

A SYSTEMS APPROACH TO CELLULAR
SIGNAL TRANSDUCTION

Jeremy E. Purvis

A DISSERTATION

in

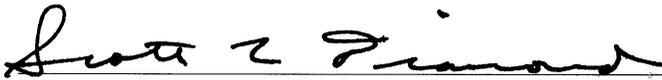
Genomics and Computational Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

2009



Scott Diamond, Professor of Chemical and Biomolecular Engineering and Dissertation Supervisor



Ravi Radhakrishnan, Assistant Professor of Bioengineering and Dissertation Co-Supervisor



Maja Bucan, Professor of Genetics and Graduate Chairperson

Dissertation Committee

Lawrence Brass, Professor of Medicine and Pharmacology

Talid Sinno, Associate Professor of Chemical and Biomolecular Engineering

Lyle Ungar, Associate Professor of Computer and Information Science

Acknowledgements

I wish to gratefully acknowledge, in no particular order, those people who have worked alongside me during my four years at the University of Pennsylvania and have contributed, in either word or deed, to the body of work presented in this dissertation: Scott Diamond, Ravi Radhakrishnan, Skip Brass, Talid Sinno, Lyle Ungar, Warren Ewens, Shane Jensen, Junhyong Kim, Maya Bucan, Andrew Shih, Neeraj Agrawal, Yingting Liu, Shannon Telesco, Josh Weinstein, Sean Maloney, Manash Chatterjee, Matt Flamm, Dan Jaeger, Keith Neeves, Praveen Sethupathy, and Bill Denney.

ABSTRACT

A SYSTEMS APPROACH TO CELLULAR SIGNAL TRANSDUCTION

Jeremy E. Purvis

Scott Diamond

Ravi Radhakrishnan

Vital cellular processes such as growth, gene expression, and homeostasis depend on the correct transmission of molecular signals within and between cells. The vast complexity of these molecular signaling networks has necessitated the use of mathematical methods to model, characterize, and predict cellular responses. The work presented in this dissertation shows how computational methods were used to elucidate two clinically-relevant cellular signaling responses: *(i)* phosphotyrosine signaling through the epidermal growth factor receptor (EGFR), a receptor tyrosine kinase that is commonly overexpressed or structurally altered in human cancers; and *(ii)* phosphoinositide and calcium signaling in human platelets—the key cellular mediators of hemostasis and pathological thrombus formation. The kinetic model of EGFR-mediated signaling in wild-type and mutant cells showed how mutant forms of the receptor use an irregular pattern of tyrosine phosphorylation that preferentially activates the survival oncoprotein, Akt. By quantifying the amount of signal flow through diverging pathways downstream of the receptor, our calculations provided a mechanistic explanation for the clinical observation that therapeutic tyrosine kinase inhibitors can control tumor growth in cells bearing certain EGFR mutations. In the second major study, a kinetic model of ADP-stimulated calcium release in human platelets was used to make precise, quantitative predictions about the molecular makeup and structural properties of the platelet. Specifically, we found that the resting structure of platelets places strong restrictions on several biophysical quantities, such as the resting concentration of free inositol 1,4,5-trisphosphate, the ratio of calcium ATPase pumps to release channels, and the size of the calcium storage compartment. Notably, the model also demonstrated that the irregular calcium spiking behavior observed in single ADP-stimulated platelets is due to the extremely small cellular volume. A novel method for constructing kinetic signaling networks, based on restricting the steady-state properties of the model, is also presented. Future applications and extensions of the systems approach to signal transduction modeling are discussed in the final chapter.

Contents

Acknowledgements	ii
1 Introduction	1
1.1 Cellular Signals	1
1.2 Emergent Properties with Molecular Causes	3
1.3 The “Systems Approach” to Cellular Signaling	5
1.4 Application of the Systems Approach	6
1.4.1 Signaling through the Epidermal Growth Factor Receptor	6
1.4.2 Intracellular Signaling in Human Platelets	7
1.4.3 Novel Methods for Construction and Analysis of Signal Transduction Models	8
2 Phosphotyrosine Signaling in the Epidermal Growth Factor Receptor	9
2.1 Summary of Methods	13
2.1.1 Molecular Dynamics (MD)	13
2.1.2 Molecular Docking	13
2.1.3 Signal Transduction	13
2.1.4 Intramolecular Diffusional Timescales of C-terminal Tail Tyrosines in Auto- and Trans-phosphorylation	16
2.1.5 Spatial Stochastic Model for Endocytotic Vesicle Nucleation	16
2.2 Results	17
2.2.1 Dimer Mediated Activation of EGFRTK	17
2.2.2 Constitutive Activation of the L834R Mutant	18
2.2.3 Ligand and Substrate Binding Affinities for EGFRTK	19
2.2.4 Timescales Associated with Diffusion of Substrate Tyrosines Binding Prior to Auto- and Trans-phosphorylation and with Endocytotic Vesicle Nucleation	20
2.2.5 Differential Signaling through EGFRTK	23

2.3	Discussion and Conclusions	25
3	Efficacy of Tyrosine Kinase Inhibitors in Mutants of the Epidermal Growth Factor Receptor	28
3.1	Methods	30
3.1.1	Systems-level Model Calibration using Genetic Algorithm	30
3.1.2	Model Sensitivity using Principal Component Analysis	31
3.1.3	Differential Signaling through Y1068 and Y1173	31
3.2	Results and Discussion	32
3.2.1	Differential phosphorylation	32
3.2.2	Differential Signaling	32
3.2.3	Robustness of Signaling	33
3.2.4	EC ₅₀ for Inhibition of EGFR in the cell	35
3.2.5	Inhibition of Downstream Activation (EC ₅₀)	35
3.3	Conclusion	37
4	Calcium and Phosphoinositide Signaling in Human Platelets	39
4.1	Introduction	39
4.2	Materials and Methods	40
4.2.1	ADP-stimulated Ca ²⁺ release assay	40
4.2.2	Platelet Image Analysis	41
4.2.3	Model Construction and Simulation	41
4.2.4	Module Calibration and Analysis	41
4.3	Results	42
4.3.1	Model Overview	42
4.3.2	Platelet Ca ²⁺ Balance	43
4.3.3	PI Metabolism	47
4.3.4	Signal Attenuation	49
4.3.5	Receptor Activation	49
4.3.6	Signaling in Platelet Populations	51
4.3.7	Signaling in Individual Platelets	51
4.4	Discussion	54
5	Steady-State Kinetic Modeling	59
5.1	Introduction	59

5.2	Materials and Methods	60
5.2.1	Model Definitions and Requirements	60
5.3	Homeostasis and the steady-state assumption	63
5.4	Results and Discussion	63
5.4.1	Reduction of Modular Kinetic Networks	63
5.4.2	Merging Steady-State Modules	66
5.4.3	Applying the Method To Monitor Cellular Resting States	67
5.4.4	Method Application, Computational Efficiency, and Extensions	69
6	Conclusions and Future Directions	73
6.1	Summary of Results	73
6.2	Extensions and Future Work	74
6.2.1	<i>Extension:</i> Modeling the Platelet Response to Combinatorial Agonists	75
6.2.2	<i>Extension:</i> Quantifying the role of store-operated calcium entry in platelet signaling	76
6.2.3	<i>Future Work:</i> A signal processing model to predict apoptotic response	79
6.2.4	<i>Future Work:</i> Predicting global regulators of transcription in “oncogene-addicted” cell lines	81
6.3	Future Outlook	82

List of Tables

2.1	Summary of multiscale simulations	12
2.2	Reaction mechanisms and parameterization for EGFR phosphorylation reactions . .	15
2.3	Diffusional time for C-terminal tyrosines to reach the active site of EGFR TK	22
3.1	Parametric differences between WT and mutant EGFR systems	30
3.2	Hyper-sensitive parameters in EGFR network	35

List of Figures

2.1	Scheme for hierarchical multiscale simulations	11
2.2	Phosphorylation model for EGFR	14
2.3	Visualization of the stabilizing residues, dimer interface residues, and clinical mutations of both the active and inactive EGFR tyrosine kinases.	18
2.4	Binding modes for ATP and the optimal peptide sequence in the EGFR TK domain	21
2.5	Epsin-induced endocytotic vesicle nucleation.	23
2.6	2-D scan of ERK and Akt activation in wild-type and mutant EGFR	24
3.1	Differential signaling for WT and L834R	32
3.2	Effect of peptide specificity on activated ERK and Akt levels	34
3.3	Relative inhibition of receptor phosphorylation	36
3.4	Inhibition of ERK and Akt activation in systems with normal receptor expression	38
4.1	Schematic of platelet model compartmentalization and reactions	43
4.2	Detailed reaction schemes for platelet signaling modules	44
4.3	Estimation and analysis of platelet Ca^{2+} regulation.	46
4.4	Comparison of PI and P2Y_1 modules to experimental data	48
4.5	ADP dose response for the full platelet model	50
4.6	Modular calibration and feature extraction	52
4.7	Stochastic simulation of a single platelet	53
4.8	Calculation of Ca^{2+} spiking frequency distribution	54
4.9	Increased P2Y_1 copy number enhances Ca^{2+} release in platelet model	55
4.10	Initial condition map for 10 best-fit platelet models	57
5.1	Example structure of model topology, ODEs, and concentration space	62
5.2	Steps in dimensionality reduction of steady-state modules and example from platelet signaling model	63

5.3	Assembly of full model from steady-state modules	68
5.4	Shifts in steady-state profiles caused by kinetic perturbations	70
6.1	Pairwise agonist scanning to predict global calcium response in human platelets . . .	77
6.2	Experimental confirmation of global 6-dimensional platelet calcium response to com- binatorial agonists	78
6.3	Schematic of kinetic model for quantifying platelet calcium regulation with mecha- nisms for store-operated calcium entry	79
6.4	Neural network-based prediction of signal transduction motifs	80
6.5	Information flow for predicting transcription regulatory networks	82

Chapter 1

Introduction

The concept of signal transduction is inherently difficult because it comprises the interaction among a group of objects and not the objects themselves. One cannot simply point to a single member of this interacting group and say: *there* is the signal. Rather, the signal must be understood as a higher-level phenomenon that emerges from the coordinated interactions among the group—a phenomenon that is more easily observed from a distance than by close examination. Under this pretext, then, the most concrete discussion we may have about a signal pertains to its *medium*: those physical entities that are coordinated in space and time to propagate the signal. For an ocean wave, this medium is water. For cells, the signaling medium comprises a vast assortment of biomolecules that range from relatively simple species such as water, ions, small organic molecules, lipids, proteins, and sugars—to large macromolecular assemblies of proteins, polysaccharides, lipids, and nucleic acids. Individually, these molecules display a wide range of physical and chemical properties that have been fairly well documented. When organized as cellular signaling systems, however, they form a staggering array of temporal and spatial configurations with holistic properties that are profoundly and qualitatively different than their molecular components. It is through these networks of molecular interactions that cells are able to send and receive information within their environment.

1.1 Cellular Signals

Cellular signals can be transmitted in a variety of ways. For example, the release of calcium from intracellular stores involves the transport of calcium ions from one compartment of a cell into another (e.g., from the endoplasmic reticulum to the cytoplasm). Once arriving in the new compartment, the calcium ions can physically interact with a new set of biomolecules [130] and

alter their functional properties. In this example, the signal is carried by calcium ions through the intracellular space and depends on biochemical interactions between enzyme and ion. By contrast, the activation of guanine nucleotide-binding proteins (G proteins) by an extracellular hormone is signaled through a successive relay of physical interactions proceeding from the extracellular space, across the plasma membrane, and into the cell's interior. Each of these interactions involves subtle conformational changes within the extracellular ligand, the G protein-coupled receptor, neighboring lipids within the surrounding membrane, and a host of intracellular proteins and nucleotides [101]. Unlike the release of calcium ions, this relay of events occurs without any bulk transport of the molecules involved—a type of signaling that more closely resembles what information theorist John Pierce describes as a “discernible indication of what is not itself directly perceptible” [125]. Whether signals are transmitted directly or indirectly, cellular signal transduction is driven by a sequence of precisely coordinated kinetic and biophysical interactions. Studying the nature of these interactions, on a component-by-component basis, has formed the basis of cellular signal transduction research for the past 30 years.

Because so many vital cellular processes depend on the correct transmission of molecular signals, it is perhaps not surprising that errors in signaling can have devastating pathological consequences. In some cases, dysregulated signals are caused by a single molecular defect. As we shall see in Chapters 2 and 3, one amino acid change in the epidermal growth factor receptor can destabilize the receptor's native conformation and cause changes in its pattern of autophosphorylation. These small perturbations at the receptor level can alter the concentrations of key downstream proteins, thwarting the cell's normal proliferatory and survival responses. In other cases, disease phenotypes arise from the combined effect of multiple molecular alterations. For example, it is well established that disruption of signaling pathways involved in DNA repair and DNA damage checkpoints can lead to an increased mutational load on the cell and the accumulation of additional mutations that cause cancer [100]. Consider the two oncogenes, *c-myc* and *ras*^D. A mutation in either gene alone does not pose an observable threat to mammalian cells; when these mutations are encountered together, however, they can rapidly transform the cell into a malignant state [151]. Signaling defects affecting a wide range of cellular processes that lead to disease have been reported [90], from which two fundamental observations have become universally appreciated: first, that signaling behaviors are driven by molecular events—which are themselves governed by the physico-chemical properties of the molecules involved—and that alterations in the physico-chemical properties of participating molecules can affect the signaling phenotype; second, that disease phenotypes often arise from the distortion—rather than the complete abolition—of “normal” signal transduction. Clearly, understanding the molecular mechanisms that underly signaling behaviors is of paramount

importance if we are to direct signal transduction research toward the treatment and prevention of disease. But, as we shall see, even a complete knowledge of these molecular properties is not sufficient to obtain a full understanding the complexities of cellular signaling.

1.2 Emergent Properties with Molecular Causes

Thus far, we have reasoned that the logical starting point for understanding cellular signal transduction is a mechanistic description of the individual molecular events that produce an observable signaling phenotype¹. Unfortunately, the relationship between individual molecular components and the observable signaling phenotype they comprise is rarely straightforward. Part of this difficulty can be attributed to the sheer complexity of such large networks; they contain so many “moving parts” that their behaviors cannot be predicted by intuition alone. The other difficulty, which is much more subtle, is that the observable signaling behavior is often described, measured, or conceptualized at a different time or length scale than the molecular events that drive it. The term *emergent properties* has been used to refer to such high-level patterns that arise out of the interactions among a set of *basic elements*². In a seminal paper appearing in *Science* in 1999 [13], Bhalla and Iyengar examined a variety of well-known signaling pathways to demonstrate that interacting pathways have complex and nonintuitive emergent properties. Strikingly, their analysis showed that—even with complete knowledge of the kinetic properties of each component in the system—the emergent signaling behaviors were qualitatively unique and unpredictable. To help solidify the concept of emergent properties, let us now consider three examples of emergent behavior from cellular biology:

1. **Lipid vesicle.** Phospholipids, such as those that comprise cellular membranes, will naturally form spherical vesicles because of the properties of lipid membranes. The physical interactions between neighboring lipids in the micelle can alter the fluidity, spacing, and permeability of the membrane. Thus, macroscopic properties such as the diameter of a vesicle—which may be measured in, say, nanometers—are dependent on the chemical properties of its surfactant molecules such as hydrophobicity, degree of saturation, and charge.

¹As an alternative strategy, many studies seek to draw statistical associations between molecular properties and phenotypic outcomes. For example, genome wide association studies relate the occurrence of a particular polymorphism to the risk for disease. While these types of correlations are useful for prediction or prognosis of disease outcomes, they do not explain the molecular basis for the observed outcome thus cannot offer any straightforward opportunities to rationally intervene in pathological scenarios.

²The concept of emergence is quite old and predates its application to cellular biology by over a thousand years. See [39] for an historical overview of emergence.

2. **Molecular oscillation.** In systems of interacting proteins, oscillatory behavior can result from a negative feedback loop between two proteins. A classic example of oscillation occurs when the tumor suppressor protein p53 activates transcription of its negative regulator, Mdm2. Mdm2, in turn, promotes degradation of p53 through ubiquitination [94]. Emergent properties of the system, such as the period of oscillation of activated p53 (measured in hours), are dependent on the various kinetic characteristics of the system such as the rate of protein degradation or the affinity of p53 for the promoter region of Mdm2.

3. **Clotting blood.** The formation of a blood clot in response to endothelial injury is coordinated by clotting factors in flowing blood, procoagulant proteins released from the subendothelium, and platelets. In healthy individuals, a clot must grow large enough to stop bleeding but not so large as to occlude the vessel and obstruct flowing blood. Slight alterations or deficiencies in just one of these coagulation factors, such as the integrin $\alpha\text{IIb}\beta\text{3}$ on the surface of platelets, can prevent normal bridging between platelets and retard formation of a stable clot [118]. Thus, macroscopic properties such as the diameter or density of a blood clot are indirectly but intimately related to the dynamic contributions of thousands of kinetic interactions among participating molecules.

In many cases, the emergent behavior of a system can be easily measured (e.g., the time between p53 peak levels) whereas detailed knowledge about the basic elements in the system (kinetic rate constants and cellular concentrations) is limited or unknown. Inferring these properties is highly desirable but can be extremely difficult. Even if one is able to reproduce, or simulate, the behavior of the system using some nominal set of basic elements, there are usually multiple sets of nominal elements that fulfill this requirement. In mathematical terms, this is known as an *inverse problem* because it seeks to find the unknown cause that produces a known consequence. In general, inverse problems are difficult to solve because they involve several (often, infinite) alternative possibilities that can each generate the same observed behavior. The standard method of solving inverse problems is quite dull: one must “try out” many different possibilities of basic elements, simulate their interactions, and determine which set of elements best explains the emergent behavior. Obviously, this step is greatly aided by computer simulations, which can rapidly alter and test potential solutions. Even for relatively simple signaling networks, however, the number of possible molecular configurations quickly becomes very large and is usually impossible to exhaustively analyze even for the fastest computers.

1.3 The “Systems Approach” to Cellular Signaling

Motivated by the increasing computational power available to end users and the growing wealth of biochemical and genetic data, various computational strategies have been employed to understand and predict the behavior of biological signaling systems. In general, these methods fall into one of two categories based on the intended outcome of the analysis. The first type of analysis is intended to make accurate predictions of a system’s behavior based on some set of initial conditions and quantitative knowledge about the system. These type of studies are often content to establish correlations between, say, a certain genetic mutation and an observable phenotype, or to define disease risks associated with a given phenotype. The second type of systems approach aims to find a mechanistic explanation for the observed behavior of a system. Here, the ability of a model to make accurate predictions is used to retroactively to examine the unique features of that model that are conferring the predictive power.

One very popular method for analyzing signaling networks is ordinary differential equation (ODE) models, [38] which provide excellent mechanistic detail for modeling time-dependent behaviors and are particularly well-suited for kinetic systems. ODEs have been used extensively to model GPCR dynamics [101], the ErbB family of signaling networks [86, 128], cell cycle regulation [38], and various other cellular signaling systems. ODE modeling is used extensively in this work to model signal transduction through the epidermal growth factor receptor (Chapters 2 and 3) as well as phosphoinositide and calcium signaling in human platelets (Chapter 4). If mechanistic detail of a system is lacking, purely “data-driven” approaches such as clustering, principal component analysis, and partial least squares regression have been successfully used to derive meaningful biological insights from large data sets [73]. These methods utilize correlations between measurements within large data sets to extract high-level processes that govern the behavior of the system. These approaches have been used to identify groups of molecules that influence cell decisions, such as whether or not to enter apoptosis [71]. Other statistical approaches to modeling signal transduction, such as Bayesian network analysis [122], have been cleverly employed to discover previously unknown interactions between signaling molecules. These methods work by methodically perturbing signals in individual cells and looking for changes in the dependence between pairs of molecules. Finally, graph theory and transcription factor activation assays have been combined to establish molecular regulators that connect transmembrane signaling and gene regulation [20]. In Chapter 6, we show how artificial neural networks can be used to model cellular signal networks and contain both topological and time-dependent features.

1.4 Application of the Systems Approach

In the chapters that follow, we apply a computational systems approach to study two clinically-relevant mammalian signal transduction systems: signaling through the epidermal growth factor receptor (EGFR), a receptor tyrosine kinase that is commonly overexpressed or structurally altered in human cancers; and phosphoinositide and calcium signaling in human platelets—small, anucleate cells that undergo dramatic signaling in response to vascular injury in order to stop bleeding.

1.4.1 Signaling through the Epidermal Growth Factor Receptor

The first analysis presented in this dissertation is focused on signaling through the epidermal growth factor receptor (EGFR/ErbB1/HER1), a receptor tyrosine kinase that activates a multilayered signaling network mediating crucial pathways leading to cell survival, proliferation, differentiation, and motility [180]. The ErbB signaling network is commonly dysregulated in a large fraction of human cancers. Specifically, ErbB overexpression and/or mutation occurs in almost half of non-small-cell lung cancers (NSCLCs), which comprise the largest portion of cancer-related deaths and will kill about a quarter of a million people in the United States this year [78]. On a molecular level, amplification of EGFR can lead to delayed and prolonged receptor tyrosine phosphorylation, constitutive Akt activation [3], and resistance to anti-tumor drugs [180]. The ErbB signaling network is arguably the most well studied signaling system from a computational systems-biological perspective³ and has spurred a great deal of computational and conceptual advances in the field of systems biology that extend beyond it the scope of just the ErbB family of receptors [95]. An enormous body of literature surrounds this receptor, its signaling, and its relationship to cancer (reviewed by Citri and Yarden [33]).

Our effort to model EGFR signaling was motivated by recent clinical evidence showing that cancer patients who harbor certain mutations in the EGFR kinase domain have a rapid and dramatic response to anti-tumor therapy [106, 120]. Seeking a molecular explanation for this phenomenon, we studied the effects of mutations in the EGFR tyrosine kinase domain on the receptors kinetic behavior and downstream signaling responses [128]. Specifically, we monitored the activation of downstream proteins Akt and ERK as indicators of preferential signaling through two phosphotyrosine residues on the EGFR C-terminal tail. By modeling signal flows through these branching pathways, we showed that the EGFR mutants had increased binding to tyrosine kinase inhibitors (TKIs), enhanced phosphorylation of substrate tyrosine residues, and preferential activation of the Akt signaling pathway, a critical pathway for cell growth, survival, and motility. These findings are

³See the “exhaustive” EGFR model constructed by Oda et al. [119]

discussed in detail in two subsequent chapters focusing on the multiscale aspects of the modeling approach (Chapter 2) and the implications of preferential Akt activation on cell survival (Chapter 3). Recent thinking has likened the ErbB signaling network to a bow-tie-configured network of dynamic interactions that display modularity, redundancy, and control circuitry [33]. Under this framework, identifying the role and significance of drug sensitizing mutations of ErbB receptors would require a systems-level understanding of the signaling network [71–73]. This would prove to be of enormous value in effectively designing multikinase inhibitors for the development of cancer therapies that target several points along the signaling pathway simultaneously [48,77].

1.4.2 Intracellular Signaling in Human Platelets

The second signaling system examined in this dissertation is the ADP-induced calcium response of human platelets. Platelets are small, disc-shaped cell fragments that circulate innocuously in the blood until encountering exposed endothelial tissue or activating agonists. When soluble agonists in the blood bind to their cognate receptors on the platelet surface, this engages a host of intracellular signaling events, including receptor activation, G protein signaling, second messenger generation, calcium release, granule secretion, and cytoskeletal rearrangement [17]. Platelet activation plays a central role in cardiovascular disease, which is the leading cause of mortality worldwide and accounts for 1.74 million heart attacks and strokes, 1.115 million angiograms, and 0.652 million stent placements in the United States each year [103].

Like the epidermal growth factor receptor signaling network, platelets are also prone to aberrant signaling. Defects in platelet signaling can lead to hemophilia or pathological thrombus formation [57]. Pathologies of platelet receptors can impair the platelet response to a single agonist [117, 118, 131]. Defects in G-protein activation result in mild bleeding disorders, abnormal aggregation and secretion responses to a number of agonists, or diminished GTPase activity on activation. Impaired calcium signaling has been observed in several patients with impaired aggregation and secretion, and has been correlated with thrombus formation [18], diabetes [166] and hypertension [47].

