Kelleher, and G. Stephanopoulos, *Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements.* Metab Eng, 2006. **8**(4): p. 324-37.
273. Metallo, C.M., J.L. Walther, and G. Stephanopoulos, *Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells.* Journal of biotechnology, 2009. **144**(3): p. 167-74.
274. Hofmann, U., et al., *Identification of metabolic fluxes in hepatic cells from transient 13C-labeling experiments: Part I. Experimental observations.* Biotechnology and bioengineering, 2008. **100**(2): p. 344-54.
275. Munger, J., et al., *Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy.* Nature biotechnology, 2008. **26**(10): p. 1179-86.
276. Fernandez, C.A., et al., *Correction of 13C mass isotopomer distributions for natural stable isotope abundance.* J Mass Spectrom, 1996. **31**(3): p. 255-62.
277. Rosenblatt, J., et al., *Stable isotope tracer analysis by GC-MS, including quantification of isotopomer effects.* The American journal of physiology, 1992. **263**(3 Pt 1): p. E584-96.
278. Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry*. 5th edition ed2002, New York: W. H. Freeman.
279. Wise, D.R., et al., *Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of alpha-ketoglutarate to citrate to support cell growth and viability.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(49): p. 19611-6.
280. Metallo, C.M., et al., *Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia.* Nature, 2012. **481**(7381): p. 380-4.
281. Mullen, A.R., et al., *Reductive carboxylation supports growth in tumour cells with defective mitochondria.* Nature, 2012. **481**(7381): p. 385-8.
282. Holleran, A.L., et al., *Glutamine metabolism in AS-30D hepatoma cells. Evidence for its conversion into lipids via reductive carboxylation.* Molecular and cellular biochemistry, 1995. **152**(2): p. 95-101.
283. Silvera, D. and R.J. Schneider, *Inflammatory breast cancer cells are constitutively adapted to hypoxia.* Cell cycle, 2009. **8**(19): p. 3091-6.
284. Clark, J.B., *N-acetyl aspartate: a marker for neuronal loss or mitochondrial dysfunction.* Developmental neuroscience, 1998. **20**(4-5): p. 271-6.

285. Hascalik, S., et al., *Clinical significance of N-acetyl-L-aspartate resonance in ovarian mucinous cystadenoma.* International journal of gynecological cancer : official journal of the International Gynecological Cancer Society, 2006. **16**(1): p. 423-6.
286. Kolwijck, E., et al., *Ovarian cyst fluid of serous ovarian tumors contains large quantities of the brain amino acid N-acetylaspartate.* PloS one, 2010. **5**(4): p. e10293.
287. Fong, M.Y., J. McDunn, and S.S. Kakar, *Identification of metabolites in the normal ovary and their transformation in primary and metastatic ovarian cancer.* PloS one, 2011. **6**(5): p. e19963.
288. Richardson, A.D., et al., *Central carbon metabolism in the progression of mammary carcinoma.* Breast cancer research and treatment, 2008. **110**(2): p. 297-307.
289. Wiechert, W., *13C metabolic flux analysis.* Metab Eng, 2001. **3**(3): p. 195-206.
290. Young, J.D., et al., *An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis.* Biotechnology and bioengineering, 2008. **99**(3): p. 686-99.
291. Schmidt, K., et al., *Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices.* Biotechnology and bioengineering, 1997. **55**(6): p. 831-40.
292. Wiechert, W., *An introduction to 13C metabolic flux analysis.* Genetic engineering, 2002. **24**: p. 215-38.
293. Wiechert, W. and A.A. de Graaf, *Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments.* Biotechnology and bioengineering, 1997. **55**(1): p. 101-17.
294. Mollney, M., et al., *Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments.* Biotechnology and bioengineering, 1999. **66**(2): p. 86-103.
295. Schmidt, K., et al., *13C tracer experiments and metabolite balancing for metabolic flux analysis: comparing two approaches.* Biotechnology and bioengineering, 1998. **58**(2-3): p. 254-7.
296. Wiechert, W., et al., *Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems.* Biotechnology and bioengineering, 1999. **66**(2): p. 69-85.
297. Wiechert, W., et al., *Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis.* Biotechnology and bioengineering, 1997. **55**(1): p. 118-35.
298. Srour, O., J.D. Young, and Y.C. Eldar, *Fluxomers: a new approach for 13C metabolic flux analysis.* BMC systems biology, 2011. **5**: p. 129.
299. Quek, L.E., et al., *OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis.* Microb Cell Fact, 2009. **8**: p. 25.
300. Wiechert, W., et al., *A universal framework for 13C metabolic flux analysis.* Metabolic engineering, 2001. **3**(3): p. 265-83.
301. Zubay, G.L., *Biochemistry*. 4th ed1998, Dubuque, IA: Wm.C. Brown Publishers. 990, [95] p.
302. van Winden, W.A., et al., *Metabolic-flux analysis of Saccharomyces cerevisiae CEN.PK113-7D based on mass isotopomer measurements of (13)C-labeled primary metabolites.* FEMS yeast research, 2005. **5**(6-7): p. 559-68.
303. Chang, G.G., et al., *Kinetic mechanism of the cytosolic malic enzyme from human breast cancer cell line.* Archives of biochemistry and biophysics, 1992. **296**(2): p. 468-73.
304. Forbes, N.S., et al., *Estradiol stimulates the biosynthetic pathways of breast cancer cells: detection by metabolic flux analysis.* Metabolic engineering, 2006. **8**(6): p. 639-52.

305. Gaglio, D., et al., *Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth.* Molecular systems biology, 2011. **7**: p. 523.
306. Wynn, M.L., P.M. Kulesa, and S. Schnell, *Computational modelling of cell chain migration reveals mechanisms that sustain follow-the-leader behaviour.* Journal of the Royal Society, Interface / the Royal Society, 2012. **9**(72): p. 1576-88.
307. Li, C., et al., *BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models.* BMC Syst Biol, 2010. **4**: p. 92.
308. Bluthgen, N. and H. Herzel, *How robust are switches in intracellular signaling cascades?* J Theor Biol, 2003. **225**(3): p. 293-300.