In Chapter 4, we address the platelet signaling response to ADP, an important agonist that triggers platelet shape change and aggregation. Through our analysis, we quantify how various molecular mechanisms are integrated to maintain platelet homeostasis and allow responsiveness to ADP. This work presents the first earnest attempt to model intracellular signaling in the platelet featured the ADP receptor, PI metabolism, calcium release, and PKC activity. Existing kinetic information for 77 reactions, 132 fixed kinetic rate constants, and 70 species was combined with electrochemical calculations, measurements of platelet ultrastructure, novel experimental results, and published

single-cell data. The model accurately predicted (*i*) steady-state resting concentrations for intracellular calcium, inositol 1,4,5-trisphosphate, diacylglycerol, phosphatidic acid, phosphatidylinositol, phosphatidylinositol phosphate, and phosphatidylinositol 4,5-bisphosphate, (*ii*) transient increases in intracellular calcium, inositol 1,4,5-trisphosphate, and GqGTP in response to ADP, and (*iii*) the volume of the platelet dense tubular system. A more stringent test of the model involved stochastic simulation of individual platelets, which display an asynchronous calcium spiking behavior in response to ADP. Simulations accurately reproduced the broad frequency distribution of measured spiking events and demonstrated that asynchronous spiking was a consequence of stochastic fluctuations due to the small volume of the platelet. The model also provided insights into possible mechanisms of negative-feedback signaling, the relative potency of platelet agonists, and cell-to-cell variation across platelet populations.

1.4.3 Novel Methods for Construction and Analysis of Signal Transduction Models

The formidable size and complexity of these two kinetic models prompted us to devise a new strategy for constructing large-scale kinetic networks. This method, described in Chapter 5, involves building the model in stepwise fashion, beginning with small “resting” subnetworks that are combined to form larger networks with complex time-dependent behaviors. A key step in the method is the enforcement of a “homeostasis requirement” on all subnetworks, which ensures that the concentration for each species does not change with time even though species may undergo a gradual turnover through coupled reactions. Interestingly, we found that only a minor fraction of potential model configurations were compatible with resting behavior in an example signaling system. These reduced sets of configurations were used to limit the search for more complicated solutions that also captured the dynamic behavior of the system. Using an example model constructed by this approach, we show how a cell’s resting behavior adjusts to changes in the kinetic rate processes of the system. This strategy offers a general and biologically intuitive framework for building large-scale kinetic models of steady-state cellular systems and their dynamics. The final chapter summarizes the main findings and implications of our systems analysis and proposes a new method for constructing and analyzing models of cellular signaling using artificial neural networks.

Chapter 2

Phosphotyrosine Signaling in the Epidermal Growth Factor Receptor

The Erb family of receptors—the epidermal growth factor receptor (EGFR or ErbB1 or HER1), ErbB2 (or HER2), ErbB3, and ErbB4—activates a multi-layered network mediating crucial pathways leading to cell proliferation, differentiation, migration and altered metabolism [33,67,141,180] in response to activation of the receptor by the epidermal growth factor (EGF), transforming growth factor- α (TGF- α), and several other related peptide growth factors [33]. Erb family of receptors are receptor tyrosine kinases (RTKs); RTKs are transmembrane glycoproteins with a cysteine rich ligand-binding extracellular domain, a transmembrane domain important in dimerization, and an intracellular tyrosine kinase domain with a tyrosine rich C-terminal tail [67,141]. Activation of the ligand-binding domain by an extracellular ligand leads to homodimerization or heterodimerization among the members of the Erb family. The process of ligand-induced dimerization is now quite well understood, based on crystallographic studies [22,54]. Precisely how this ligand-induced dimerization event is coupled to, and leads to, activation of the intracellular tyrosine kinase domain has also emerged from recent crystallographic studies [184]. The crystal structure of the EGFRTK domain was first published in 2002, both alone and in complex with an inhibitor erlotinib [159], and displayed several unique features. While structural studies of the kinase domains from other RTKs (and protein kinases in general) have shown them to be catalytically inactive until an activation event such as ligand-induced dimerization causes them to become autophosphorylated within the

activation loop [67] and subsequently become locked into a characteristic active conformation [115], members of the EGFR family appear to be unique in not requiring activation loop phosphorylation for their activity [60]. The crystal structure of an unphosphorylated form of the EGFR kinase domain was found to be constitutively locked in the active configuration [159]. In a recent pioneering crystallographic study, it is suggested that the kinase domain of the EGFR is regulated by a direct allosteric stabilization of the active conformation [184]: specifically, the kinase–kinase contact through an asymmetric dimer interface allosterically promotes the active conformation. Indeed, such an interface is inferred from (and experimentally realized by) the presence of a high concentration of the kinase domain in the crystals. The study also reported a crystal structure of the V924R mutant of EGFR TK in an inactive form which aligned closely with a previously solved structure of wildtype EGFR TK bound to the inhibitor lapatinib [179]. The mutation replaces a hydrophobic residue at the dimer interface with a charged group thereby destabilizing the dimer interface. As a consequence, the EGFR TK domain retains its inactive state and relaxes into an alternative crystallographic lattice which does not support the activating asymmetric dimer interface.

It is well established that alteration or overexpression of EGFR and ErbB2—two (of four) members of the ErbB family of RTKs—is correlated with a variety of clinical cancers [109,149]. Hence, small molecule RTK inhibitors for EGFR TK and ErbB2 RTK are also of significant interest as cancer therapeutic drugs. Small molecule tyrosine kinase inhibitors (e.g., anilinoquinazoline compounds such as gefitinib and erlotinib), which are ATP analogues [5, 6, 75, 113, 133], have been shown to exhibit growth inhibitory effects against a wide variety of human tumors [28–32]. The RTK inhibition approach has shown promise in some clinical trials, but results have been quite mixed [40]. In particular, despite the promise in a small demographic sub-population with significantly advanced disease, clinical responses to gefitinib and erlotinib varied among population samples. Recently, somatic mutations L834R, L837Q, G685S, del L723-P729 ins S¹ in the EGFR TK domain were reported to correlate with the clinical response of gefitinib and erlotinib [23, 89, 106, 120]. In vitro, these EGFR mutants demonstrated enhanced tyrosine kinase activity in comparison to wildtype EGFR and increased sensitivity to inhibition [6, 27, 41, 42, 106, 120, 121, 157]

Although the underlying biochemical basis for the observed drug sensitivity of the mutations is still not clear, the remarkable sensitivity of cell lines carrying gefitinib-sensitizing mutants appears not to be centered around inhibitor or ATP binding affinities, but rather on other biochemical mechanisms regulating the EGFR TK activity and EGF-mediated signaling [48]. In a recent review, the ErbB signaling network is likened to a bow-tie-configured, evolvable network, displaying modularity,

¹there is an alternate clinical numbering scheme used in the literature with an offset of 24 residues according to which these mutations are denoted by L858R, L861Q, G719S, del L747-P753 ins S

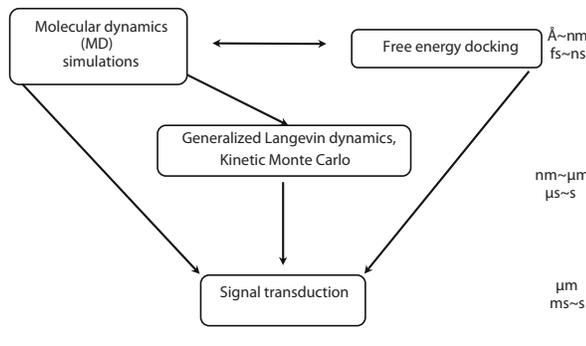


Figure 2.1: Scheme for hierarchical multiscale simulations.

redundancy, and control circuitry [33]. Under this framework, identifying the role and significance of drug sensitizing mutations of ErbB receptors would require a systems-level understanding of the signaling network [71–73]. This would prove to be of enormous value in effectively designing multikinase inhibitors for the development of cancer therapies that target several points along the signaling pathway simultaneously [48, 77]. Evidence emerging from biochemical studies points to differential phosphorylation rates associated with different tyrosine phosphorylation sites in the receptor [1, 80, 157, 178]. Moreover, recent structural studies have indicated that the clinically identified mutations are able to alter regulatory elements within the receptor so as to bypass the dimer-mediated activation step and lead to constitutive activity. Therefore, understanding the effect of such mutations on preferential signaling can be crucial in shaping the signaling response in cell lines derived from cancer patients.

In this chapter, we adopt a multiscale modeling approach (see Figure 2.1 and Table 2.1) to study the dimer-mediated receptor activation characteristics of EGFRTK. Through these modeling approaches, we are able to extend the prior modeling of EGF-mediated signal transduction by considering specific EGFRTK docking interactions mediated by differential binding and phosphorylation of different C-terminal peptide tyrosines on the RTK tail. By modeling signal flows through branching pathways of the EGFRTK resolved on a molecular basis, we are able to transcribe the effects of molecular alterations in the receptor (e.g., mutant forms of the receptor) to differing kinetic behavior and downstream signaling response.

Scale	Method	Biological scope	Input	Output	Experimental validation
Å-nm, fs-ns	Molecular dynamics (MD) simulations	Protein structure simulation	Structure, force-field, system parameters	Trajectory, hydrogen-bond tables, dominant motion, free energy	Crystal structure, mutation studies
Å-nm, fs-ns	Free energy docking	Proteinligand interaction	Protein conformations from MD simulation	Binding modes, binding affinities for inhibitor, ATP, peptide substrates	Crystal structure; K_M , K_I values from affinity studies
nm-m, μ s-s	Generalized Langevin dynamics, kinetic Monte Carlo	Protein intramolecular large-scale motion, proteinmembrane interaction	Elastic Properties of membranes, persistence length of disordered protein chains	Time scale for C-terminal substrate tyrosine binding and for internalization	Micropipette, confocal microscopy, dielectric relaxation
μ m, ms-s	Transient system dynamics	Signal transduction	Initial expression levels, rate laws, kinetic parameters	Time course profiles, signal flow, sensitivity analysis	Pharmacological studies, mutant systems

Table 2.1: Summary of multiscale simulations.

2.1 Summary of Methods

2.1.1 Molecular Dynamics (MD)

Dynamics trajectories of fully atomistic, explicitly solvated systems of wildtype and L834R mutant EGFR are obtained and analyzed for specific stabilizing interactions such as hydrogen bonds and salt-bridges.

2.1.2 Molecular Docking

Molecular docking is used to predict ligand binding in the absence of a ligand-bound crystal structure and functional affinity data. We employ AutoDock, an automated docking tool designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3-dimensional structure [112]. The binding free energy is calculated based on the intermolecular energy between protein and ligands and torsional and solvation free energy of the ligands [112]. We perform a global conformational search using a multiple conformation docking strategy, in which the protein flexibility is taken into account implicitly.

2.1.3 Signal Transduction

EGF stimulation in a cell results in the simultaneous activation of multiple pathways that are functionally interlinked [80,141,142]. In the kinetic model we have employed, signaling through the EGFR is modeled by combining three published models and augmented by our own set of reactions and calculations. Phosphorylation and docking reactions are modeled according to Kholodenko et al. [86]; the MAP kinase pathway reactions are modeled after Schoeberl et al. [147]; Akt and PI3K activation are incorporated into the model as described in Brown et al. [21]. The similar parameterization and topology in these models allowed us to construct a consistent, stable, and comprehensive system with results in good agreement with published experimental data [140]. Seventeen of these reactions are novel to this work and represent enhanced molecular resolution and detail in EGFR activation, phosphorylation, and docking reactions (2.2 and 2.2). Using the system model, we calculate the state levels of well-known downstream indicators of cell growth and proliferation, such as phosphorylated ERK and Akt. Each output state is quantified according to peak response, total (integrated) state level, and the time elapsed until peak response.

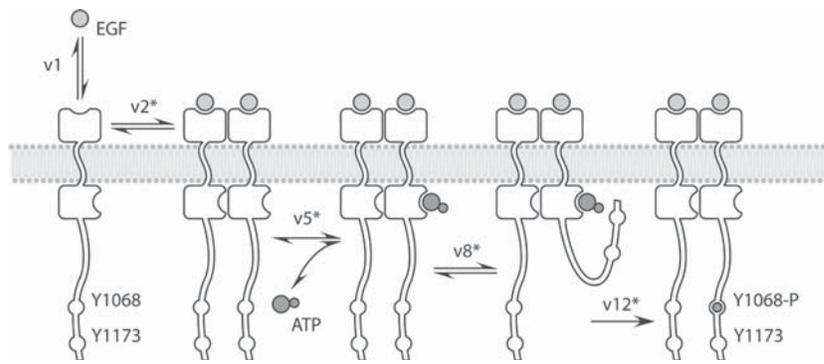


Figure 2.2: Phosphorylation model for EGFR. (v1) EGF binds reversibly to EGFR, which increases the receptors affinity for other ligand-bound receptors (v2). Basal levels of spontaneous dimerization also occur (v3, v4; not shown). ATP binds reversibly to the catalytic site of the receptor (v5). The transient ATP-bound EGFR must then encounter a free unphosphorylated peptide on either cytoplasmic tail in a diffusion-limited step (v8v11). (v12) The receptor catalyzes autophosphorylation at either Y1068 or Y1173 (for clarity, only the Y1068 site is shown). Reactions with an asterisk denote more than one reaction involving similar kinetics but differing substrates (see 2.2).

Reaction	Mechanism	Parameter values	Reference
v1	$[\text{EGFR}] + [\text{EGF}] \rightleftharpoons [\text{EGFR}:\text{EGF}]$	$k_f = 310^{-7} M^{-1} s^{-1}$	[86]
v2	$[\text{EGFR}:\text{EGF}] + [\text{EGFR}:\text{EGF}] \rightleftharpoons [\text{EGFR}_2]$	$k_r = 0.1 s^{-1}$	[86]
v3 ^a	$[\text{EGFR}] + [\text{EGFR}] \rightleftharpoons [\text{EGFR}_2]$	$k_f = 310^{-7} M^{-1} s^{-1}$	[86]
v5	$[\text{EGFR}_2] + [\text{ATP}] \rightleftharpoons [\text{EGFR}_2:\text{ATP}]$	$K_m = 2.410^{-8} M^{-1}$	[19]
v8	$[\text{EGFR}_2:\text{ATP}] \rightleftharpoons [\text{EGFR}_2:\text{ATP}:\text{Y1068}]$	$k_f = 6.5110^5 s^{-1}$	[51]
v10	$[\text{EGFR}_2:\text{ATP}] \rightleftharpoons [\text{EGFR}_2:\text{ATP}:\text{Y1173}]$	$k_f = 1.9810^5 s^{-1}$	[51]
v12	$[\text{EGFR}_2:\text{ATP}:\text{Y1068}] \rightarrow [\text{EGFR}_2^{\text{Y1068-P}}]$	$k_{cat} = 0.29 s^{-1}$	[51]
v14	$[\text{EGFR}_2:\text{ATP}:\text{Y1173}] \rightarrow [\text{EGFR}_2^{\text{Y1173-P}}]$	$k_{cat} = 0.25 s^{-1}$	[51]
v16	$[\text{EGFR}_2^{\text{Y1068-P}}] \rightarrow [\text{EGFR}_2]$	$V_{max} = 4.510^{-7} M^{-1} s^{-1}$	[86]
		$K_m = 5.010^{-8} M^{-1}$	[86]

Table 2.2: Reaction mechanisms and parameterization for EGFR phosphorylation reactions. ^aAn identical rate equation was employed for spontaneous dimerization of EGFR and EGFR:EGF (v4).

2.1.4 Intramolecular Diffusional Timescales of C-terminal Tail Tyrosines in Auto- and Trans-phosphorylation

Auto- as well as trans-phosphorylation of specific tyrosine sites in the C-terminal tail of the receptor involves diffusion of the particular tyrosine residue in the C-terminal tail to the active site of the EGFR TK. We model the tail diffusion as a freely joined chain (FJC) consisting of Kuhn segments [44]. The diffusion coefficient of the tyrosine residue is then calculated using a reptation model [44]. The reciprocal of the diffusional time of a specified tyrosine residue to reach the active site is taken as the k_{on} for substrate binding.

2.1.5 Spatial Stochastic Model for Endocytotic Vesicle Nucleation

Clathrin mediated endocytosis is a major mechanism for the down regulation of activated receptor proteins. The assembly of a clathrin coat proceeds via a polymerization of clathrin trimers [143], which is observed to coat the endocytotic vesicle. However, it is believed that the clathrin coat alone may not be capable of inducing the required curvature in the membrane. Another protein, epsin, in fact plays a crucial role in inducing the curvature of the membrane [52, 56]. The protein epsin has binding sites in the clathrin lattice and also directly interacts with the membrane. This dual interaction together with the dual interaction of an adaptor protein Ap180 with clathrin and the membrane (Ap180 tethers the clathrin coat to the membrane) is believed to induce the vesicle nucleation and anchor the clathrin lattice to the budding vesicle. Recently, we developed the KMC-TDGL (kinetic Monte Carlo- time-dependent Ginzburg Landau) heterogeneous multiscale approach for studying membrane dynamical processes by combining two different phenomenological theories [172]. Simulations performed using this approach yields the time evolution of the membrane undulations and deformation in response to the diffusion of membrane bound epsins and a timescale for the nucleation of an endocytotic vesicle. We calculate the timescale for epsin-induced endocytotic vesicle nucleation from our membrane simulations performed using the KMC-TDGL approach [172]. We combine this timescale for nucleation with the timescales for signal transduction calculated using the network simulations to estimate an overall timescale for phosphorylated receptor internalization.

2.2 Results

2.2.1 Dimer Mediated Activation of EGFRTK

We performed two 10 ns molecular dynamics (MD) simulations of the EGFRTK system in the active and the inactive conformations, respectively. Based on a comparison of the crystal structures—active: erlotinib-bound, wildtype [159], AMP-bound wildtype [184]; inactive: lapatinib-bound wildtype [179], V924R mutant [184]—we conclude that the conformational switching from an inactive to an active conformation involves a rotation of the α C-helix of EGFRTK and the shifting of the activation loop (A-loop) residues to make way for peptide and ATP binding. The recent study of Zhang et al. [184] has implicated a dimer-mediated activation mechanism for the EGFRTK. That is, loss of interactions upon RTK dimerization leads to switching of the conformational state of the enzyme from inactive to active conformations. On this basis, we hypothesize that specific interactions (namely, H-bonds and salt-bridges) between the residues of the α C-helix and those of the A-loop need to reorganize in the course of the conformational switching from inactive to active states. To elucidate the nature of these interactions, a hydrogen bond analysis focusing on the residues surrounding the A-loop, α C-helix, catalytic loop, and nucleotide-binding loop was carried out using CHARMM6 and VMD30 programs. The MD trajectories were first analyzed in CHARMM with a hydrogen bond cutoff of 2.4 Å and a cutoff angle of 150°. We generated a list of hydrogen bonds that were present in at least 60% of the trajectory. These hydrogen bonds were then visualized in VMD and any non-significant bonds were removed to reveal the stabilizing H-bonds. During the H-bond analysis, acidic and basic residues formed strong hydrogen bonds, similar to salt-bridges. The A-loop and α C-helix conformations were stabilized by internal hydrogen bonds (i.e., both the residues participating in the hydrogen bond are within the loop/helix), as well as external hydrogen bonds (i.e., one residue is in the loop/helix, the other is not). The residues participating in these stabilizing interactions are predominantly acidic and basic residues. A comparison of the specific interactions showed that the number of external interactions stabilizing the A-loop and α C-helix in the active conformations far out-numbered those in the inactive conformations (Figure 2.3 and Table 3 of ref. [102]). Intriguingly, the small number of stabilizing interactions holding the kinase in the inactive are predominantly located proximal to the dimer-interface residues (Table 3), which lends further credibility to the allosteric activation mechanism [184] of the wildtype. Namely, that loss of dimer-interface residues will alter the surrounding stabilizing bond network, leading to the destabilization of the inactive conformation and triggering the conformational shift to the active conformation.

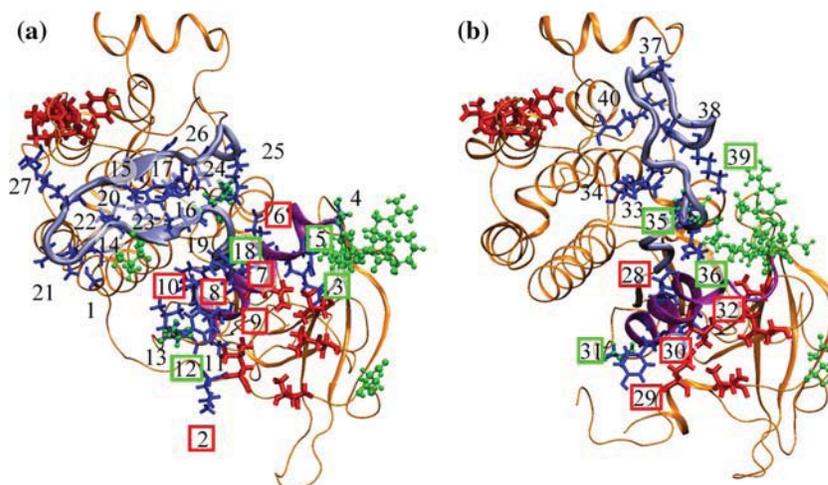


Figure 2.3: Visualization of the stabilizing residues external to A-loop and α C-helix (shown in blue), dimer interface residues (shown in red), and clinical mutations (shown in green) of both the active (panel a) and inactive (panel b) EGFR tyrosine kinases. The A-loop is highlighted in sky blue while the α C-helix is shown in purple, the remainder of the protein is colored orange. The external stabilizing residues for the EGFR tyrosine kinases are numbered from 140, with 127 corresponding to residues in the active state while 2840 correspond to the stabilizing residues in the inactive state. The stabilizing residues that are potentially affected by dimerization are highlighted in red boxes, and the residues that are potentially affected by the clinical mutations are highlighted in green boxes (See also Table 3 of ref. [102]). The inactive conformation has fewer stabilizing bonds holding the conformation stable: 13 stabilizing residues as opposed to 27 stabilizing residues. Therefore, the loss of interactions triggered by either mutation or dimerization will have a greater effect upon the stability of the inactive kinase and allow a conformation change. The numbered labels correspond to the residue numbers as follows: 1:672 2:679 3:721 4:728 5:732 6:734 7:737 8:738 9:740 10:741 11:743 12:744 13:745 14:808 15:810 16:812 17:832 18:834 19:836 20:838 21:840 22:843 23:845 24:848 25:851 26:867 27:932 28:738 29:740 30:742 31:744 32:753 33:812 34:813 35:834 36:836 37:846 38:848 39:851 40:865.

2.2.2 Constitutive Activation of the L834R Mutant

The clinically relevant mutations del724-729 and L834R are located in distinct parts of the enzyme but both mutations lead to enhanced receptor activity [25, 106, 120, 157]. Our delineated pattern of stabilizing interactions in Table 3 in ref. [102] serves as a platform for unifying the effects of these mutations at a structural level. The small number of stabilizing interactions holding the kinase in the inactive conformation also overlaps with the residues associated with several clinically relevant mutations. The deletion mutant interferes directly with the dimer interface, and both the R substitution of L at 834 as well as R substitution of Q in 837 destabilize the specific (external H bonds) interactions associated with A-loop and α C-helix in the inactive but not the active conformation. Thus, these mutations have a deleterious effect (see Table 3 of ref. [102]) in stabilizing A-loop and α C-helix conformations in the inactive state—hence, they stabilize the

active conformation. Indeed, in one of our MD trajectories initiated from the active conformation of the wildtype monomer RTK, we observed the motion of the enzyme towards the inactive state in 10 ns. In this case, a key interaction between two residues in the α C-helix and the A-loop (salt-bridge between GLU851 and LYS734, see also Table 3 of ref. [102]) was broken. In stark contrast, a 12 ns MD trajectory of the L834R mutant system (under similar conditions) remained stable in the active conformation. The collective evidence from these observations points to a constitutively active L834R mutant RTK system (i.e., activated in monomer and dimer states in the presence or in the absence of an EGF ligand) and a dimer-mediated activation of wildtype RTK (i.e., activated only in the dimer state in the presence of the EGF ligand).

Experimental observation of constitutive activation in the gefitinib-sensitive EGFR mutants has been recorded independently in several studies, which report significantly elevated basal phosphorylation (in the absence of the stimulating ligand) of the mutant systems in comparison to the wildtype [25, 26, 120, 157]

2.2.3 Ligand and Substrate Binding Affinities for EGFRTK

We determine the ATP and erlotinib binding modes and binding affinities based on our multiple conformations docking strategy and analyze several low-energy clusters of ligand conformations. For both wildtype and L834R mutant EGFRTK systems, one lead erlotinib binding conformation is found to be closely aligned with that in the crystal structure: the hydrogen bonds between the N1 of the quinazoline and Met769 are present and so is the interaction between Thr766 and the quinazoline nitrogen atom (N3). The binding modes for ATP and erlotinib are also consistent with the existing pharmacophore model for ATP binding to kinases [164, 165, 183]. Significantly, a second binding mode, which has a higher affinity to the mutant kinase L834R, is revealed in our calculations. This mode still blocks the ATP binding pocket and in addition shows an additional interaction between erlotinib and the R834 residue. The effect is a higher calculated binding affinity for erlotinib binding to L834R (-9.4 \pm 0.4 kcal/mol) in comparison to wildtype (-7.3 \pm 0.5 kcal/mol). In order to explore the C-terminal peptide-substrate-bound pre-catalytic conformation, we also dock EGFRTK substrate peptides to the catalytic site. First, we focus on an optimal peptide sequence EEEYFEL as a control. This optimal sequence is selected for a maximal value in k_{cat}/K_M among EGFRTK substrate peptides [156]. Based on 5001000 different genetic algorithm runs for docking of this peptide to EGFRTK, we select the best binding conformation for binding as one that yields the lowest binding free energy simultaneously with the correct orientation for the tyrosine binding into the active site. The latter criterion is essential in ensuring that k_{cat} is not significantly compromised in obtaining a stronger binding. We depict in Figure 2.4 the peptide

conformation bound to the active site, which is seen to be ideally poised for a two-metal-ion catalyzed phosphoryl transfer from ATP to the bound tyrosine [110,160]: i.e., the relative positions of the ATP, Mg^{2+} ions, and the catalytic aspartates are in good alignment. Encouraged by this result, we performed peptide docking simulations for two substrate peptides (each being seven amino acids long) corresponding to the Y1068 (VPEYINQ) and Y1173 (NAEYLRV) sites of the C-terminal tail for both wildtype and L834R mutant structures. Based on a similar criteria of lower binding free energy and optimal catalytic geometry, we calculated the peptide binding affinity corresponding to sites Y1068 and Y1173 for wildtype and L834R mutant systems. Upon mutation, the calculated K_M for the Y1068 peptide decreased 20-fold (calculated $\Delta\Delta G$ [wildtype \rightarrow L834R] = -2.2 kcal/mol), while the calculated K_M for the Y1173 peptide increased 4-fold (calculated $\Delta\Delta G$ [wildtype \rightarrow L834R] = $+0.9$ kcal/mol). Analysis of the binding modes reveals that the relative alignment of the tyrosine with respect to the catalytic site is ideal (2.2 \AA in the observed distance ξ between Tyr OH oxygen and Asp 813 $O_{\delta 2}$) for the two peptide tyrosines binding to wildtype, with the two mutant systems showing a 0.15 \AA deviation. The differences can translate into differences in k_{cat} values associated with the tyrosine phosphorylation. Assuming a harmonic energy profile along the reaction coordinate ξ with a force constant K_ξ , the work done in reducing this distance is given by $1/2K_\xi\xi^2$. The value of K_ξ is obtained by recording the fluctuations of ξ in the MD simulations ($K_\xi = k_B T / \sigma_\xi^2 = 0.77 \text{ kcal/mol/\AA}^2$ for the wildtype and $1.01 \text{ kcal/mol/\AA}^2$ for L834R), where σ_ξ is the standard deviation associated with the fluctuations in ξ in the trajectories). The normalized values of k_{cat}/K_M can then be compared for the four systems at the same value of $\xi = 2.2 \text{ \AA}$; the relative increase in k_{cat}/K_M^{Y1068} for the L834R in comparison to the wildtype is 20-fold (increase by 1-order of magnitude), while the relative decrease for Y1073 is 200-fold (decrease by 2-orders of magnitude). This differential sensitivity translates into differences in phosphorylation of the associated tyrosine sites, which ultimately leads to differential signaling and down stream activation in the EGFR network: the Y1173-phosphorylation is associated Shc mediated ERK activation and Y1068 phosphorylation is associated with Akt activation [80,157].

2.2.4 Timescales Associated with Diffusion of Substrate Tyrosines Binding Prior to Auto- and Trans-phosphorylation and with Endocytotic Vesicle Nucleation

C-terminal Tail Diffusion

We calculate the diffusional time scale for each of the seven tyrosine residues present on the C-terminal tail of EGFRTK to reach the active site Table 2.3. The timescales for diffusion are in μs

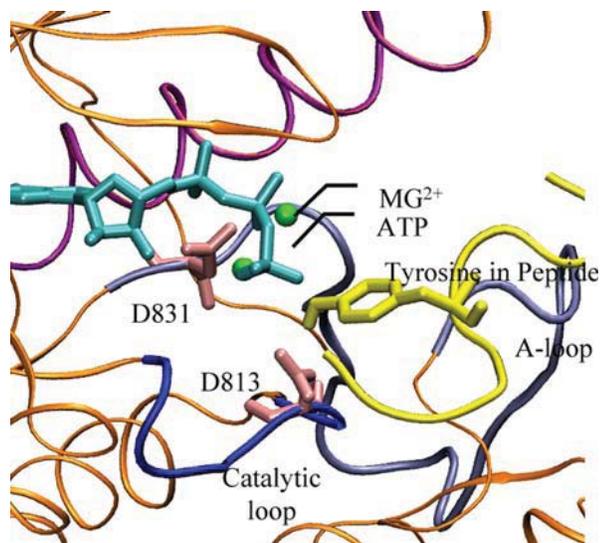


Figure 2.4: Binding modes for ATP (cyan) and the optimal peptide sequence (yellow) in the EGFR TK domain. The geometry of the active site is consistent with that prescribed for an ideal two-metal-ion catalyzed mechanism and reveals a state that is poised for the phosphoryl transfer reaction. The alignment of the two divalent Mg^{2+} ions as well as the catalytic aspartates D813 and D831 are also shown.

and are fairly similar for the different tyrosine sites to bind to the active site prior to auto- (head) or trans- (tail) phosphorylation. For each case, the reciprocal of the corresponding timescale is taken as the on-rate (k_{on}) of the tyrosine binding to the active site. The off-rate is calculated as $k_{off} = k_{on} \times \exp(+\Delta G/k_B T)$, where ΔG is the binding free energy of the substrate to the active site; here it is assumed that the binding process is diffusion-limited with no significant additional barriers. The values for ΔG for each C-terminal tyrosine is obtained from the analogous data on peptide substrate affinities K_M published in the literature [19, 51]: $k\Delta G/k_B T = -\ln(K_M)$. These calculated values are used in the model for signal transduction described in Section 2.1.3 and provided in Table 2.2.

Endocytotic Vesicle Nucleation

In our earlier work, we identified different regimes conducive to nucleation of an endocytotic vesicle by exploring a range of parameters associated with the membrane adsorbed proteins, such as surface density of the protein on the membrane, and the intrinsic curvature of the membrane induced by the protein-membrane interaction [172]. Here, we report our results specifically for epsins interacting with the membrane. Consistent with the experimental values of Ford et al., we assume that the intrinsic membrane curvature induced by the epsin molecules falls in the range of $10\mu m^{-1}$. We further assume that the distance from a given epsin molecule that this curvature effect decays

Tyrosine residue	Diffusional time to the active site (μs)	
	Head EGFRTK (auto-phosphorylation)	Tail EGFRTK (trans-phosphorylation)
Y992	0.34	0.65
Y1045	1.72	2.54
Y1068	2.64	3.68
Y1086	3.49	4.7
Y1101	4.26	5.61
Y1148	7.26	9.08
Y1173	9.16	11.22

Table 2.3: Diffusional time for C-terminal tyrosines to reach the active site of EGFRTK.

is four times the epsin diameter. A depiction of our findings for an epsin surface density (on the membrane) of 60 per μm^2 is given in Figure 2.5, in which in the time course of 250400 s, the nucleation of a vesicle is observed (see snapshots in panels BD). Panels E and F provide the corresponding spatial and orientational correlations [172]. The necessary condition for nucleation is the persistence of orientational (hexagonal) ordering of the membrane-adsorbed epsin molecules (see caption) [172]. Our results have suggested an interesting mechanism for the epsin-mediated membrane invagination. Namely, the invagination occurs through a nucleation event, which involves the colocalization and hexagonal ordering of several epsin molecules on the membrane. Based on our results (Figure 2.5), the timescale for nucleation of such an event is 250400 s.

As described in Section 2.1.3, receptor activation, which is eventually followed by receptor internalization (via endocytosis), results in triggering many signal transduction cascades—one of which is the PLC- β pathway [13, 180]. The PLC- β pathway is accompanied by a transient release of calcium ions from intracellular stores. Models for the transient release of calcium have been developed on the basis of several published experimental studies [90, 154]. We work under the hypothesis that the signal transduction in the endocytosis pathway is coupled to membrane deformation via two functional interactions: (1) the direct induction of curvature via the interaction of the membrane lipids with epsin [52, 123, 144, 158]; and (2) through the assembly of the clathrin coat [144, 181], which is triggered by an increase in the local concentration of Ca^{2+} ions released in the PLC- β pathway [13]. In the latter case, above a threshold value of Ca^{2+} concentration the clathrin monomers spontaneously self-assemble to form a lattice [144, 181]. This leads to the following mechanism for endocytosis. The transient increase in calcium concentration at the site of the activated receptor helps promote clathrin polymerization and coat assembly. The adapter protein Ap180 tethers the clathrin lattice on the membrane. The vertices of the hexagonal lattice of the clathrin coat provide binding sites for epsins. Thus, in the vicinity of the activated receptor, the membrane

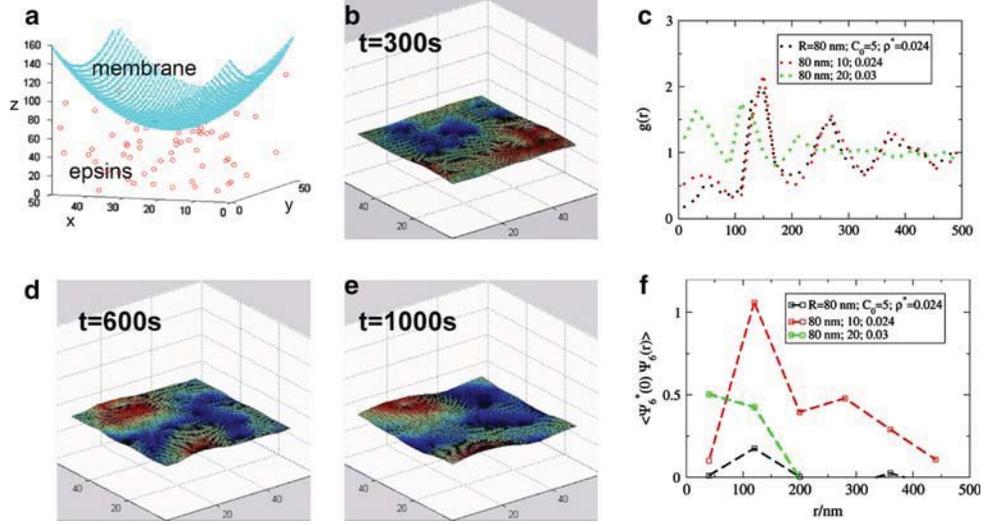


Figure 2.5: (a) Depiction of the system of epsin molecules interacting with the membrane. (bd) Snapshots of the system at different time points (the epsin positions are not shown for clarity). The blue regions are recessed and red regions are elevated in the membrane. The nucleation of a budding vesicle in the membrane is evident in panel d. (c) Spatial correlation function for the membrane adsorbed epsins. (f) Orientational correlation function for the membrane adsorbed epsins. The correlation functions are provided for 3 simulation runs at different epsin densities (ρ^* , dimensionless) and values of the induced curvature (c_0 expressed in μm^{-1}). The value of a dimensionless density ($\rho^* = 0.024$ corresponds to 60 epsins per μm^2). The snapshots correspond to the red symbols in panels e and f ($\rho^* = 0.024$ and $c_0 = 10\mu\text{m}^{-1}$). For this case, there is persistence of orientational correlations (hexagonal ordering), see panel f. For the other two cases in panels e and f (for which snapshots are not shown), no nucleation and no significant orientational ordering is observed.

adsorbed epsins assume a spatial conformation with persistent orientational correlations induced by the clathrin lattice. This provides the condition for a nucleation event leading to formation of the endocytotic vesicle. (We note that according to this mechanism, unactivated receptors can also internalize; however, in the absence of the transient calcium surge, which only occurs upon receptor activation, the clathrin coat assembly may represent a slower step).

An estimate for the timescale of receptor internalization $\tau(\text{internalization})$ is therefore obtained by the sum of timescales for receptor phosphorylation at Y1045, PLC- β activation, calcium release, clathrin coat assembly, and nucleation of epsin-mediated vesicle formation.

2.2.5 Differential Signaling through EGFRTK

By employing the deterministic model described in Section 2.1.3 (see Figure 2.2 and Table 2) we have calculated transient evolution of the EGF-mediated signaling network under a variety of initial conditions and parameter values. In particular, we have focused on two pools of normal

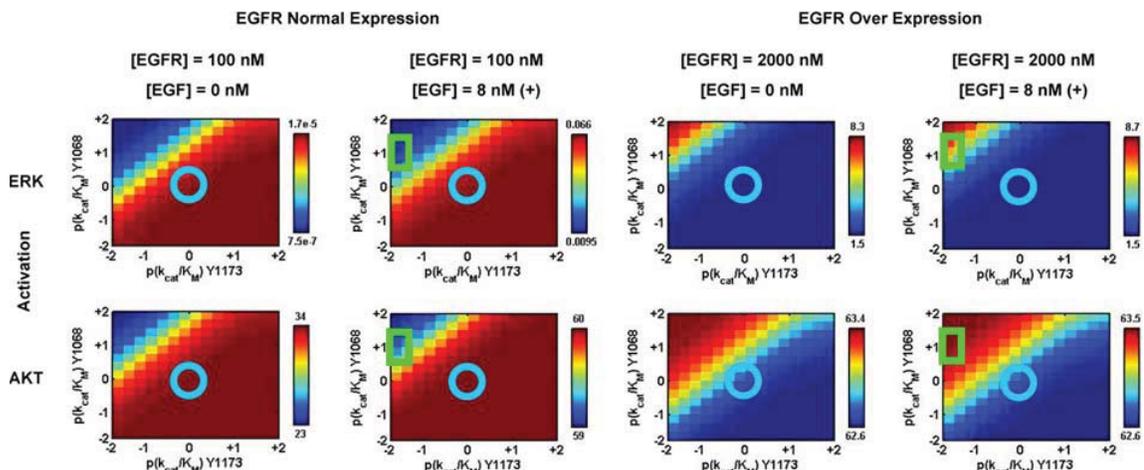


Figure 2.6: Calculated ERK and Akt phosphorylation levels in units of nM (peak levels over the time course of 1000 s) under serum starved (EGF-) and serum cultured (EGF+) conditions for cell types with normal EGFR expression and EGFR overexpression. In each panel the Y1173 rate of phosphorylation (k_{cat}/K_M) is varied along the x-axis (value relative to wildtype EGFR TK) and the Y1068 rate of phosphorylation (k_{cat}/K_M) is varied along the y-axis. In each panel, the blue circle denotes the wildtype levels and the green square denotes those relevant to the L834R mutant cell lines.

([EGFR] = 100 nM or 33,000 receptors per cell) and overexpressed ([EGFR] = 2000 nM or 660,000 receptors per cell) EGFR concentrations, each in the absence (serum starved, [EGF] = 0 nM) and presence (serum cultured, [EGF] = 8 nM) of the EGF ligand. Since our molecularly resolved model (Figure 2.2) is currently focused on differential signaling through Y1068 and Y1173 phosphorylation sites of EGFR TK, we track the integrated response of ERK and Akt phosphorylation levels in our simulations. Each output state was quantified according to the peak level of phosphorylation over the simulated time of 1000 s. Specifically, we perform a two-dimensional scan over k_{cat}/K_M values associated with Y1068 and Y1173 phosphorylation in which the respective k_{cat}/K_M values are allowed to deviate from their default (wild-type) value over a logarithmic range of five log units. The result is a two-dimensional matrix in which each element represents output from a single simulation involving a unique pair of parameters (Figure 2.6).

The simulated response of the EGFR network also enables us to estimate a timescale for receptor internalization $\tau(\text{internalization})$. The typical timescale for EGFR TK phosphorylation is in the range of 10 s, while that for PLC- β activation is 50 s. The calcium release from intracellular stores through the activation of IP₃R (IP₃ receptors) is reported to occur within a short timescale of 10 s [154]. While the timescale for clathrin coat self-assembly under high calcium concentration is not known, it is assumed here that this is a rapid process. According to our results from the spatial stochastic simulations (Section 2.2.4), the epsin mediated endocytotic vesicle nucleation in

the presence of the clathrin coat is calculated to occur in a timescale 250400 s. Further, assuming that the stabilization of the budding endocytotic vesicle and the vesicle pinch-off process catalyzed by dynamin are not rate-limiting, the estimated time for vesicle internalization upon EGF ligand binding is ≈ 450 s (≈ 8 min). We note that these timescales from our simulations are ball park estimates and we have not yet attempted a careful comparison with timescales from experimental measurements.

2.3 Discussion and Conclusions

Based on an integration of a large body of knowledge accumulated over the years, the modular nature of EGFR-mediated signaling has become apparent [33, 71]. Through a hierarchical multiscale modeling scheme, we have realized a molecular resolution to this system-level view that can potentially augment our predictive as well as comprehensive capabilities and rationalize the changes in early signaling characteristics of altered and mutant cell lines. Our overall findings are summarized below.

Analysis of stabilizing interactions in our molecular dynamics simulations have corroborated a dimer-mediated activation mechanism for the EGFR TK. That is, loss of key stabilizing interactions in the inactive conformation of the kinase upon dimerization results in destabilizing the inactive state and triggering a conformational change to the active state. Moreover, our results point to a constitutively active L834R mutant. That is, even in a monomer state, the RTK of the L834R mutant system is stabilized in an active conformation and can trigger signaling through the EGFR network by auto-phosphorylation of its C-terminal tail tyrosine residues in an EGF (ligand) independent manner. These results are consistent with the existing biochemical data [25, 106, 120, 157, 184] that have qualitatively shown that the phosphorylation of C-terminal tail tyrosines of the L834R mutant in the absence of EGF stimulation is comparable to wildtype values in the presence of EGF stimulation at EGF concentrations of 30100 ng/ml.

Our results from the free energy docking simulations suggest enhanced binding of the inhibitor erlotinib to the L834R mutant system in comparison to wildtype binding. Our results also reveal preferential binding and phosphorylation of certain C-terminal tyrosine substrates (but not others) to the L834R mutant in comparison to the wildtype: that is, the calculated change in the turn over rate (k_{cat}/K_M) upon the L834R mutation is markedly different for the Y1173 peptide in comparison to the Y1068 peptide. This results in a differential sensitivity at the receptor level. Specifically, the estimated increase in k_{cat}/K_M for the Y1068 peptide is ≈ 10 -fold in comparison to a decrease of ≈ 100 -fold for the Y1173 peptide. Thus, we expect to see a net increase in phosphorylation of Y1068

and a net decrease in phosphorylation of Y1173 for the L834R mutant, in comparison to wild-type levels. The biochemical results of Sordella et al. [157] and Chen et al. [25] corroborate this view qualitatively; in these studies, the immunoblots measuring the degree of phosphorylated receptor at the Y1173 site show a modest decrease for the L834R mutant in contrast to a strong increase for the Y1068 site, both in comparison to their respective wild-type measurements.

Our systems level simulation illustrates how the altered affinities at the receptor level for the wildtype and L834R translates into changes in the downstream response. In Figure 2.6, we explored how the variation in Y1173 and Y1068 phosphorylation affects the peak responses of ERK and Akt for two different EGFR concentrations, each in the absence or presence of EGF stimulation. The calculated responses for the wildtype values are highlighted by the blue circles at the origin of each of the panels. The responses for L834R are identified by the green square centered at a $(k_{cat}/K_M)/(k_{cat}/K_M)_{wildtype}$ value of 1/100 (-2 in \log_{10} units) for Y1073 and 10 ($+1$ in \log_{10} units) for Y1068. Owing to the constitutively active characteristic of the L834R mutant, the signals relevant for the L834R system (even in the absence of EGF) are analogous to those in the panels with EGF stimulation. As indicated by the color maps in these two scans, increasing the degree of phosphorylation of the Y1173 site (positive x-direction) leads to an overall increase in ERK response. This finding is consistent with our model which is based on Y1173 being an important docking site for the Shc and Grb-2 adaptors that lead to Ras activation and subsequent ERK phosphorylation. For normal EGFR expression, the calculated ERK response in the absence of EGF is low for the L834R mutant (0.01 nM) but negligible (approximately zero) for the wildtype. In the presence of EGF the ERK response for L834R is 7-fold lower than that for the wildtype. For over expressed EGFR, the ERK activation is 7-fold higher in L834R compared to wildtype both in the presence as well as in the absence of EGF stimulation. Thus, ERK stimulation is higher in EGFR overexpressed cells. The Akt response relative to the wildtype is 2-fold higher for the mutant cells under no EGF stimulation and about the same under EGF stimulation for normal EGFR expression. The Akt response does not seem to be affected by mutation for high EGFR expression either in the presence or absence of EGF.

Our calculated responses for ERK and Akt signaling for normal EGFR expression agree with the qualitative experimental observations of Sordella et al. [157] and Tracy et al. [163]: namely, that the ERK stimulation levels decrease and the Akt levels are maintained for the mutant cells with normal EGFR expression levels. This preferential activation of Akt in L834R mutant cell lines observed in these experiments is consistent with the systems level predictions from our simulations. Moreover, our calculations predict that the preferential Akt activation is not featured in EGFR overexpressed cells.

In light of the overall agreement between simulations and experiment, we can begin to explore the origins of inhibitor sensitivity in L834R mutant cell lines. The enhanced sensitivity of the L834R mutant to erlotinib arises from a combination of several effects: (1) the mutant system has a higher binding affinity to the inhibitor; (2) the constitutive activity of L834R renders the mutant to be a more potent target for therapeutic inhibition; (3) the differences in the predicted downstream response (between the wildtype and the mutant) accounted for by early signaling events can potentially lead to a divergent long-term behavior in the two systems causing a differential sensitivity to inhibition. More conclusive inferences from our modeling results can only be obtained after we consider the effect of other receptor states (namely, one in which Y845 is phosphorylated) and interactions with other Erb partners (see below) on down-stream signal response.

In conclusion, we have described the role of a hierarchical multiscale computational approach in resolving at a molecular level the systems-level model of EGFR-mediated signal transduction, to help rationalize the collective information emerging from biochemical, cell biology, and clinical studies. Our approach enables us to transcribe differences in molecular mechanisms as well as functional activity originating at the single molecule level into tangible differences in early signaling events. Therefore, subject to the well appreciated limitations of computational modeling, this predictive capability is useful in differentiating signaling characteristics of mutant cell lines from the native wild-type case. This molecular-based refinement of the overall EGFR signal transduction model is a first attempt at untangling the differential nature of early signaling events triggered by receptor activation and their consequences in shaping the signal transduction in altered cell lines derived from different pathologies. A complete model, which is currently being pursued, will not only require resolving the differential characteristics of all of the tyrosine phosphorylation sites in EGFR and their associated substrate recognition properties in EGFR, but also the extension to other Erb family members in the context of homo- and hetero-dimers. Transactivation of EGFR occurring through ligand-induced receptor heterodimerization [171] combined with a potential for differential signaling adds a palette of finer control elements in the ErbB family signaling network. Indeed, recent studies have identified possible mutations in ErbB2 that may correlate with drug sensitivity in different cancers [97,98]. The computational tools described here are ideal for assessing the likely effect of novel EGFR and ErbB2 mutations and determining whether the drug-sensitizing mutations implicated in non-small-cell lung cancer also occur in other cancers. We believe that our model driven approach will in the long-term significantly impact the optimization of future small molecule therapeutic inhibition strategies.

Chapter 3

Efficacy of Tyrosine Kinase Inhibitors in Mutants of the Epidermal Growth Factor Receptor

In Chapter 2, we employed a hierarchical multiscale computational strategy to study the dimer-mediated receptor activation characteristics of the Erb family receptors, through which we were able to transcribe the effects of molecular alterations in the receptor (e.g., mutant forms of the receptor) to differing kinetic behaviors and downstream signaling responses [102]. Here, we extend this approach to study the effects of EGFR inhibition through TKIs. By employing molecular docking in combination with network modeling, we are able to quantify changes in the EC_{50} of receptor phosphorylation (i.e. 50% inhibition in the cellular context) and EC_{50} for the inhibition of downstream markers (ERK and Akt) upon treatment with the TKI erlotinib, in cell-lines carrying both wildtype (WT) and mutant forms of the receptor. Based upon our results, we conclude that the remarkable efficacy of the inhibitor erlotinib in the L834R mutant cell line can be attributed to a relative gain in efficiency over the WT in inhibiting the Akt response.

How such constitutive activation is precisely linked to increased sensitivity to inhibition and the efficacy of some TKIs remains an interesting question. Recently, Noro et al. [116] reported that the majority of inhibitor hypersensitive cell lines showed Akt phosphorylation without ligand stimulation while a small fraction of the resistant cell lines displayed this characteristic. On this basis,

they concluded that Akt phosphorylation without ligand stimulation may play a key signaling role in TKI hypersensitivity. Sharma et al. [148] proposed the concept of an oncogenic shock, which attributes the hypersensitivity to inhibition in such cancer cell lines to differential attenuation rates of pro-survival (triggered by Akt phosphorylation) and pro-apoptotic signals emanating from EGFR following its inactivation upon inhibitor treatment. Their surmise, like the Noro et al. finding, is based on preferential Akt activation, and relies on the assumption that pro-survival signals such as Akt-(p) dissipate quickly upon oncoprotein inactivation, whereas pro-apoptotic signals linger sufficiently long to commit the cell to an apoptotic death. This differential effect may contribute to the rapid and dramatic clinical responses observed in some cancer patients treated with selective TKIs. Hendriks et al. [65] sought to provide a mechanistic basis for enhanced Akt signaling in gefitinib hypersensitive cell lines and suggested on the basis of experimental and computational results that the gefitinib sensitivity is a marker of a reliance on Akt signaling for cell survival that may be brought about by impaired EGFR internalization in the inhibitor-sensitive cell lines.

Here we describe another possible mechanism that leads to preferential Akt signaling. Specifically, by introducing a molecularly resolved branched network systems model (the molecular resolution is introduced through a hierarchical multiscale computational strategy to transcribe the effects of molecular alterations in the receptor, e.g., mutant forms of the receptor, to differing downstream short-term signaling responses (see Chapter 2)), to study the effects of EGFR signaling and inhibition in WT and mutant cell lines, we are able to quantify: (1) differences in short-term signaling in downstream ERK and Akt activation, (2) the changes in the cellular inhibition EC_{50} associated with receptor phosphorylation (i.e., 50% inhibition of receptor phosphorylation in the cellular context), and (3) EC_{50} for the inhibition of activated downstream markers ERK-(p) and Akt-(p), where (p) denotes phosphorylated, upon treatment with the TKI erlotinib in cell lines carrying both WT and mutant forms of the receptor. This model also suggests a possible mechanism for preferential Akt activation in the oncogenic mutants of EGFR implicated in non-small-cell lung cancer as well as an enhanced efficacy of the inhibitor erlotinib in ablating the Akt response. Using a simple phenomenological model to describe the effect of Akt activation on cellular decisions, we discuss how this preferential Akt activation is conducive to cellular oncogene addiction and how its disruption can lead to the observed hypersensitivity of the inhibitors. We also illustrate that key network nodes of our model, which were identified through our sensitivity analysis as those rendering the network hypersensitive to enhanced ERK-(p) and Akt-(p) levels, have a striking correlation with species implicated in oncogenic transformations in human cancers as well as identified in drug resistance mechanisms for the inhibitors gefitinib and erlotinib in non-small-cell lung cancer therapy.

Parameter	WT	L834R	Reference
1	1	1	1
K_M^{ATP}	5.0 μM	10.9 μM	[23]
$K_I^{erlotinib}$	17.5 nM	6.25 nM	[23]
K_M^{Y1068}	265 μM	13.3 μM *	[51, 102]
K_M^{Y1173}	236 μM	944 μM *	[51, 102]
k_M^{Y1068}	0.29 s^{-1}	0.14 s^{-1} *	[23]
k_{cat}^{Y1173}	0.25 s^{-1}	0.010 *	[23]

Table 3.1: Parametric differences between WT and mutant EGFR systems. *Calculated from molecular simulations.

3.1 Methods

Models for the WT and L834R systems each comprised 74 species, 140 parameters, and 77 reactions and are described in detail in Chapter 2. The model parameters are derived from previously published models [85] and *de novo* molecular level calculations in our lab [102]. One important difference between the WT and L834R mutant network is that the WT RTK initiates phosphorylation of C-terminal tail substrate tyrosines only as a dimer, whereas the mutant—owing to constitutive activation—can initiate phosphorylation as a monomer as well as a dimer [184]. Our molecularly resolved signal transduction network incorporates differential signaling through Y1068 and Y1173 phosphorylation sites of the EGFR TK (see Section 3.1.3). To model competitive inhibition by an ATP analog, we employed the following reactions in our model.

Consistent with our computational studies (our molecular docking studies described in Chapter 2 revealed that the ATP binding affinity of the RTK in the inactive state is 100-fold lower than that in the active state), the above reactions are implemented as several analogous reactions, such that an inhibitor molecule may bind to the activated receptor kinase when the substrate tyrosines are unphosphorylated, or already phosphorylated at either Y1068 or Y1173. To model the kinetic behavior in the presence of the TKI erlotinib, we used appropriate values for the inhibition constant (K_I) of erlotinib, binding constant for ATP, and constants for describing the phosphorylation kinetics of Y1068 and Y1173 peptides for WT and mutant (L834R) receptors. These parameters are obtained from our molecular-level simulations and from experiments. The parameters are summarized in Table 3.1.

3.1.1 Systems-level Model Calibration using Genetic Algorithm

Model parameters were refined by calibrating simulation output with data from published cell-base assays and time-resolved mass spectrometry [185]. A genetic algorithm (GA) was used to identify parameter sets that produced the closest fits to experimental measurements, using a linear least

squares distance measure between computed and measured time course data points as the objective function. Initial estimates of parameters (default values) were taken from published studies, or computed by molecular docking and stochastic simulations as described previously [102]. In the GA, each parameter value was sampled from a log-normal distribution ($\mu = \text{default value}$, $\sigma = 1$ log unit) to allow the parameters (rate constants or initial conditions, except those for [EGFR], [EGF], and [TKI]) to range continuously over several orders of magnitude but with strong central tendency for the default value. The GA was implemented with a constant-sized population of 100 parameter sets, mutation rate $\mu = 0.15$, and crossover rate $\mu = 0.15$, for 1000 generations. Fitness scores for all individuals were evaluated at each generation by computing the objective function.

3.1.2 Model Sensitivity using Principal Component Analysis

In order to obtain a quantitative measure of the sensitivity and robustness of the signaling network to rate constants and initial concentrations, we employed a Monte Carlo protocol to perturb the default parameter sets. This was done by generating a random vector of elements ζ_i from a normal distribution with mean zero and variance (σ) 0.3, 1.0, and 3.0, then perturbing the default parameters P_i as $P_i \times \exp(\zeta_i)$. Network simulations were repeated 2000 times with new (perturbed) parameter sets and the sets yielding top 500 values for ERK and Akt activation were stored. A principal component analysis in parameter space (over the 500 stored sets) was carried out in order to determine the combinations of parameters (i.e., the constitution of the principal eigenvectors), whose perturbation are likely to render the signaling hyperactive to ERK and Akt phosphorylation, that is, phosphorylation of ERK/Akt in the perturbed system is greater than in the unperturbed system. These calculations were performed in two contexts: (1) by choosing the set of rate constants as the parameters P_i , and (2) by choosing the set of initial concentrations as the parameters P_i .

3.1.3 Differential Signaling through Y1068 and Y1173

Our molecular resolution to the systems model is enabled by two features in our research design: (1) the *de novo* estimation of key binding and rate constants associated with the receptor through molecular simulations, we are able to quantify the molecular contributions to signaling and hence differentiate signaling in WT and mutant cells; (2) our systems model introduces the concept of differential signaling, where phosphorylation at the Y1068 and Y1173 sites transduces signals through different pathways. This differential effect was initially incorporated based on experimental evidence [157] by assuming perfect (or 100/0) specificities for the interactions of the Y-phosphorylated docking sites: i.e., the phosphorylated Y1068 binds only to PI3K and Grb2 and

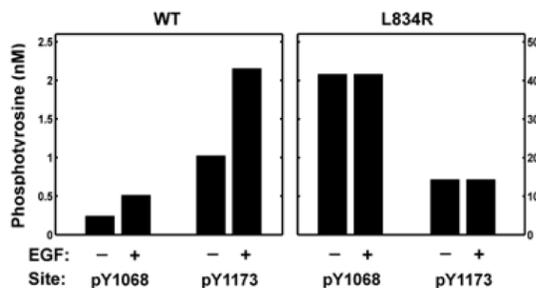


Figure 3.1: Differential signaling for WT and L834R. Total phosphotyrosine (pY1068 or pY1173) was calculated after 60 s, with (8 nM) and without ligand (EGF).

not PLC- β and Shc, and phosphorylated Y1173 binds only to PLC- β and Shc and not to PI3K and Grb2. The alternative approach is to assume no (50/50) specificity, i.e., the two sites 1068 and 1173, when phosphorylated, serve as docking sites to all four substrates in our model, namely PI3K, Grb2, PLC- β and Shc. This would reduce to a single-site model, where the docking sites are not distinguished. A realistic model for incorporating this differential effect lies between these two extremes. The degree to which the docking interactions of the phosphotyrosines are specific can be determined by molecular docking calculations similar to those described in the section below (these calculations are ongoing in our laboratory). However, in order to gauge the sensitivity of the downstream response (ERK vs. Akt activation) on the degree of specificities, we performed network simulations with varying degrees of specificities ranging from 100/0 to 50/50.

3.2 Results and Discussion

3.2.1 Differential phosphorylation

We first assessed the differential phosphorylation patterns of WT EGFR and L834R mutant systems, with normal receptor copy number (initial [EGFR] = 100 nM or 30,000 receptors per cell), with 8 nM (50 ng/ml) EGF or without EGF stimulation (Figure 3.1). As expected from the altered K_M and k_{cat} values, L834R has a stronger preference for Y1068 phosphorylation compared to the WT receptor, while our results predict the opposite trend for Y1173 phosphorylation.

3.2.2 Differential Signaling

To examine the effects of our branched signaling through pY1068 and pY1173 on the downstream response, we explored a series of cases in which differential signaling was fully implemented, i.e. 100/0 (see methods section), partially active (75/25), or completely absent (50/50). In the latter

case, docking peptides (e.g., PLC- β , PI3K, Gab1) could bind indiscriminately to either phosphorylated tyrosine sites. Simulations were performed for 15 min (900 s) with 8 nM EGF present. The resulting total pERK and pAkt response are reported in Figure 3.2.

For both systems, implementing either a partially or completely specific branching quantitatively yields the same downstream response, regardless of the degree of differential signaling. This observation validates our hypothesis that the differences in the downstream signaling between the WT and mutant receptors stem from changes in the efficiency of C-terminal tail phosphorylation (i.e., k_{cat}/K_M values for the substrate tyrosines) rather than the specificities of the docking proteins to phosphotyrosines. One notable feature is the 10-fold lower pERK levels in the mutant in the presence of a differential branched signaling, suggesting that specificity of the Y1173 peptide induces a sensitivity in the ERK activation, but the degree of specificity (when present) is unimportant.

3.2.3 Robustness of Signaling

In a recent review, the ErbB signaling network is described as a bow-tie-configured (or hour-glass shaped), evolvable network, displaying modularity, redundancy, and control circuitry [33]. This framework suggests that identifying the targets proteins for effective inhibition and the effects of the mutation landscape would require a systems level understanding of the signaling network. Based on the protocol outlined in the methods section, we perturbed the network using our Monte Carlo strategy and performed principal component analysis in the space of rate-constants as well as on initial concentrations. In both cases, the top three modes constituted 99% of the scatter in the parameter space. We then identified the components of the three principal eigenvectors (i.e., which particular combination parameters render the network hyper-sensitive according to our measure. There were several surprising features in our findings: (1) The components of the eigenvectors were non-zero only as singlets or pairs, suggesting that the perturbation of only one or at most a combination to two parameters were driving the network hyper-sensitive; (2) The identified components (Table 3.2), were insensitive to the degree of perturbation (i.e., the same components resulted irrespective of the extent of perturbation, σ of 0.3, 1.0, and 3.0; (3) The resulting components in Table 3.2 clearly reflect the bow-tie-configuration of the network; namely, the system is susceptible to perturbation only at the top (and bottom, these are our measures for hyper-sensitivity), with the middle (or core) layers robust to perturbations; (4) The top principal components identified almost exclusively comprised of ATP binding and C-terminal tyrosine phosphorylation, thereby justifying why the inhibition of receptor phosphorylation is likely to be an optimal target, and suggesting why mutations changing the (differential) phosphorylation kinetics can profoundly impact the downstream response.

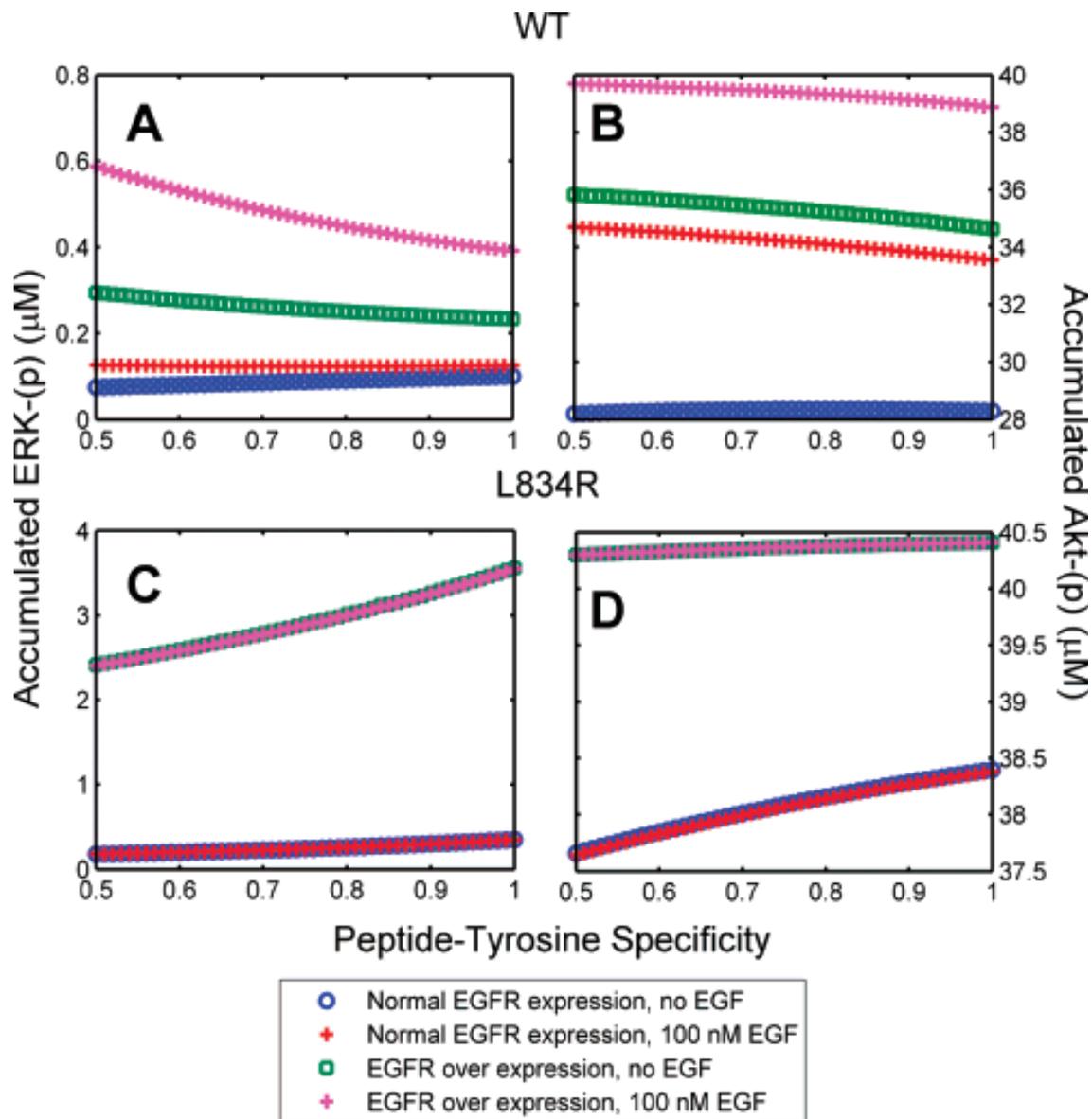


Figure 3.2: Effect of peptide specificity on activated ERK and Akt levels. The WT (A and B) and mutant (C and D) systems were simulated for 15 min under the conditions indicated. Accumulated (time-integrated) ERK-(p) (A and C) and Akt-(p) (B and D) levels were calculated as a function of peptide specificity; 0.5 denotes 50/50 or no specificity (peptide binds equally to either tyrosine); 1.0 denotes 100/0 or perfect specificity (peptide binds only to its cognate tyrosine as described in the text).

Rate Constants	Initial Concentrations
k_f : Y1068	[Raf·Ras·GTP]
k_f : Y1173	phosphatase for pERK
K_M : ATP·RTK	[pMEK]
K_M : GAB·pEGFR	[PI3K inactive]
.	[MEK·Raf active]
.	[EGFR·Shc·Grb2·SOS·RasGTP]

Table 3.2: Parameter variations constituting the top three principal components for network hyper-sensitivity.

3.2.4 EC_{50} for Inhibition of EGFR in the cell

Next, we examined the sensitivity of WT and mutant systems to inhibition in the cellular context by calculating receptor phosphorylation over a range of erlotinib concentrations. Simulations were performed for both “normal” ($[EGFR] = 100$ nM) and “over-expressed” ($[EGFR] = 1000$ nM) systems. All simulations were performed with (50 ng/ml or 8 nM) or without ligand (EGF) stimulation. The EC_{50} (inhibitor concentration at which 50% of the activity is suppressed in the cellular context; this is different from IC_{50} because the nonlinear and temporal effects due to signal-transduction is accounted for) for L834R was 50-fold lower (Figure 3.3a,b) than that of the WT (100 nM vs. 5000 nM) in the presence as well as absence of ligand. There were also no prominent differences between the inhibition at 1068 and 1173 sites. In over-expressed systems, there was no significant difference in EC_{50} among the four groups (Figure 3.3c,d). All had an EC_{50} near 1000 nM.

3.2.5 Inhibition of Downstream Activation (EC_{50})

We then examined the sensitivity of downstream signaling molecules ERK and Akt to inhibition for a range of erlotinib concentrations (see Figure 4: a-d). In normal expression systems (Figure 3.4a,b), there was a nearly 7-fold increase in the efficiency of pERK inhibition for L834R ($EC_{50} = 100$ nM) compared to WT ($EC_{50} = 700$ nM) with and without ligand present. With respect to Akt activation, there was a 4-fold decrease in EC_{50} for L834R compared to WT (300 nM vs. 1200 nM) with and without ligand present. For the case of over-expressed receptors (Figure 3.4c,d), inhibition of pERK is 4-fold more efficient for L834R compared to WT, while pAkt inhibition is almost 10-fold more efficient for the L834R mutant, although significantly more erlotinib is required to achieve inhibition ($EC_{50}=40$ M). This result is consistent with the dramatically elevated Akt levels expected in systems that bear both the L834R mutation and a higher receptor count.

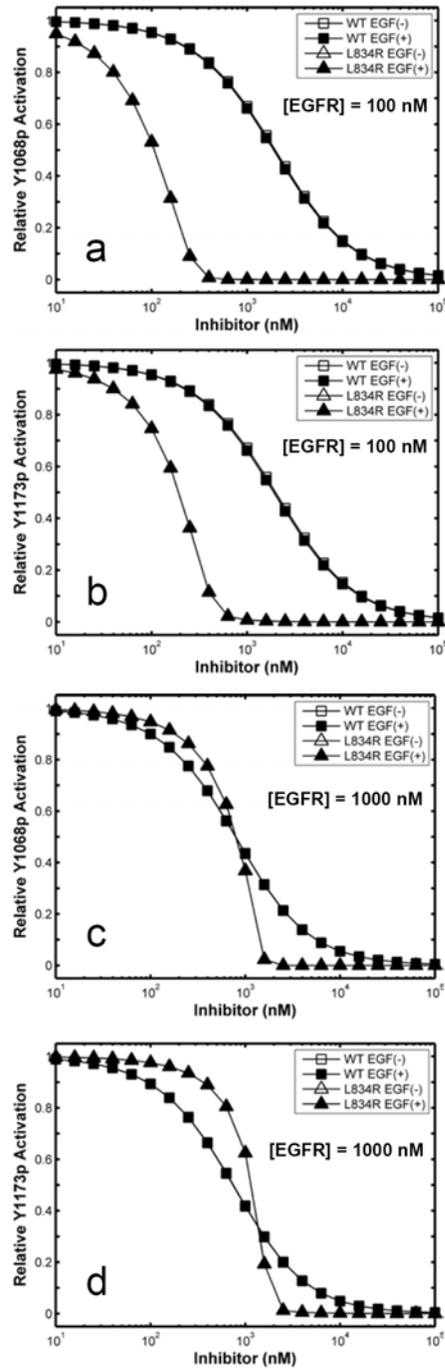


Figure 3.3: Relative inhibition of receptor phosphorylation: Tyr phosphorylation levels relative to that without inhibitor are plotted.

3.3 Conclusion

We have presented a framework to interpret the effect of tyrosine kinase inhibitors on the system-wide cellular response. Through multiscale modeling, we have also shown how this framework can be utilized to predict the efficacy of inhibition in cell-lines, especially those harboring mutations in the receptor. We find that the mutant cell line is more susceptible to inhibition by TKIs through the curbing of receptor phosphorylation and downstream (ERK and Akt) activation. Considering that the absolute pAkt levels are 5-fold higher than those for pERK in the WT and 100-fold higher in the mutant (Figure 3.1), the remarkable effect of the drug in non-small-cell lung cancer cell lines carrying the mutation can be attributed to the gain in efficacy with respect to Akt inhibition. The multiscale computational framework described here is ideal for assessing mutation landscape on signal transduction. We believe that our model driven approach will in the long-term significantly impact the optimization of future small molecule therapeutic inhibition strategies.

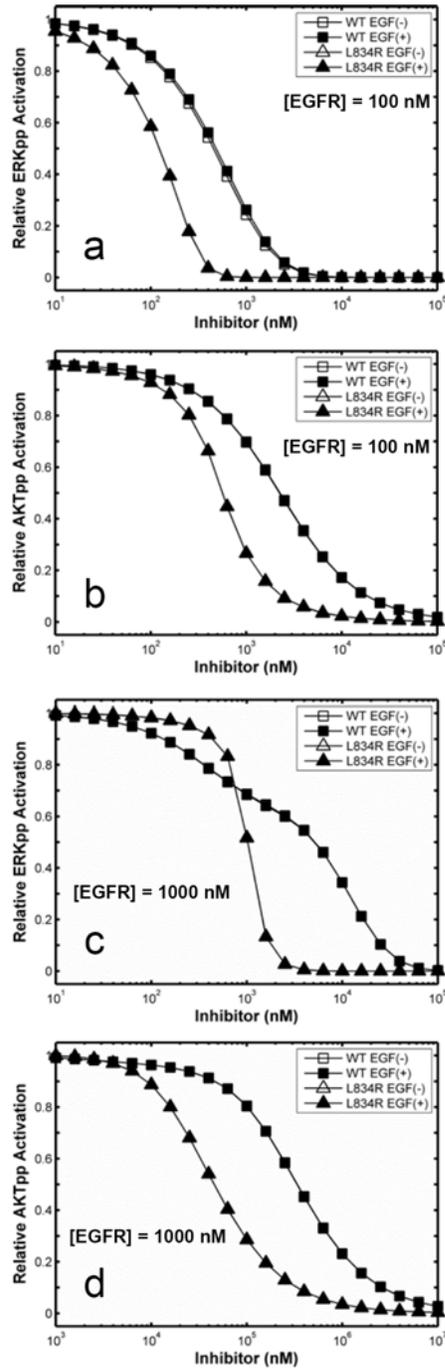


Figure 3.4: Inhibition of ERK and Akt activation in systems with normal receptor expression. Responses plotted relative to that without inhibition under the same conditions.

Chapter 4

Calcium and Phosphoinositide Signaling in Human Platelets

4.1 Introduction

Platelets respond to endothelial injury or activating agonists by engaging a host of intracellular signaling events, including receptor activation, G protein signaling, second messenger generation, Ca^{2+} release, granule secretion, and cytoskeletal rearrangement [17, 150]. While the molecular details underlying individual activation events are continually refined through focused studies, it has been difficult to develop a unified and integrated view of platelet metabolism since it involves a large number of simultaneously interacting molecular components. To this end, computational models help provide an integrated view of the many interacting components involved in cellular signaling [4]. Successful applications of the systems biology approach have led to an improved understanding of cell surface receptors [15, 86], quantitative prediction of protease cascades in blood coagulation [66, 91, 105], and the discovery of novel molecular interactions [136]. When based on reliable data sets, accurate models have the potential to not only explain previously observed behaviors, but to make experimentally verifiable predictions about how cells process biological signals [74].

In many ways, the platelet provides an ideal test system for human systems biology. Platelets from normal donors or patients are readily available for *in vitro* diagnostic research and clinical monitoring. Isolated platelets, platelet rich plasma, or whole blood are amenable to high throughput experiments to study signal transduction and associated coagulation protease cascades. Additionally, numerous genetic mutations in humans [118] as well as knockout and transgenic mice [79, 137]

are known that predispose to bleeding phenotypes. From a modeling perspective, the platelet is particularly well-suited for study since it lacks a nucleus, allowing one to avoid the challenges of describing whole-genome transcriptional regulation. Finally, the well-appreciated importance of platelets in mediating thrombosis and hemostasis, as well as their contribution to systemic disorders such as inflammation and cancer [36], place a high practical value on an accurate platelet model. Such a model would be useful both as a basic research tool for predicting normal activation behavior and as a strategy for the rational design of patient-specific pharmaceutical therapies [92]. In this study, we describe a computational model of the human platelet that accurately predicts both resting and activated system behaviors. The model is based on 24 peer-reviewed studies spanning three decades of platelet research. Because of the inherent complexity in a model of this size, we first constructed a set of four distinct signaling modules (see Chapter 5) to describe (i) Ca^{2+} release and uptake, (ii) phosphoinositide (PI) metabolism, (iii) P2Y1 G-protein signaling, and (iv) protein kinase C (PKC) regulation of phospholipase C- β (PLC- β). Using fixed reaction equations and fixed kinetic parameters, each module was tuned to match a relevant set of experimental data by selecting allowable values for the resting concentrations of the module species. For the final analysis, the four modules were merged into a single kinetic model and fit to Ca^{2+} release data from ADP-stimulated platelets. Dynamic traces of intracellular Ca^{2+} represent multiple steps in a complex signaling pathway from extracellular ligand to intracellular activation response, incorporating dynamic contributions from each module in the system. The fully integrated model was capable of reproducing population and single-cell Ca^{2+} release data as well as measurements of PI turnover and G protein activation.

4.2 Materials and Methods

4.2.1 ADP-stimulated Ca^{2+} release assay

Human blood was collected from healthy donors via venipuncture and anticoagulated with sodium citrate (9 parts blood to 1 part sodium citrate). Platelet-rich plasma was incubated for 45 min with Fluo-4 NW (Invitrogen) per the manufacturers instructions. PRP was diluted to a final concentration of 12% on a 384-well plate. A separate plate containing varying concentrations of ADP was prepared on a PerkinElmer Janus. Intracellular Ca^{2+} concentration was quantified by measuring $F(t)/F_0$ using a Molecular Devices FlexStation. Results are the average time-series measurements of 6 replicates. Phlebotomy was conducted in accordance with the Declaration of Helsinki and under University of Pennsylvania Institutional Review Board approval.

4.2.2 Platelet Image Analysis

High-resolution electron micrographs of platelets stained with glucose-6-phosphatase obtained from Ebbeling et al. [46] were used to quantify the relative area of the dense tubular system (DTS) with respect to cytoplasmic area. Stained areas were delineated by using a threshold value for pixel intensity. To determine the sensitivity of the technique to the chosen threshold value, we calculated the DTS area using a maximally inclusive threshold (all stained regions and some of periphery included in delineation) and a minimally inclusive threshold value (delineation entirely within stained portion of image). The range of results obtained by both threshold values did not differ by more than a factor of two.

4.2.3 Model Construction and Simulation

Ordinary differential equations (ODEs) were used to describe reaction rates corresponding to published reaction mechanisms and kinetic parameters (see Table 1 in ref. [129]). To efficiently estimate the dynamic range of initial conditions (ICs), the abundance of each estimated species or compartment size was estimated over a base-10 logarithmic scale (as in Figure 4.3). For reactions occurring at compartmental interfaces (e.g., cytosol and plasma membrane), reaction rates were adjusted by scaling the concentrations of reactants to the bulk compartment as described by Kholodenko et al. [86]. Within each of the five compartments, species were assumed to be well-mixed. While we do not discount the importance of spatial gradients in Ca^{2+} signaling [177], previous studies have shown that changes in $[\text{Ca}^{2+}]_i$ within platelets are effectively instantaneous [64]. Simulations were performed using the SBToolbox for MATLAB [146] Numerical integration was performed with an absolute tolerance of 10^{-45} and a relative tolerance of 10^{-9} .

4.2.4 Module Calibration and Analysis

The model topology was defined by a set of reaction equations, rate laws, and kinetic constants (see ref. [129] for a complete list of model reactions and rate constants). The concentrations of all enzymes and metabolites in the resting platelet at homeostasis comprise the initial condition (IC) space of the model. The topology of the Ca^{2+} signaling network and almost all other kinetic parameters are known from literature [129] except for those describing P2Y_1 activation, which were calibrated from measurements of P2Y_1 -reconstituted proteoliposomes (see 4.3). A subset of IC space for each module was established by use of homeostasis constraints and experimental data from platelets. The homeostasis constraint requires that a resting platelet remains resting such that the initial condition is also a steady-state solution of a given set of ODEs (see Chapter 5). We

performed a dense sampling of unknown ICs over fixed reaction topologies to identify combinations of IC values that were consistent with known steady-state concentrations.

Calculation of Ca^{2+} spiking frequency distribution. As described by Heemskerk et al. [64], strong release events were identified by taking the derivative of the Ca^{2+} flux traces and setting a noise threshold above which peaks were considered to be significant (Figure 4.8). Video-imaged platelets were found to be quite variable both in peak interval times (4 to 40 s) and peak amplitudes (20 to 300 nM). We also observed significant variability in the platelet model, with peak amplitudes ranging from 100 to 300 nM (Figure 4.7C). We were unable to estimate a single dominant frequency using Fourier transform analysis of the simulated Ca^{2+} traces because of the substantial noise and irregularity of spiking. Half-maximal agonist dose (1 μM) was used to generate a frequency distribution for desensitized platelets treated with antagonists prostaglandin I_2 , apyrase, and aspirin prior to 20 μM ADP.

4.3 Results

4.3.1 Model Overview

The full model (Figure 4.1, ref. [129]) comprises four interlinked kinetic modules (Figure 4.2A-D). The Ca^{2+} module (Figure 4.2A) spans five compartments and functions to maintain a low intracellular Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) by pumping ions across the PM into the extracellular space, or across the membrane of the DTS into platelet stores using a sarcoplasmic/endoplasmic reticulum Ca^{2+} ATPase (SERCA). Inositol trisphosphate receptor (IP_3R) channels release Ca^{2+} from the DTS and are regulated by inositol 1,4,5-trisphosphate (IP_3) and Ca^{2+}_i . In the PI module (Figure 4.2B), membrane-derived signaling intermediates such as IP_3 are continually recycled between the plasma membrane (PM) and cytosol by a series of phosphorylation, dephosphorylation, synthesis, and hydrolysis reactions [129]. Among its hydrolysis products, $\text{PLC-}\beta$ generates PM-bound diacylglycerol (DAG), which, along with Ca^{2+}_i , modulates the activity of PKC. PKC dampens GPCR-mediated signaling by phosphorylating $\text{PLC-}\beta$, reducing its hydrolytic activity. In a module for receptor activation (Figure 4.2D), free ADP in the extracellular compartment binds to the G protein-coupled receptor (GPCR) P2Y_1 , causing a subsequent rise in activated G protein ($\text{G}_q\text{-GTP}$). $\text{G}_q\text{-GTP}$ has a basal rate of hydrolysis that is accelerated when it is bound to $\text{PLC-}\beta$. The transient complex $\text{PLC-}\beta\text{G}_q\text{-GTP}$ ($\text{PLC-}\beta^*$) is the hydrolytically active form of the enzyme and regulates transmembrane signaling in the PI module.

Clearly, the model does not include all of the reactions that govern platelet homeostasis and activation. Such a list is both unavailable and intractable to deploy until a core set of modules can be

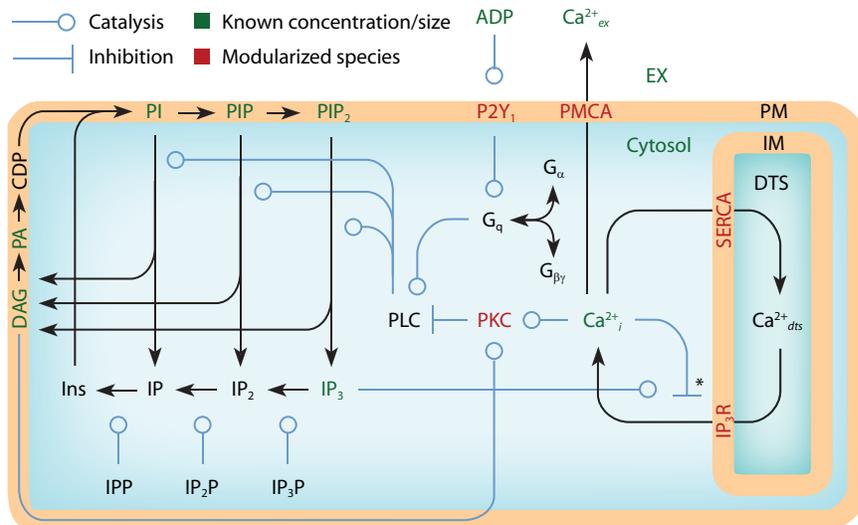


Figure 4.1: Schematic of platelet model compartmentalization and reactions. The full model comprises 5 compartments, 70 species, 77 reactions, and 132 kinetic parameters [129]. 3D compartments: extracellular (EX), cytosolic, and DTS. 2D compartments: PM and DTS inner membrane (IM). Abbreviations: IPP, Inositol phosphate phosphatase; IP_2P , Inositol bisphosphate phosphatase; IP_3P , inositol trisphosphate phosphatase; DAG, diacylglycerol; PA, phosphatidic acid; CDP, CDP-DAG.

validated, which is the focus of this study. We have focused on the subset of reactions that govern ADP-mediated phosphoinositide signaling and mobilization of intracellular Ca^{2+} [129]. The $P2Y_1$ receptor is coupled primarily to the G_q family of G proteins,20,21 which activate the beta-isoform of phospholipase C [17,82]. Previous work has shown that this pathway accounts for more than 90% of the ADP-mediated Ca^{2+} signaling in platelets [138]. To eliminate Ca^{2+} influx and the need to model store-operated Ca^{2+} entry, we used experimental data from studies in which extracellular Ca^{2+} was removed by EDTA or another chelating agent. EDTA was also used in our own experiments. ADP releases Ca^{2+} only from the DTS [104] and thus modeling the acidic store is not required at this point.

4.3.2 Platelet Ca^{2+} Balance

The Ca^{2+} module addresses a fundamental question of what resting level of IP_3 is needed to balance the open probability of IP_3R channels with SERCA in a platelet of a given DTS volume and $[Ca^{2+}]_{dts}$. Our strategy for modeling platelet Ca^{2+} signaling can be summarized in two steps: First, we considered the kinetic properties of the IP_3R and SERCA [45,154,186], the resting Ca^{2+}_i concentration [150], and the volume of the platelet [174] to be accurate as reported. These values were held fixed in the module. Second, we generated more than 10^9 combinations of the unknown

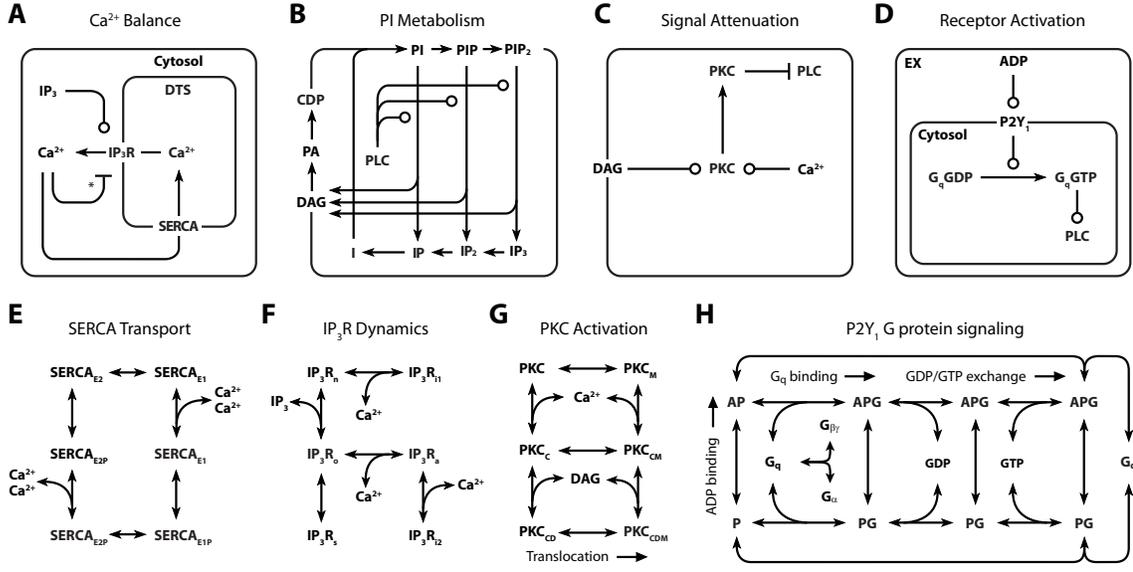


Figure 4.2: Detailed reaction schemes for platelet signaling modules. **(A)** Ca^{2+} Module: cytosolic and DTS compartments are separated by the DTS membrane, which contains the IP_3R and SERCA. **(B)** PI Module: PM-bound PIs are cleaved by PLC- β to form diffusible inositol phosphates and DAG, which are substrates for resynthesis of PIs. **(C)** PKC Module: Ca^{2+}_i and DAG activate PKC, which migrates to the PM where it phosphorylates PLC- β . **(D)** P2Y₁ Module: extracellular ADP binds to and activates P2Y₁. Active P2Y₁ accelerates guanine nucleotide exchange on bound G_q . G_q -GTP binds and activates PLC- β , which increases the GTPase activity of G_q -GTP. **(E)** SERCA catalytic cycle [45]: Subscripts: E1, facing cytosol; E2, facing DTS; P, phosphorylated. **(F)** IP_3R dynamics [154]: Subscripts: n, native; i1, inhibited; o, open; a, active; s, shut, i2, inhibited. **(G)** PKC activation: Active kinase is bound to Ca^{2+}_i and DAG and located at the PM. Subscripts: M, located at the PM; C, Ca^{2+} -bound; D, DAG-bound. **(H)** P2Y₁ activation module: Rate equations describing the interactions among ADP, P2Y₁, and G_q were modeled according to the ternary complex model described in Kinzer-Ursem et al [87]. For clarity, DTS membrane and cytosolic compartments are not delineated. Abbreviations: A, ADP; P, P2Y₁; G, G_q . * Ca^{2+} both activates and inhibits IP_3R [154].

values in the module (number of IP₃R/platelet, SERCA pumps/platelet, [Ca²⁺]_{dts}, volume of the DTS) to find a set of “configurations” that reached an equilibrium state with the known resting [Ca²⁺]_i (see 4.2). Each configuration may be thought of as a unique platelet model with a characteristic compartmental structure and molecular makeup (i.e., different numbers of SERCAs, IP₃Rs, etc). Note that more than one module configuration can produce the same resting [Ca²⁺]_i. Existing kinetic, electrochemical, and physiological data was incorporated into the module as follows: Resting platelets maintain a [Ca²⁺]_i between 40-100 nM [64, 150], while storing concentrated Ca²⁺ ([Ca²⁺]_{dts}) in the DTS [150]. Pumping of Ca²⁺ by SERCA isoforms (Figure 4.2E) was modeled according to a kinetic study of the type 3 SERCA isoform [45], which is abundant in human platelets [93]. For the IP₃R [154], each subunit of the type 2 receptor exists in one of six states: native, open, shut, active, or two inactive states (Figure 4.2F), where state transitions depend on [IP₃] and [Ca²⁺]_i. The channel open probability (P_o) is determined by the number of IP₃R tetramers with all subunits in either open or active conformations [129]. We used the Nernst equation [2] to relate P_o to the release of Ca²⁺ from the DTS:

$$\frac{d[\text{Ca}^{2+}]_i}{dt} = NP_o\gamma e \frac{RT}{zF} \ln \left(\frac{[\text{Ca}^{2+}]_{dts}}{[\text{Ca}^{2+}]_i} \right) \quad (4.1)$$

Here, N is the total number of channels per platelet, γ is the single-channel conductance (10 pS) of the platelet IP₃R [186], and e is the number of elementary charges (z) per second per Ampere (6.24×10^{18}). The cytosolic volume of the platelet is ≈ 6 fL [174].

The unknown quantities in the module were the number of IP₃R and SERCA per platelet, [Ca²⁺]_{dts}, and the volume of the DTS. The range of values sampled for each quantity is shown in Figure 4.3A. A dense sampling ($n = 10^9$) of these quantities generated 100,000 unique configurations of the Ca²⁺ module (with [Ca²⁺]_i = 100 ± 10 nM) that were further divided into three groups (low-, mild-, and high-response) based on the estimated [Ca²⁺]_{dts}, an indicator of the configurations response to an increase in [IP₃] (Figure 4.3B). To examine any molecular or structural differences among low-, mild-, and high-response groups, we compared the levels of SERCA, IP₃R, IP₃, and relative DTS volume that were estimated for each group. These values are presented as probability distributions (Figure 4.3C), where regions of high density reflect frequently occurring values for the estimated concentrations and compartment sizes. For example, low response configurations tended to have ≈ 1000 IP₃R channels per platelet as indicated by red regions (Figure 4.3B). We note that, of 10⁹ sampled configurations, only 0.005% satisfied the dual constraints of steady-state Ca²⁺ homeostasis and IP₃ responsiveness (mild- or high-response configurations). This observation indicated that the kinetic properties of Ca²⁺-regulating enzymes such as IP₃R and SERCA alone place strong constraints on the physical size and molecular makeup of the platelet.

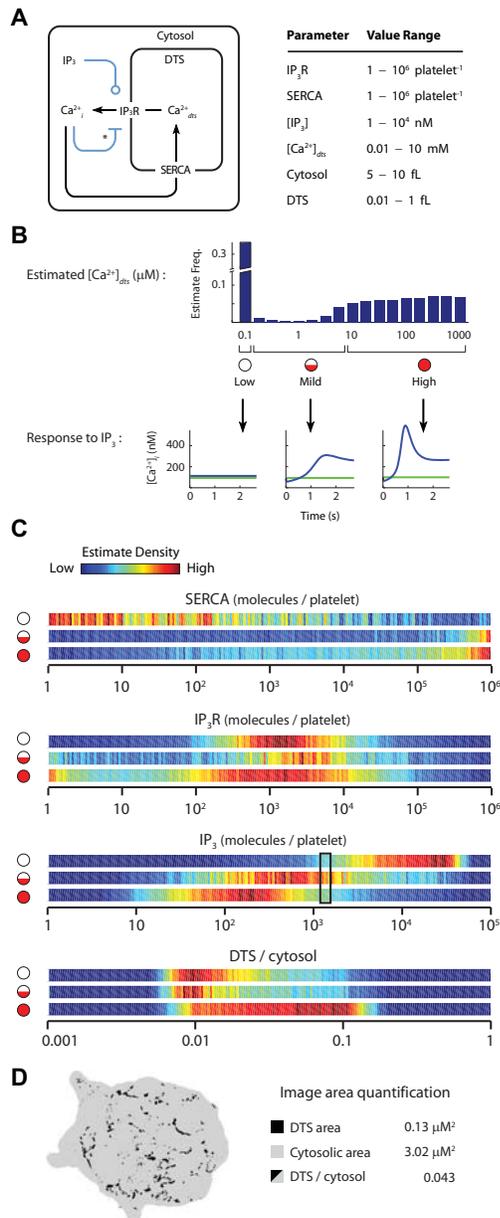


Figure 4.3: Estimation and analysis of platelet Ca²⁺ regulation. Unknown concentrations and compartment sizes in the platelet Ca²⁺ module were estimated by randomly sampling the unknown quantities in panel **A** and selecting only those molecular “configurations” that produced a resting [Ca²⁺]_i of 100 ± 10 nM. **(B)** This reduced set of configurations was divided into 3 groups based on the estimated [Ca²⁺]_{dt_s} (top), which corresponded to the [Ca²⁺]_i release behavior after a 10-fold increase in [IP₃] (bottom). **(C)** Probability distributions for estimates of SERCA, IP₃R, IP₃, and the relative DTS volume in low-, mild-, and high-response configurations ($n = 100,000$). The measured IP₃ level from resting platelets [161] is boxed. **(D)** Image obtained from glucose-6-phosphatase stain of the DTS from a human platelet [46] was used to quantify the cytosolic (gray) and DTS (black) area from a 2D section of the platelet.

Configurations with a negligible gradient ($[Ca^{2+}]_{dts} \approx 100nM$) lacked IP_3 -responsiveness and had relatively low SERCA levels (67% had ≤ 5000 copies) with a median estimate of 1600 IP_3R channels/platelet. By contrast, mild-response configurations ($[Ca^{2+}]_{dts} \leq 10 \mu M$) responded to elevated $[IP_3]$ with a transient rise in $[Ca^{2+}]_i$ and were characterized by high SERCA levels ($\approx 10^6$ /cell) and higher IP_3R abundance (median = 2000 channels/cell). High-response configurations ($[Ca^{2+}]_{dts} \geq 10 \mu M$) clustered near 10^6 SERCA pumps but showed a broad distribution of IP_3R abundance. The calculated IP_3R /SERCA ratio for low-response configurations had a loosely defined distribution, while the mild- and high-response configurations favored 10^3 - 10^5 more pumps than channels (1:6900 for mild-response and 1:5200 for high-response). Thus, the model predicted a very low IP_3R /SERCA ratio for functioning platelets, a hypothesis supported by noting that a single SERCA3b pump operating at a membrane with a 1000-fold Ca^{2+} gradient transports $\approx 0.4 Ca^{2+}$ ions/s [45], while a single type 2 IP_3R channel in the same membrane at 37° conducts $\approx 3,000 Ca^{2+}$ ions/s, assuming a steady-state P_o of 0.0005 [154].

The low-response configurations harbored high resting concentrations of IP_3 , inconsistent with a resting platelet [16]. The median IP_3 count for mild-response configurations was 750 molecules/cell, similar to the measured value (≈ 1200 molecules/platelet or $\approx 200 nM$) [161] in resting platelets. More than 80% of high-response configurations harbored ≤ 1000 IP_3 molecules/cell in resting platelets. The relative volume of the DTS compartment to cytosol, as predicted from the estimation procedure, was quite constrained at 0.5-5% (median = 2%) of the non-granular intracellular volume. To compare this result to a direct physical observation, we quantified the relative area of the platelet DTS as revealed by glucose-6-phosphatase staining [46]. This calculation gave a DTS/cytosol fraction of 4.3%. Assuming the stained region marks the true DTS border, this estimate is accurate within a factor of two (see Section 4.2). Thus, estimates obtained from independent kinetic calculations or image analysis of stained platelets gave similar values for the DTS compartment volume fraction.

4.3.3 PI Metabolism

PIs are continually interconverted in platelets, even under resting conditions [16, 150, 176]. To capture this behavior, we employed a set of synthesis and degradation reactions that continually recycles PIs in the absence of a stimulating dose of agonist (Figure 4.2B, [129]). Failure to account for PI resynthesis would have prevented prediction of steady-state concentrations. Using the same strategy for modeling platelet Ca^{2+} regulation (see Section 4.3.2), we generated 10,000 unique PI module configurations that were consistent with the measured resting concentrations of PtdIns,

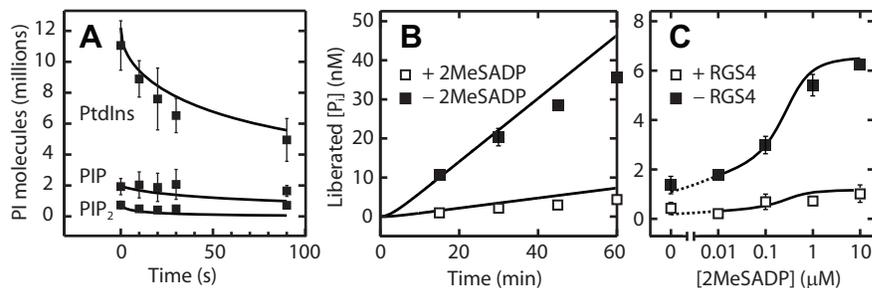


Figure 4.4: Comparison of PI and P2Y₁ modules to experimental data. (A) At $t = 0$, platelets were treated with 27 nM thrombin and the levels of PtdIns, PIP, and PIP₂ were measured (data obtained from Wilson et al. [176]). To simulate this turnover in the PI module, [PLC- β^*] was raised to 1 μ M and PI levels were calculated over time. Average results from 100 simulations are plotted as solid lines. (B,C) Kinetic parameters in the P2Y₁ module were fit to GTP hydrolysis measurements from proteoliposome assay containing 300 nM P2Y₁, 1 μ M G α -GDP, 3 M G $\beta\gamma$, 2 μ M GTP, RGS4, and 2MeSADP (data obtained from Waldo et al. [169]). Best-fit simulation results are plotted as solid lines. (B) Time-course GTPase activity is shown with (*filled squares*) or without (*open squares*) 100 nM RGS4. (C) P2Y₁ dose-response to 2MeSADP in the absence (*open squares*) or presence (*filled squares*) of 100 nM RGS4. In the final platelet model, K_d and k_{cat} values for ADP-binding and PLC- β -mediated hydrolysis, respectively, were substituted for the 2MeSADP- and RGS4-mediated activities used in the calibration procedure shown here.

PIP, PIP₂, DAG, and PA [126,168,176]. Thus, each configuration represents a potential molecular arrangement of PIs in the platelet that is consistent with published kinetic rates and resting measurements for key membrane phospholipids.

To examine the transient behavior of the PI module under activating conditions, we set the concentration of PLC- β^* in the module to 1 μ M (\approx 100-fold above basal levels). Figure 4.4A compares the measured and simulated changes in the platelet PIs after treatment with thrombin or elevated PLC- β^* , respectively (both thrombin and ADP signal through G_q-mediated stimulation of PLC- β 1). The simulated time course for PtdIns degradation was consistent with experiment throughout the 90 s interval following activation. Simulated levels of PIP and PIP₂ were consistent with experiment [176] through the first 25 s (Figure 4.4A). After this time, PIP and PIP₂ continued to be degraded in the simulation but were sustained, or slightly elevated, in thrombin-stimulated platelets. A recent study in murine megakaryocytes [170] has confirmed that resynthesis of PIP₂ is necessary to replenish basal levels following agonist stimulation, a feature that is included in the model. However, because the rate of hydrolysis exceeds the rate of PI synthesis, we observed a monotonic decrease in PIs when the level of PLC- β^* was held fixed. To address this discrepancy between model and experiment, we introduced a negative-feedback module (see Signal Attenuation, Figure 4.2C) in which the activity of PLC- β^* is attenuated by activated PKC (PKC*) through phosphorylation of the phospholipase.

4.3.4 Signal Attenuation

As noted above, the degradation of PIs in platelets following PLC- β activation does not decrease monotonically [150, 176]. Furthermore, the sharp rise in $[\text{IP}_3]$ and $[\text{Ca}^{2+}_i]$ following GPCR stimulation returns to near-basal levels within 30-60 s [43]. This negative-feedback mechanism was modeled by allowing direct inactivation of PLC- β by PKC through a phosphorylation reaction, which has been observed in several cell types [55, 135]. In the PKC module (Figure 4.2C), activation of the kinase requires association with Ca^{2+}_i and DAG and translocation to the PM [13]. Activated PKC (PKC*) phosphorylates PLC- β , rendering it unable to bind $\text{G}_q\text{-GTP}$. Clearly, this represents only one of several potential mechanisms for regulation of G_q -dependent PI hydrolysis in platelets. Other mechanisms include receptor internalization [11], deactivation of G proteins [17], and accelerated dephosphorylation of IP_3 [37]. While these mechanisms will provide useful contributions to future iterations of the platelet model, it is interesting that the use of PLC- β regulation was sufficient to reproduce the observed attenuation in PI signaling following agonist stimulation (see Figure 4.5C,D). Specifically, the gradual accumulation of Ca^{2+} and DAG during the first 10-15 seconds provides the appropriate time-delay for shutting off PIP_2 hydrolysis.

4.3.5 Receptor Activation

P2Y_1 is a GPCR expressed on the surface of human platelets (≈ 150 copies/platelet [12]) that is essential for ADP-induced platelet shape change and aggregation [82]. Although other ADP receptors are involved in platelet Ca^{2+} signaling (e.g., P2X_1 and P2Y_{12}), P2Y_1 contributes $\approx 90\%$ of the Ca^{2+} signal [138]. In our representation of P2Y_1 activation (Figure 4.2D,H), binding of extracellular ADP to P2Y_1 leads to activation of G_q through GDP/GTP exchange reactions. $\text{G}_q\text{-GTP}$ is a substrate for GTPase activating proteins (GAPs) such as PLC- β and RGS4, which can accelerate $\text{G}_q\text{-GTP}$ hydrolysis over 1000-fold [114]. Bornheimer et al. showed that the combination of G proteins, active GPCRs, and GAPs can form kinetic “ternary” modules with distinct signaling patterns [15]. Thus, we used results from a published *in vitro* assay of P2Y_1 activity to construct a ternary model of the P2Y_1 receptor. In that assay, Waldo et al. combined purified human P2Y_1 , ADP, Ga and $\text{G}\beta\gamma$ subunits, PLC- β or RGS4, and radiolabeled GTP in large phospholipid vesicles and measured the agonist binding and G_q -stimulating activities of the receptor [169]. GTPase activity was measured with and without agonist (Figure 4.4B). ADP dose response was measured with and without GAP protein (Figure 4.4C). Using identical simulated conditions, we obtained the kinetic parameters in the P2Y_1 module from these time-course measurements. The kinetic

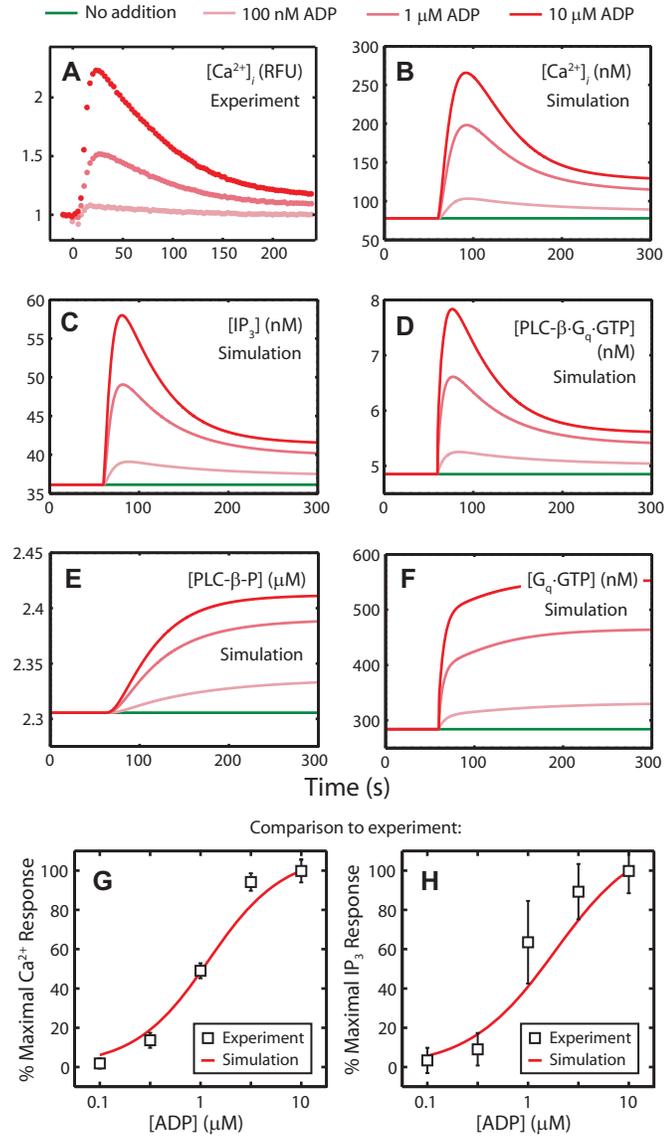


Figure 4.5: ADP dose response for the full platelet model. **(A)** Relative fluorescence of Fluo-4-loaded human platelets was measured after treatment with 100 nM, 1 μ M, or 10 μ M ADP. **(B-F)** After 1 min of simulated rest, the platelet model was activated by setting the extracellular [ADP] to 100 nM, 1 μ M, or 10 μ M. Deterministic computation of **(B)** $[Ca^{2+}]_i$, **(C)** $[IP_3]$, **(D)** $[PLC-\beta^*]$, **(E)** deactivated $[PLC-\beta]$, and **(F)** activated $[G_q-GTP]$. **(G, H)** Gel-filtered human platelets were activated with ADP and maximal changes in **(G)** $[Ca^{2+}]_i$ and **(H)** $[IP_3]$ from basal levels were measured [43]. For panels **G** and **H**, simulated responses are plotted in red.

rate constants, given in Table 1 of ref. [129], accurately reproduced both sets of measurements (Figure 4.4B,C).

4.3.6 Signaling in Platelet Populations

The four modules described above were merged into a single kinetic model using a bootstrap method (Figure 4.6). With a fixed reaction network (Figures 4.1 and 4.1) and fixed kinetic rate constants [129], initial conditions were fit to dose-response $[Ca^{2+}]_i$ time-course data from ADP-stimulated platelets (Figure 4.5A,B). In the presence of basal ADP levels similar to human venous plasma [8], the model maintained a resting $[Ca^{2+}]_i$ of 75 nM (Figure 4.5B). At higher ADP concentrations, $[Ca^{2+}]_i$ reached peak levels ≈ 20 s after addition of agonist with a decay constant of ≈ 1 min. We found the synchronous $[Ca^{2+}]_i$ peak-response at all agonist doses to be a complex and unique feature in the data, particularly well-suited for testing model validity. Achieving this characteristic shape required a transient (non-monotonic) rise in $[IP_3]$. In ADP- and thrombin-stimulated platelets [43, 132], $[IP_3]$ increases rapidly after agonist addition, peaks around 15 s, and then decreases to near-basal levels. The model predicted this behavior accurately (Figure 4.5C). We observed that PLC- β^* , which comprised $\approx 0.2\%$ of the total PLC- β pool in the model, strongly controlled this rise and fall of $[IP_3]$ (Figure 4.5C,D). Phosphorylation of inactive PLC- β by PKC* (Figure 4.5E) was sufficient to dampen G_q -mediated PI hydrolysis. The simplicity of this negative-feedback model suggests a mechanism whereby the phosphorylation of inactive PLC- β , present at much higher abundance than PLC- β^* , fine-tunes the hydrolytic activity of the enzyme by sequestering it from G_q -GTP. Addition of ADP caused a sharp burst in $[G_q \cdot GTP]$ that saturated within 5-10 s for all doses (Figure 4.5F). This trend was consistent with the observed time-scale for G protein activation [87]. Maximal $[Ca^{2+}]_i$ and $[IP_3]$ responses to increasing ADP was predicted quantitatively (Figure 4.5G,H).

4.3.7 Signaling in Individual Platelets

Deterministic simulations, such as those in Figure 4.5B-F, resemble data obtained from suspensions of platelets (Figure 4.5A), in which a million or more cells may be assayed simultaneously. These conditions often produce average or “smooth” responses that are typical of cell populations. However, when platelets are monitored individually, the addition of agonist can produce a series of sharp, asynchronous Ca^{2+} spikes [63, 64]. To test the single-platelet response in the model, we used the Stochastic Simulation Algorithm [59] to simulate Ca^{2+} release in a single platelet during rest or activation by ADP. Using the estimated sizes of the platelet compartments, species concentrations were converted to integer values for use in the stochastic simulation. Note that any species present

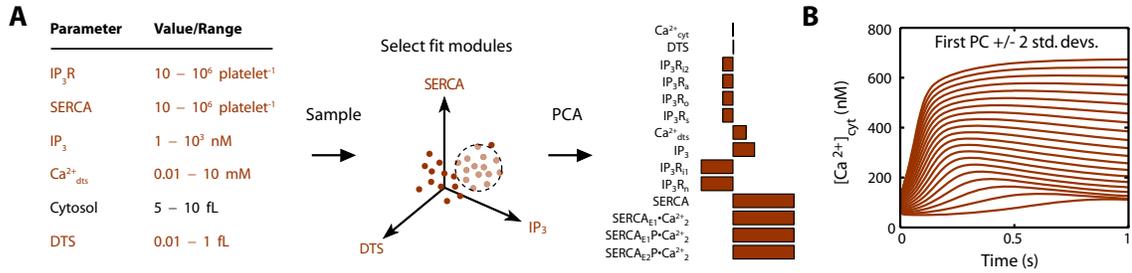


Figure 4.6: Modular calibration and feature extraction. **(A)** Example construction of a functional module, showing calibration of platelet Ca^{2+} balance. A fixed reaction network containing one or more unknown ICs (e.g., the concentration of a protein or the size of a compartment) was simulated to equilibrium using a large number of random value-sets. The output of each simulation was compared to experimentally measured values, and value-sets with low error ($\pm 10\%$ of experimental values) were selected as “points” in the module space. Principal components analysis (PCA) was then used to transform these points to a new coordinate set that maximally describes the variability in the module. **(B)** Simulated release of intracellular Ca^{2+} for sampled points along the first principal component. Traces show the first second of Ca^{2+} release after a ten-fold increase in IP_3 . See Chapter 5 for a detailed description of the method.

at 1 nM in a 6 fL platelet constitutes less than 4 molecules, a level that can give rise to stochastic fluctuations.

Dramatic, asynchronous $[Ca^{2+}]_i$ spikes increased in frequency and amplitude after activation with ADP (Figure 4.7A,B). Before agonist stimulation, the resting platelet model showed occasional Ca^{2+} spikes with relatively small amplitude. Similar behavior was noted in a third of recorded platelets [63], which spiked on average 1.6 ± 0.2 /min without ADP stimulation. The onset of Ca^{2+} release after agonist addition was rapid in the model (≈ 5 s) and persisted for several minutes, also consistent with experimental results [63,64]. Key drivers in this asynchronous spiking were the channel open probability, the low copy number of IP_3R channels, the size of the cytosolic compartment (8 fL in the model), and the large Ca^{2+} gradient at the DTS membrane. To determine the frequency distribution of Ca^{2+} spiking, we followed exactly the empirical strategy used by Heemskerk et al. [64] to distinguish individual Ca^{2+} release events in a detailed study of video-imaged platelets (see Figure 4.8). There was a striking resemblance between simulation and experiment for the peak interval distribution (Figure 4.7C), with the most common intervals occurring at 6-8 s or 11-13 s. As platelet volume was artificially increased 100-fold in the stochastic simulation (Figure 4.7D), the Ca^{2+} spiking was lost and typical, smooth ADP responses (as in Figure 4.5B) for populations emerged. Thus, the asynchronous spiking appeared to be a result of the fundamentally stochastic nature of signal transduction in cells as small as human platelets.

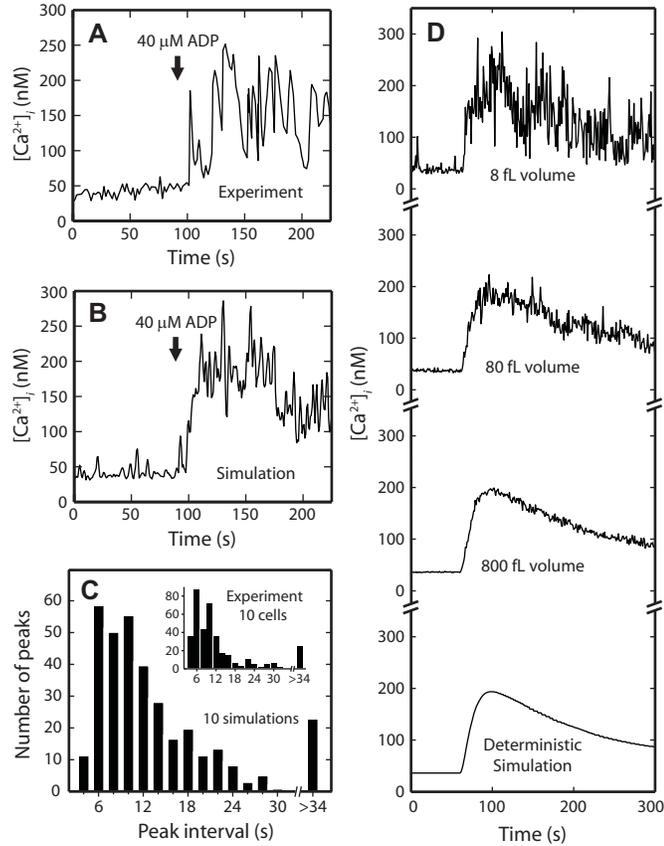


Figure 4.7: Stochastic simulation of a single platelet. (A) A single, fura-2-loaded platelet was immobilized on a fibrinogen-coated coverslip and activated with $40 \mu\text{M}$ ADP at $t = 90$. Ca^{2+}_i trace was obtained from Heemskerk et al [63]. (B) After 90 s of simulated rest, the platelet model was activated by setting extracellular [ADP] to $40 \mu\text{M}$. $[Ca^{2+}]_i$ was sampled every 0.1 s. (C) Single platelets were simulated for 10 min and $[Ca^{2+}]_i$ was sampled once per second. Ca^{2+} peaks were identified as described by Heemskerk [64]. Peak intervals between 5 and 30 s were counted from 10 simulations as described for video-imaging studies (see Figure 4.9). Interval times were binned in 2 s increments for direct comparison with experiment (*inset*). (D) The platelet model compartment sizes were artificially scaled to 1, 10, or 100 times their original sizes. Simulation results are shown for 300 s stochastic simulations with $20 \mu\text{M}$ ADP added at 90 s. The bottom plot shows results for the deterministic simulation.

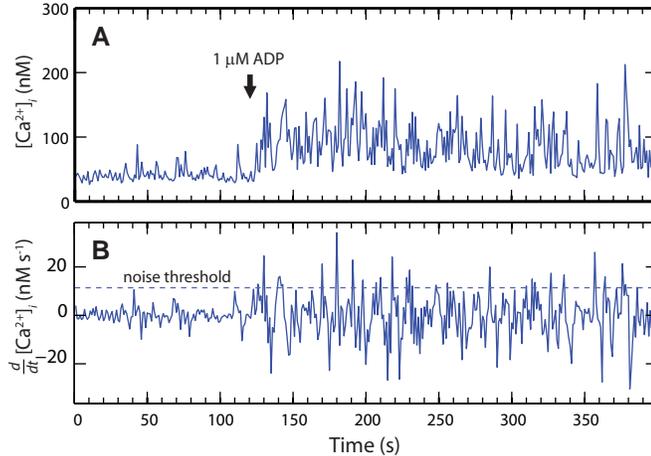


Figure 4.8: Calculation of Ca^{2+} spiking frequency distribution. **(A)** 600 s Ca^{2+}_i traces from 100 stochastic simulations were sampled every 1 s. **(B)** The derivative of each trace was calculated, and a noise threshold of 10 nM s^{-1} was used to count the peak-to-peak interval times for each trace.

4.4 Discussion

We deployed available kinetic data, electrochemical calculations, measurements of cell size, and modular organization of signaling function to estimate the physical structure and dynamical properties of ADP-stimulated Ca^{2+} release in human platelets. For the first time, a single model has accounted for several fundamental attributes of known platelet behavior: *(i)* In the presence of basal ADP levels, the model remained at steady state and accurately predicted the resting concentrations of Ca^{2+}_i , IP_3 , DAG, PA, PI, PIP, and PIP_2 . Under these conditions, the model underwent slow turnover of signaling molecules (e.g., IP_3 , PLC^* , and $\text{G}_q\text{-GTP}$), yet the rates of synthesis and degradation of these metabolites were exactly balanced. At the same time, Ca^{2+} was continually leaked from the DTS and actively pumped back into stores, maintaining a resting platelet $[\text{Ca}^{2+}]_i$ near 100 nM. This “active” homeostatic state is representative of the situation *in vivo*, since inhibition of SERCA pumps causes a gradual depletion of Ca^{2+} stores [150]. *(ii)* Increased [ADP] caused a transient rise in $[\text{Ca}^{2+}]_i$ and $[\text{IP}_3]$ with the correct dose response and time-course profiles. *(iii)* Stochastic simulation of the platelet model generated noisy Ca^{2+} release and reuptake behavior with a nearly instantaneous onset. *(iv)* Calculation of peak-to-peak interval times revealed a distribution of intervals favoring 6-8 and 11-13 s gaps. These results were strikingly similar to what was observed in video-imaged platelets and demonstrate that both averaged and stochastic behaviors may be accurately captured by the same molecular model. Importantly, the model fulfilled the dual requirement of maintaining homeostasis under resting conditions while remaining capable of generating an activation response in the presence of sufficiently high agonist concentration. This is

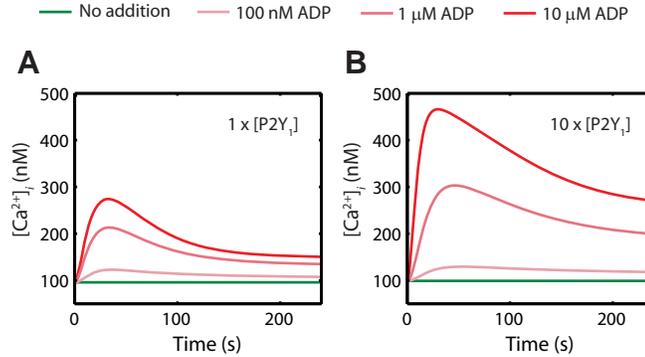


Figure 4.9: Increased P2Y₁ copy number enhances Ca²⁺ release in platelet model. The surface concentration of P2Y₁ in the best-fit platelet model was increased 10-fold and steady-state was reestablished. At $t = 0$, both models were activated with the indicated concentrations of extracellular ADP.

a critical quality to consider when modeling platelets, which must be sensitive enough to respond to injury without activating prematurely [17, 150].

Our analysis also produced several novel predictions: (i) The number of SERCA pumps was predicted to outnumber IP₃R channels (5000 IP₃R copies were predicted in most configurations). This was largely a consequence of the kinetic properties of the enzymes, which transport Ca²⁺ at significantly different rates. (ii) The relative volume of the platelet DTS was predicted to be 2-4% of the non-granular intracellular volume, both by kinetic modeling and image analysis of a DTS-stained platelet. (iii) Lower resting [IP₃] was associated with greater [Ca²⁺]_{dts} and more acute Ca²⁺ release. (iv) Restoration of basal PI levels following PLC-β* hydrolysis was predicted to occur in part through a negative-feedback mechanism in which PKC* phosphorylates PLC-β and thereby diminishes its hydrolytic activity. (v) By artificially increasing the volume of the simulated platelet, we were able to quench the fluctuation of Ca²⁺ spikes, achieving a response more similar to that of megakaryocytes, in which regular oscillations in [Ca²⁺]_i are observed with fixed amplitude and frequency [162]. Since mature megakaryocytes are more than 1000 times larger than platelets [69], these observations suggest that the asynchronous and stochastic Ca²⁺ response in platelets may simply be the consequence of a small intracellular volume.

The model also provided some insight into why thrombin is a more potent agonist than ADP. Both agonists act through G_q-coupled GPCRs [17], yet there are about 10 times as many thrombin receptors as ADP receptors [12, 17]. To determine whether the increased potency of thrombin may be due to greater receptor surface expression, we increased the number of P2Y₁ receptors in the model and examined the ADP response. A 10-fold increase in [P2Y₁] caused a 2-fold higher peak response and more sustained elevation of [Ca²⁺]_i (Figure 4.9), similar to the observed

difference between thrombin- and ADP-stimulated platelets. While a more thorough analysis will involve examining the coupling efficiency to G_q and the duration of the receptor active state, these calculations imply that the low copy number of $P2Y_1$ may be a limiting factor for ADP signaling.

Our estimation procedure generated 10 individual platelet models (unique sets of initial conditions representing the resting concentrations for the 70 species in the model) that each satisfied homeostasis constraints and fit all available data (Figure 4.10). These models may be thought of as individual cells in a platelet population that vary in their exact molecular makeup but are each capable of generating the same activation response. For certain species such as Ca^{2+}_i , DAG, and PIP_2 , the similarity among the models can be easily explained since these quantities were held fixed (with a small variability term) to the values reported in the literature. Other similarities in the concentrations, as for GTP and $PLC-\beta^*$, are more interesting because these values were not explicitly restricted by the estimation procedure. Rather, these uniformities reflect implicit constraints in the platelet model that emerged because of other species concentrations or signaling reactions that strongly control the production or consumption of these species. As such, these quantities are predicted to be rigid, or inflexible, “nodes” in the model [105] and are subject to further investigation as a new hypothesis revealed by the analysis. By contrast, we observed several instances of variability in concentrations (node flexibility) among the 10 models. For example, the concentrations of PKC and $G_q \cdot GTP$ both varied over 3 orders of magnitude. Similarly, the values for $P2Y_1$, SERCA, PMCA, $PLC-\beta$, Ins, IP_2 , and CDP·DAG varied appreciably. The large variability observed for many of the estimated phospholipids (e.g., Ins, IP_2 , and CDP·DAG) was due to allowable node flexibility and/or insufficient experimental data to constrain the estimation of the resting concentrations. Thus, while the essential steady-state and activation properties were accurately captured by the module as a whole (Figure 4.4A), additional experimental data will be necessary to generate reliable estimates for individual phosphoinositides. Finally, the estimated number of IP_3R channels was generally low (≈ 2500 channels/cell) as expected from the analysis of the Ca^{2+} module. Higher $[Ca^{2+}]_{dts}$ appeared to be compensated by higher SERCA levels.

Through a computational analysis of the human platelet, we have demonstrated the successful integration of heterogeneous data sets and legacy knowledge to construct a predictive model of platelet homeostasis and activation. This assembly of molecular detail represents, for the first time, a substantial description of the platelet signaling apparatus and was capable of reproducing diverse experimental observations as well as providing specific, testable hypotheses regarding platelet enzymatic regulation, the physical structure of the cell, and the dynamics of intracellular calcium release. Specifically, this is the first work to provide a quantitative molecular explanation of the asynchronous calcium spiking observed in stimulated human platelets. By incorporating additional

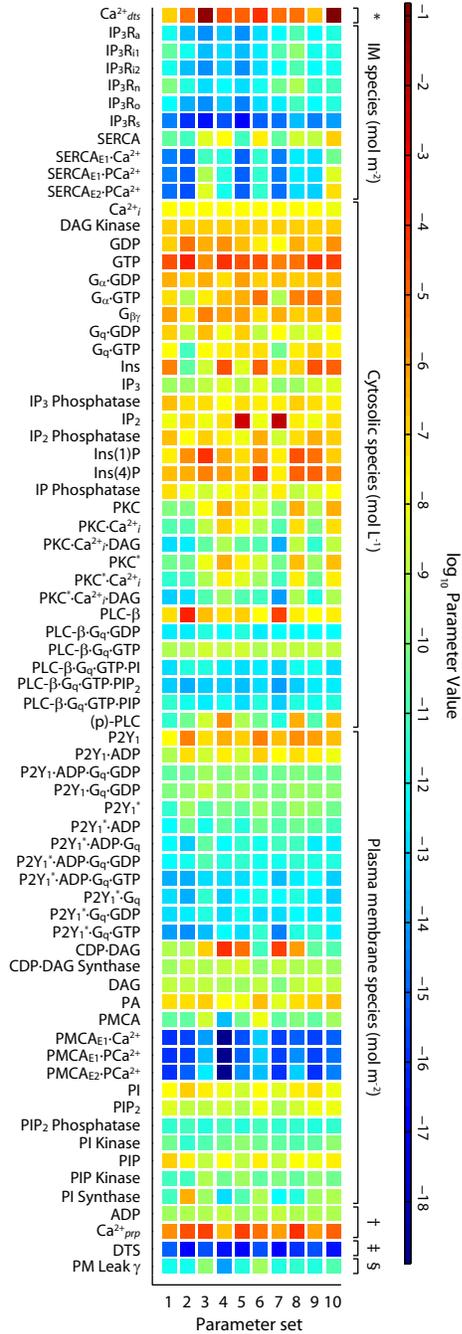


Figure 4.10: IC map for 10 best-fit platelet models. Raw values for model quantities (steady-state concentrations, compartment sizes, or kinetic parameters) are color-coded according to the indicated units for each value type. Note color bar on right-hand side of figure. In most cases, colors correspond to molar concentrations (mol/L or mol/m², see footnote). For example, Ca²⁺_{dtS} values shaded orange correspond to molar values of 110 μM. Many quantities displayed a narrow range of values (e.g., PLC-β-G_q-GTP, IP₃, DAG Kinase) that reflect either imposed restrictions based on experimental evidence or implicit constraints due to the model structure (see Section 4.4). Other quantities varied considerably (e.g., IP₂ and CDP-DAG). Species are grouped according to compartment. *DTS species (mol L⁻¹). †Extracellular species (mol L⁻¹). ‡DTS volume (L). §PM leak conductance/area (S m⁻²).

experimental data, future versions of the model can include mechanisms for store-operated Ca^{2+} influx mechanisms, the acidic Ca^{2+} store, integrin activation, granule secretion, phosphatidylserine exposure, as well as additional receptors for prostacyclin, nitric oxide, thromboxane, epinephrine, thrombin, fibrinogen, and collagen. As experimental investigation of platelet function continues, new data will be used to challenge, confirm, and expand the platelet signaling model presented here.

Chapter 5

Steady-State Kinetic Modeling

5.1 Introduction

Computational models help quantify the reaction dynamics and regulatory modes in complex biochemical systems [4,7,9,13,50], particularly when a system is so intricate that its behavior cannot be predicted by intuition alone. The building blocks for constructing large reaction networks are often available in numerous databases [10,83,108,152] and journal archives. Here, one can obtain many of the experimentally-derived elementary reaction steps, kinetic constants, or rate laws for individual steps in a given biochemical system or pathway. Despite this wealth of information, however, compiling these data to construct models with accurate system-wide behavior represents a significant challenge in systems biology [70,81]. Comprehensive models of metabolism have been successfully developed for microbial systems [49,50,139] and a few eukaryotic cell types [145,155,175]. These constraint-based models [88] are often represented by stoichiometric networks that lack an explicit description of substrate concentrations, reaction mechanisms, or the transient behavior of the system. Although various strategies have been proposed to incorporate these features into large-scale models [124,153], the task of assembling complex kinetic models with nonlinear dynamics remains a difficult problem. One of the major obstacles to building accurate kinetic models is the number of unknown parameters in the model that must be estimated using experimental datasets [153], which themselves are often massive, incomplete, noisy, and/or imperfect [76]. A number of parameter estimation methods, such as genetic programming, simulated annealing, and various gradient-based routines [111,127], have been proposed to infer unknown quantities in biochemical models. Most of these methods address the problem of estimation in purely abstract terms and do not take into account the unique mathematical features of biochemical systems, such as a well-characterized kinetic

subsystem (e.g., the dynamics properties of an ion channel [154]). Estimated parameters must still meet constraints imposed by the other experimentally measured parameters in the model.

To address these challenges, we propose a strategy for assembling large kinetic networks that retain the nonlinear dynamics governing individual reactions in the system. The key features of the method are: (i) restriction of steady-state values by subsystem kinetics, (ii) reduction of the steady-state solution space by principal component analysis (PCA), and (iii) combination of independently constructed submodels (modules). The first feature is a Monte Carlo sampling over unknown concentrations with fixed kinetic parameters derived from the literature. The opposite strategy has been used in microbial systems to restrict kinetic parameters based on species concentrations [49]. The second feature, reduction of the steady-state space by PCA, has been applied previously for metabolic systems described by a stoichiometry matrix [50, 139], but not, to our knowledge, for nonlinear systems. In the last step, a full model representation is assembled by combining PCA-reduced, steady-state solutions from each module to form a combined steady-state solution space for the entire system. This global space may then be searched for solutions with accurate time-dependent behavior using any number of established routines [61, 127].

The method exploits three properties common to many biological systems: modularity, homeostasis, and known quantitative kinetic relationships among interacting molecular components. Interestingly, this physiology-inspired approach enforces natural constraints on the range of allowable system states and allows one to monitor shifts in steady states due to kinetic perturbations. To illustrate the method with an example, we show how 77 reactions from 17 primary data sources were integrated to construct an accurate model of intracellular calcium and phosphoinositide metabolism in the resting and activated human platelet. Finally, we extend our analysis of this modeling approach by examining the steady-state characteristics of a system that is affected by changes in kinetic rate constants.

5.2 Materials and Methods

5.2.1 Model Definitions and Requirements

Our method builds upon a common representation of biochemical reaction networks [38] consisting of a system of ordinary differential equations (ODEs). In this paradigm, the concentration of each molecule in the system changes with time as a function of the instantaneous values of other concentrations and fixed kinetic parameters in the model. We separate this model description into two parts: The concentration vector (CV) of the model refers to the set of all molecule concentrations at a given instant in time and is denoted by the vector \mathbf{c} :

$$\mathbf{c} = (c_1, c_2, \dots, c_n) \quad (5.1)$$

The model *topology* refers to the entire set of kinetic parameters and rate equations that determine how these concentrations evolve with time. Mathematically, this is represented by the vector function \mathbf{f} , which defines the rate of change of \mathbf{c} with time as a function of the model concentrations and rate parameters:

$$\frac{d\mathbf{c}}{dt} = \mathbf{f}(\mathbf{c}) = \begin{bmatrix} f_1(\mathbf{c}) \\ f_2(\mathbf{c}) \\ \dots \\ f_n(\mathbf{c}) \end{bmatrix} \quad (5.2)$$

The functional form of each f_i is a sum of rate equations for each reaction that consumes or produces c_i and will generally vary for each molecule. Typical functional forms for f may include, for example, a series of Michaelis-Menten or nonlinear rate expressions. A simple reaction topology is shown in Figure 5.1A with corresponding ODEs in Figure 5.1B. It is useful to separate a large model into two or more *modules* with subset CVs that overlap at reaction edges, as shown in Figure 5.1A.

Often, the topology of a biological system is better characterized than its CV [88]. For example, the major protein-protein interactions in a signaling pathway may be deduced from mutation or knock-out studies, providing a molecular wiring diagram that links together the various components in the network. For each of these interactions, purified enzymes may be used to measure the strength of the interaction *in vitro* or to measure the rate of some enzyme-catalyzed reaction in the system. An important caveat is that the kinetic rate constants within the cellular milieu (the cell context) may be different from those obtained in an *in vitro* experiment with purified components. In contrast, it is generally more difficult to accurately measure the absolute abundance of intracellular enzymes or metabolites *in vivo*, although progress is being made in this area [107]. Our method thus assumes that the topology of a given system is known and that the unknown set of concentrations exists in a linear space of dimension n in which each species c_i comprises a separate dimension (Figure 5.1C). The ultimate goal of the method is to efficiently search this *concentration space* to find a set of values that, when combined with the fixed topology, renders the full model consistent with known resting states and experimental time-series data obtained by perturbation of the cell.

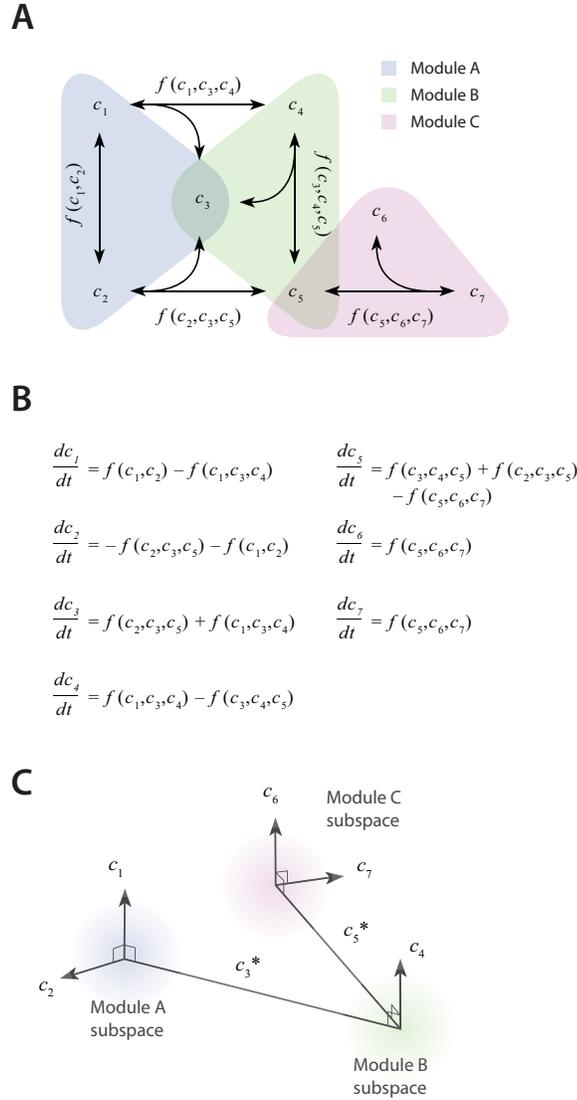


Figure 5.1: Example structure of model topology, ODEs, and concentration space. **(A)** The model topology defines the state transitions (arrows) and rate equations (f) that determine how molecules are interconverted. This example model is organized into three overlapping modules, with molecules c_3 and c_5 each occurring in two modules. Corresponding **(B)** ODEs and **(C)** concentration space for the example topology in panel A. Each of the 7 molecules occupies a separate linear dimension, with each module comprising a subspace of the full 7-dimensional space. Modules that share a common molecule have intersecting subspaces.

5.3 Homeostasis and the steady-state assumption

A special situation arises when $d\mathbf{c}/dt = 0$ in equation 5.2. Under these conditions, the model is said to be at steady state, and the vector \mathbf{c}_{ss} is a steady-state solution to the system of ODEs. If \mathbf{f} contains nonlinear terms, there may be an infinite number of steady-state solutions for the system of ODEs [38]. This set of solutions occupies some nonlinear subspace of the concentration space exemplified in Figure 5.1C. To guarantee that nonzero steady-state solutions may be found, the method requires the model topology (and all module topologies) to be balanced, meaning that the production and consumption of each molecule must be equal so that the total mass of the system is conserved. This steady-state assumption [88] is a common constraint in stoichiometric modeling and metabolic flux analysis and is conceptually related to the biological phenomenon of homeostasis [68], in which opposing processes are coordinated to maintain the stability of a cell or organism. For example, a nerve cell may maintain a constant electrochemical gradient by continually transporting ions across a lipid membrane.

5.4 Results and Discussion

5.4.1 Reduction of Modular Kinetic Networks

The first phase of the method involves generating a compact representation of the steady-state solutions for each module. The steps for module reduction are outlined in Figure 5.2A. First, conservative bounds are chosen for \mathbf{c} based on physiological and practical considerations. For example, a regulatory enzyme is expected to be present in at least one copy per cell and not to exceed an intracellular concentration of one molar. Knowledge about the physical size of the system is useful in this step to convert a raw copy number to a concentration. For small systems, this information can provide a rigid lower bound on unknown concentrations (see Chapter 4). For example, a single molecule in a 6 fL platelet has a concentration of 4 nM. Also, because molecular concentrations can span several orders of magnitude, it is often more efficient to delineate this range of values on a logarithmic scale rather than a linear scale.

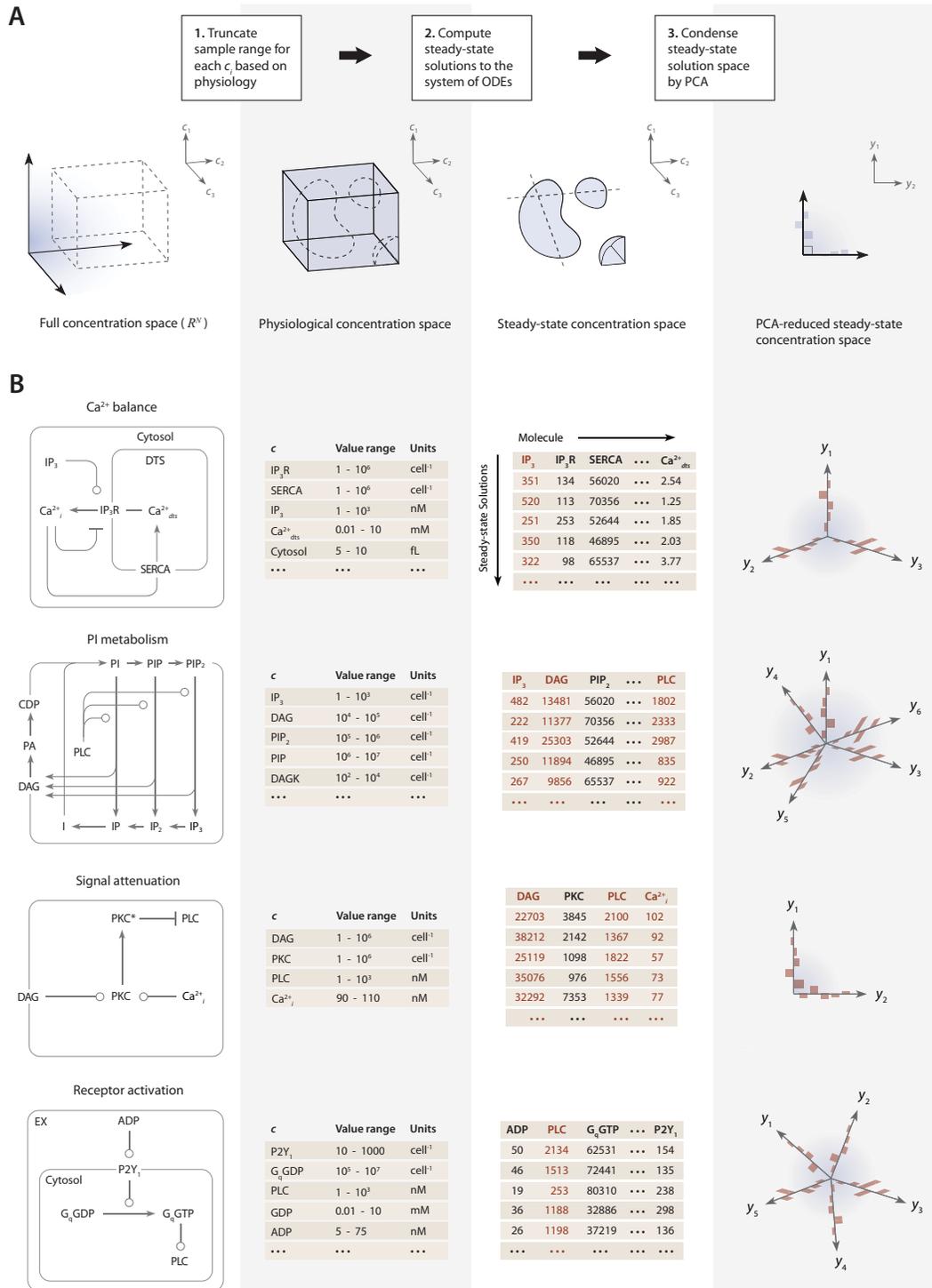


Figure 5.2: Steps in dimensionality reduction of steady-state modules and example from platelet signaling model. **(A)** Steps in dimensionality reduction of kinetic modules: (1) Restrict value ranges for each c_i to physiologically realistic ranges. (2) Compute multiple steady-state solutions to the model ODEs using initial guesses sampled randomly from the defined distribution. (3) Reduce the dimensionality of the steady-state solution set by PCA. **(B)** Results obtained from modular reduction 4 kinetic modules in a platelet signaling model. For each module, a fixed topology was combined with initial guesses from the defined distribution and simulated until equilibrium was reached ($dc/dt < 10^{-18}$) using 10^9 initial guesses for \mathbf{c}_0 . Specific concentrations within these steady-state solutions were compared to experimentally measured values, and solutions with low error ($\pm 10\%$ of known concentration values) for these elements were selected as “points” in the steady-state concentration space. PCA was then applied to transform these points to a new coordinate set that maximally covers the space of steady-state solutions.

Once the sampling distribution for \mathbf{c} has been defined, steady-state solutions (\mathbf{c}_{ss}) for each module are calculated using fixed kinetic parameters for each reaction in the module obtained from the literature [10, 108, 152], novel kinetic experiments, or estimation. For this step, each initial guess is sampled from the distribution for \mathbf{c} and combined with the predetermined topology. The combination of fixed rate equations, fixed parameters, and forms a well-posed initial value problem,

$$\frac{d\mathbf{c}}{dt} = \mathbf{f}(t, \mathbf{c}), \mathbf{c}(t_0) = \mathbf{c}_0 \quad (5.3)$$

that may be computed using a numerical solver [34]. For non-oscillating systems, steady-state solutions may be obtained by simulating the system until equilibrium is reached (i.e., until $d\mathbf{c}/dt \approx \mathbf{0}$). Alternatively, one may use any number of multidimensional root-finding routines, such as those available in the GNU Scientific Library [58], to find the closest n-dimensional root to the vector function \mathbf{f} using starting guess \mathbf{c}_0 .

In the third step, a large collection of steady-state solutions for each module is subjected to principal component analysis (PCA). A sample size of 1000 points per unknown concentration is generally sufficient to minimize error due to over-fitting [62]. PCA is then used to transform these points to a new coordinate set that optimally covers the space of steady-state solutions using the fewest number of dimensions. For example, if two molecule concentrations in the steady-state space are highly correlated due to participation in the same reaction, PCA will locate a single dimension to represent each pair of points in the transformed space. Ultimately, these new dimensions will be combined across all modules to search for global solutions that lie in the steady-state space for the fully combined network. Since PCA is a linear method, a steady-state solution space that is highly nonlinear may require more principal component vectors to accurately estimate the solutions. Nonlinear methods of dimensionality reduction, such as kernel PCA [96] or local linear embedding [134], may provide a more compact representation of steady-state solutions spaces in future iterations of the method.

The reduction procedure is illustrated with an example of a human platelet model comprising 4 interlinked signaling modules (Figure 5.2B). For each module, we used published reaction mechanisms and kinetic parameters to construct the module topologies [129]. Each topology was held fixed while the unknown CVs were sampled from empirically-defined distributions. For this step, we generated more than 10^9 sets of initial guesses (\mathbf{c}_0) for each module, computed the initial value problem for each \mathbf{c}_0 until a steady state was reached ($d\mathbf{c}/dt \approx \mathbf{0}$), and selected only those steady-state CVs (\mathbf{c}_{ss}) that were consistent with known concentrations. For example, the concentration of intracellular Ca^{2+} ($[\text{Ca}^{2+}]_i$, Figure 5.2B) in platelets is known to be $\tilde{100}$ nM. Thus, only those with $[\text{Ca}^{2+}]_i \approx 100$ nM were kept as part of the steady-state solution space for the Ca^{2+} balance module. This procedure was used to generate 10,000 steady state solutions for each module for subsequent reduction by PCA. A minimal set of principal component (PC) vectors (those capturing 90% or more of the variance in the solution set) were used as search directions in the final estimation step, in which the transient behavior of the perturbed steady-state was compared to experimental time-series data.

Interestingly, only a small fraction of initial guesses produce steady-state solutions that are also consistent with known concentration values. For example, it was previously shown that only 50,000 of 10^9 initial guesses (0.005%) in the Ca^{2+} balance module (Figure 5.2B) met both requirements and were suitable for further analysis [129]. Among this set of CVs, marginal distributions for individual molecules were often confined to a narrow range of values. As an example, 80% of steady-state solutions for the calcium module contained $\lesssim 1000$ IP_3 molecules/cell, although initial guesses were sampled uniformly between 1 and 10^6 molecules/cell. This observation shows that the kinetic topology of these molecular networks places very strong constraints on the range of concentrations that can exist at steady state. In biological terms, this suggests that fixed kinetic properties at the molecular level (e.g., IP_3R and SERCA kinetics) can affect not only the dynamical features of a biochemical system but can also determine the abundance of chemical species and the compartmental structures that contain them.

5.4.2 Merging Steady-State Modules

In the final step of the method, the full model is assembled by combining PCA-reduced, steady-state solution spaces from each module into a combined steady-state solution space for the entire system (Figure 5.3A). This global space is searched for full-length, steady-state solution vectors that satisfy both the individual steady-state requirements of each module and the desired time-dependent properties when the steady-state is perturbed (for example, by increasing the initial concentration of a signaling molecule). For the platelet signaling model, consisting of 77 reactions,

132 fixed kinetic parameters, and 70 species [129], a set of 16 PC vectors representing all 72 unknown variables (70 molecule concentrations, 1 compartment size, and 1 rate constant) in the model were used as search directions in a global optimization routine. The global solution space was searched for models with accurate dynamic behavior using experimental time-series data for ADP-stimulated Ca^{2+} release (Figure 5.3A). Equality constraints are imposed during optimization to maintain consistent concentrations of molecules that are present in more than one module. Specifically, for a steady-state space \mathbf{A} represented by m PC vectors and a steady-state space \mathbf{B} represented by n PC vectors, the projections of each space onto must be equal,

$$\mathbf{A}\mathbf{c}_i = \mathbf{B}\mathbf{c}_i \tag{5.4}$$

where \mathbf{c}_i is the unit vector for the shared molecule, c_i . This condition forms a linearly-constrained optimization problem for which a number of efficient routines exist [127]. We used the Asynchronous Parallel Pattern Search (APPSPACK) to perform a derivative-free optimization of the platelet signaling model [61]. A least-squares objective function was used to score the difference between simulated (after perturbation of steady state) and experimental time-series data points. One of the high-scoring steady-state solution vectors for the full model is shown in Figure 5.3B, along with individual steady-state vectors for each of the four modules. This 72-dimensional vector (*i*) satisfies the homeostasis constraint in that it is a steady-state solution, (*ii*) is consistent with the known steady-state levels for 8 of the molecules in the 72-dimensional space, and (*iii*) predicts the entire dynamic Ca^{2+} and IP_3 response of platelets exposed to ADP (0100 μM). Additionally, rigid and flexible nodes (steady-state concentrations) in this 72-dimensional space were readily identified when a set of allowable steady-state solution vectors are compared [129].

5.4.3 Applying the Method To Monitor Cellular Resting States

Resting systems remain in a steady state by the coordinated action of opposing but balanced kinetic processes. Thus, in general, altering one or more of these rate processes (e.g., increasing the catalytic rate of a reaction) should upset the balance of the system and cause it to adopt a new steady state. Various cell types have been shown to have altered steady-state properties because of mutations that affect the constitutive rates of reactions. For example, patients with type 1 diabetes harbor more Ca^{2+} ATPase activity in their platelets than healthy volunteers and experience high resting levels of intracellular Ca^{2+} [24]. In a separate case, a mutation within the tyrosine kinase domain of epidermal growth factor receptor causes significantly higher basal (growth factor-independent) tyrosine phosphorylation levels than the wild-type receptor [26]. Therefore, to examine the changes in steady-state properties caused by kinetic perturbations in our example

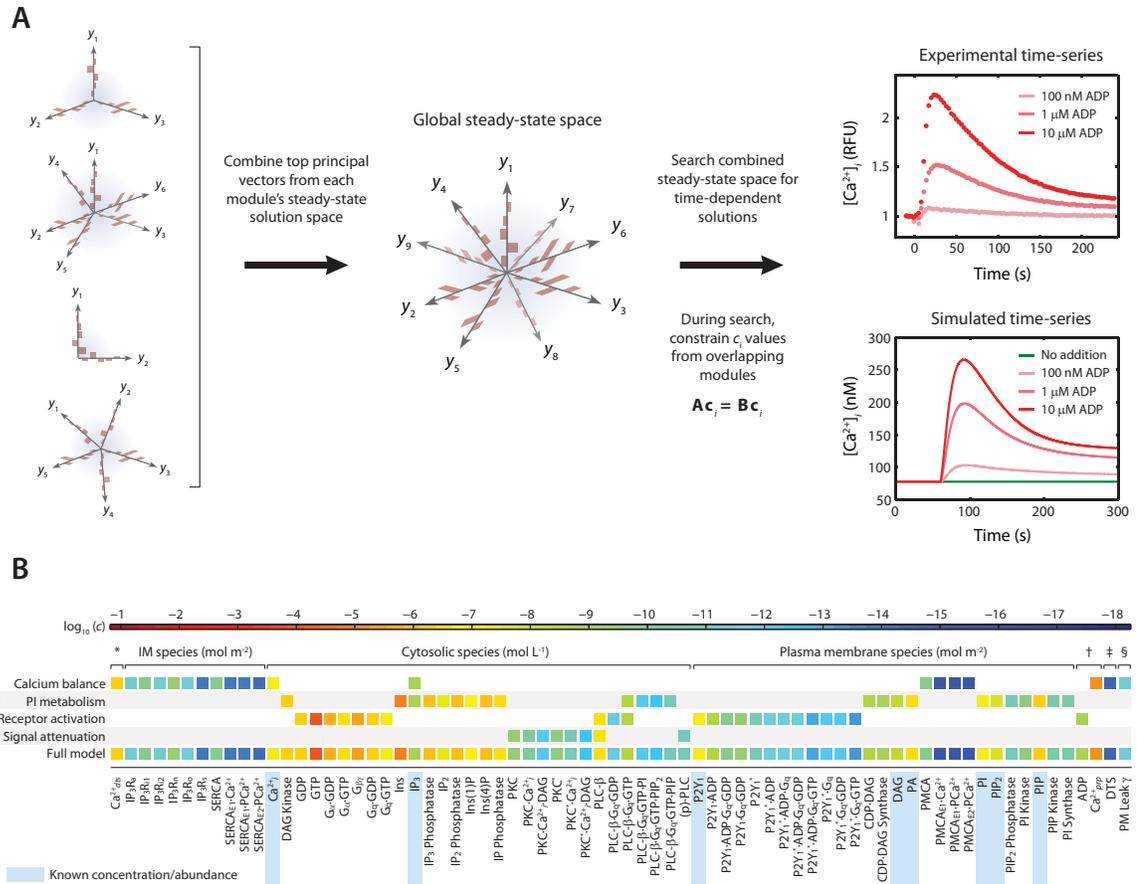


Figure 5.3: Assembly of full model from steady-state modules. (A) The full model is assembled by combining PCA-reduced, steady-state solution spaces from each module into a combined steady-state solution space. This global space is searched for full-length, steady-state solution vectors that satisfy both the steady-state requirements of each module and the desired time-dependent properties when the steady-state is perturbed (in this example, by increasing the concentration of the signaling molecule ADP and measuring the change in intracellular Ca²⁺ concentration). A simple linear constraint is imposed for every pair of modules that share a common molecule c_i to ensure that steady-state solutions are consistent. (B) To assemble the platelet signaling model, a set of 16 PC vectors representing all 72 unknown variables in the model were used as search directions in a global optimization routine. The global solution space was searched for models with accurate dynamic behavior using experimental time-series data for ADP-stimulated Ca²⁺ release. Species are grouped according to compartment. Color values correspond to molar concentrations (mol/L or mol/m²) or as indicated: * DTS species (mol L⁻¹). † Extracellular species (mol L⁻¹). ‡ DTS volume (L). § PM leak conductance/area (S m⁻²).

model, we altered the rates of 3 important regulatory reactions and observed the system response to each perturbation. Each perturbation cause a brief adjustment phase lasting $\tilde{200}$ s followed by a more gradual phase characterized by a new steady-state profile (Figure 5.4, *left*). After 1 hr of simulated time, steady-state concentrations and reaction fluxes were quantified relative to their original steady-state levels (Figure 5.4, *right*).

As expected, increasing the rate of Ca^{2+} release from intracellular stores resulted in higher cytosolic Ca^{2+} levels (7-fold increase) and 10-fold greater pumping activity by plasma membrane Ca^{2+} pumps (PMCA), although the new steady-state Ca^{2+} release flux remained relatively unchanged (Figure

5.4.4 Method Application, Computational Efficiency, and Extensions

We have presented a novel strategy for enumerating permissible steady-state solutions to fixed kinetic topologies and combining these solutions spaces to form large kinetic models. This is a practical strategy because kinetic parameters are commonly reported whereas absolute concentrations are not (see, for example, refs. [10, 108, 152]). The method extends the capability to build “genome-scale” models [50, 53, 70, 81] to include nonlinear kinetic features. Through application of the method, we have also explored the implicit restrictions on steady-state solutions that can be imposed by the underlying kinetic structures within a system [13]. This is useful from a physiological standpoint since the regulation and distribution of molecular species in living systems is largely regulated by the coordinated action of synthetic, degrading, and transporting enzymes.

The proposed method requires the model to fulfill a steady-state assumption (i.e., the model must contain nontrivial steady states) even if the system is typically characterized by transient behavior. It is precisely this requirement that allows the model to have the dual functional behavior observed in many biological contexts, such as in cellular signaling responses. At very low levels of activating signal, the model remains at rest by quenching the low level of activating signal through feedback mechanisms or futile cycling. When activating signals are increased, the system responds with the appropriate transient signaling behavior. As an example, a human platelet must remain quiescent under normal circulating conditions, tolerating a number of fluctuations in its surrounding chemical and physical environment. In the presence of the appropriate stimulus, however, it must be able to respond rapidly to bleeding conditions and trigger a precise program of molecular signaling events. Developing a mathematical model that is consistent with two or more biological behaviors is analogous to writing a single equations that has multiple solutions, each dependent on a given set of initial conditions and parameter values.

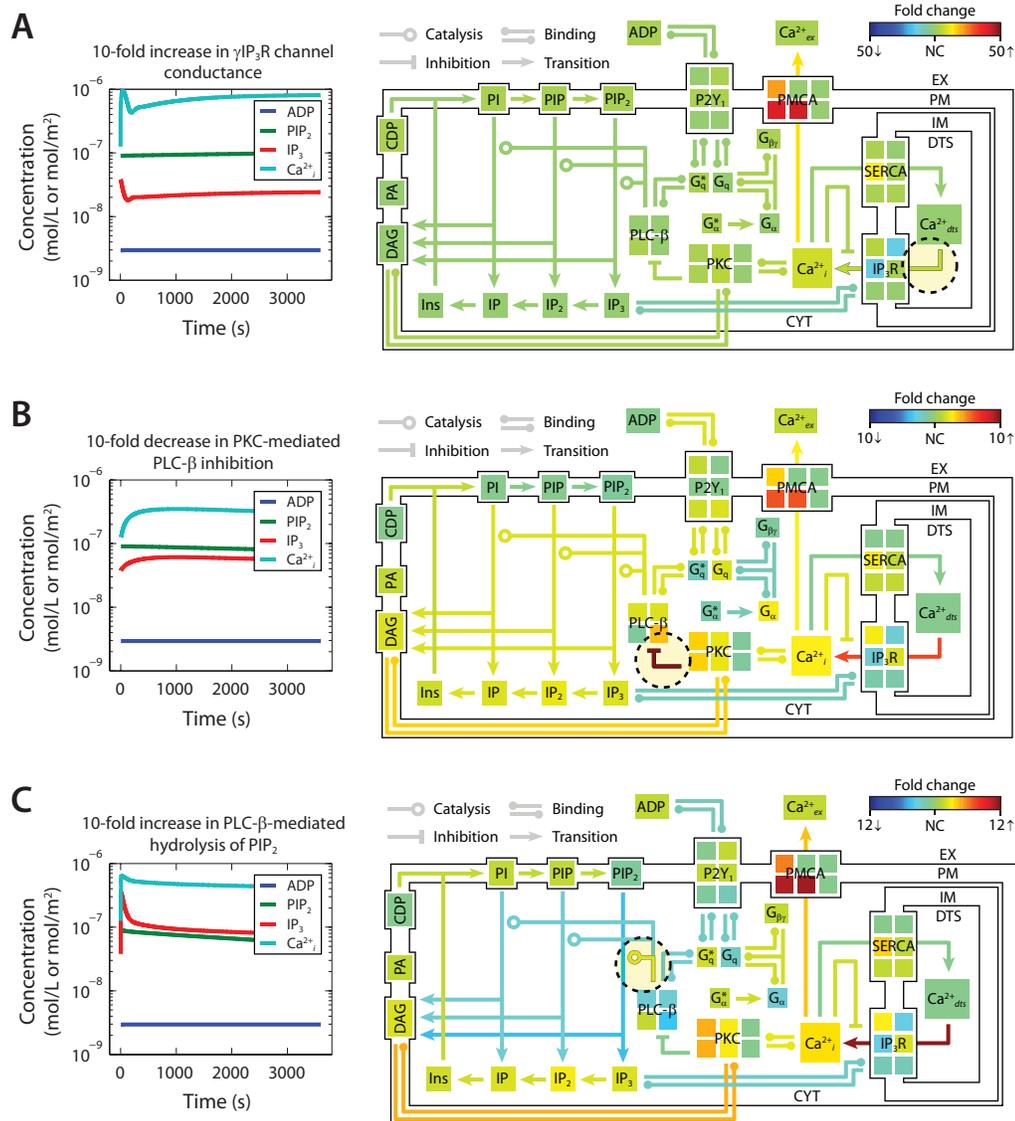


Figure 5.4: Shifts in steady-state profiles caused by kinetic perturbations. The steady-state platelet model was perturbed by changing selected kinetic parameters (± 10 -fold) and simulating for 1 h (*left panels*). After approaching a new steady state, the model concentrations and fluxes were determined relative to their original steady-state values and colored according to fold-change (*right panels*). Green indicates no change (NC) relative to initial flux/concentration. Red indicates a relative increase and blue indicates a relative decrease. Note that the color scale in each panel is normalized separately to maximize distinctions in fold change. New steady states were achieved after a (A) 10-fold increase in Ca^{2+} release through open IP₃R channels [129], (B) 10-fold decrease in PKC-mediated inhibition of PLC- β , and (C) 10-fold increase in PIP₂ hydrolysis (10-fold increase in k_{cat} of hydrolysis). Reactions with perturbed rate constants are circled and correspond to reaction mechanisms from ref [145]. (A) $\text{Ca}^{2+}_{dts} \rightarrow \text{Ca}^{2+}_i$, (B) $\text{PKC}^* + \text{PLC-}\beta \rightarrow \text{PKC}^* + \text{pPKC-}\beta$, (C) $\text{PLC-}\beta^* + \text{PIP}_2 \rightarrow \text{PLC-}\beta^* + \text{IP}_3 + \text{DAG}$.

Our approach differs critically from metabolic flux analysis and previous genome-scale metabolic network reconstructions [50,139] because it accommodates nonlinear terms that describe the dynamic behavior of each reaction in the system. Previous large-scale network reconstructions typically use a stoichiometry matrix to represent the gross flux of metabolites in the system [88]. Here, we have preserved the mathematical form of each kinetic rate equations as reported in the literature, allowing models to be built from existing data in a “bottom-up” fashion [70] while still allowing calibration to whole-system experimental data. This feature will substantially improve the accuracy of dynamical system simulation and parameter estimation.

Additional computational savings are provided through modularization. When estimating modules of modest size (5 or less unknown concentrations), we use a brute-force Monte Carlo approach to densely sample the feasible space of initial conditions. Larger networks (20 or more unknowns) cannot be efficiently searched in this brute-force manner, but can be built piecewise by combining subspaces of smaller size that have been densely sampled. Using the naive Monte Carlo approach, estimating n free parameters is exponential in n . By dividing these parameters into k independent networks, each with n/k free parameters, the estimation procedure becomes exponential in n/k and thus more tractable. By assembling the entire system from smaller, more manageable kinetic modules, data may be used to test the functionality of individual modules before incorporating them into the entire system. In several cases, this approach was shown to offer a substantial computational benefit (e.g., reducing the global search space by over 10,000-fold) by simply requiring a steady-state solution with known subcomponent values. The search space can be reduced further by principal component analysis if there is correlation between free parameters within a module. This was found to be the case for enzymes that have opposing regulatory roles; increasing the levels in one enzyme required a similar increase in the other in order to preserve homeostasis. Lastly, modules sharing common components must hold the same value for that component, which imposes an additional constraint on the steady-state solutions (equation 5.4).

As presented, the method exploits known kinetic parameters to restrict unknown concentrations due to kinetic interactions. However, the method is equally valid for estimating unknown kinetic parameters and/or utilizing known concentrations. Both concentrations and kinetic parameters appear indistinguishably as nonlinear terms in the ordinary differential equations that describe the system (Figure 5.1B). Hence, it does not matter which types of values are known and which are estimated; the procedure is valid for mixed or incomplete sets of unknown values. The use of qualitative data may also be exploited by the method. For example, beginning with a large set of steady-state solutions for a given module, the size of the set may be reduced by determining which solutions in the set contain some qualitative behavior or function. In a previous application of the

method [129], a set of 10^9 steady-state solutions representing calcium balance in a resting platelet were divided into 3 groups, according to their qualitative response to increased IP_3 concentration (low, mild, and high response). Using this technique, the functional testing of steady-state modules may be used to eliminate a large subset of the original steady-state solution set. As another example, one may use data from a Western blot to establish the relative abundance between two proteins in the model. This qualitative information may be used to filter the steady-state solutions to a reduced set that is consistent with experimental results. This kinetically-driven, constraint-based approach, which combines a homeostasis requirement with known kinetic parameters and cellular concentrations, naturally enforces numerical limits on unknown system quantities.

Chapter 6

Conclusions and Future Directions

6.1 Summary of Results

The central theme of this work is the integration of complex data sets and mathematical modeling to understand and predict intracellular signaling responses. We have used a variety of mathematical methods—including ordinary differential equation modeling, principal component analysis, stochastic markov processes, artificial neural networks, digital image processing, and numerical optimization—to better understand the mechanistic basis of signaling in both growth factor-responsive cells and in human platelets. Our analyses have not only duplicated known signaling behaviors in these cell types, but have also predicted and confirmed new behaviors.

In Chapter 2, we sought to understand how mutations that change the kinetic properties of the tyrosine kinase domain of the EGFR can affect the cell's survival response. This was accomplished by monitoring changes in two downstream signaling proteins, ERK and Akt, which showed disproportionate levels of activation in cells bearing mutated receptors. Chapter 3 extended this analysis by studying the effect of tyrosine kinase inhibitors on both receptor behavior and downstream responses. Specifically, a gain of signaling flux through the PI3K/Akt pathway in *EGFR*^{L858R} mutants was strongly downregulated when the mutant was exposed to the tyrosine kinase inhibitor, erlotinib. Using a branched signaling model to quantify phosphorylation patterns in the EGF receptor, we showed that the EGFR mutants had increased binding to tyrosine kinase inhibitors (TKIs), enhanced phosphorylation of substrate tyrosine residues, and preferential activation of the Akt signaling pathway, a critical pathway for cell growth, survival, and motility. Future computational and experimental studies will be necessary to pinpoint the exact molecular players that mediate both this increased activation of Akt and the potent efficacy of the inhibitor.

In Chapter 4, we used a kinetic model of calcium and phosphoinositide regulation in human platelets to study P2Y₁ activation by adenosine diphosphate (ADP). The model accurately predicted: (1) steady-state resting concentrations for intracellular calcium, inositol 1,4,5-trisphosphate, diacylglycerol, phosphatidic acid, phosphatidylinositol, phosphatidylinositol phosphate, and phosphatidylinositol 4,5-bisphosphate; (2) transient increases in intracellular calcium, inositol 1,4,5-trisphosphate, and Gq-GTP in response to ADP; and (3) the volume of the platelet dense tubular system. A more stringent test of the model involved stochastic simulation of individual platelets, which display an asynchronous calcium spiking behavior in response to ADP. Simulations accurately reproduced the broad frequency distribution of measured spiking events and demonstrated that asynchronous spiking was a consequence of stochastic fluctuations resulting from the small volume of the platelet. The model also provided insights into possible mechanisms of negative-feedback signaling, the relative potency of platelet agonists, and cell-to-cell variation across platelet populations. This integrative approach to platelet biology offers a novel and complementary strategy to traditional reductionist methods.

Both of these studies made extensive use of ordinary differential equations to describe the system dynamics and relied on published kinetic parameters. Building on this modeling framework, Chapter 5 presented a new method for constructing large-scale kinetic networks. This method involved building kinetic models in stepwise fashion, beginning with small “resting” subnetworks that are combined to form larger networks with complex time-dependent behaviors. The key step in the method is the enforcement of a “homeostasis requirement” on all subnetworks ($d\mathbf{C}/dt = \mathbf{0}$), which ensures that all species concentrations remain constant over time even though each species may undergo a gradual turnover through coupled reactions. This constraint was shown to not only reduced the computational cost of fitting experimental time-series data, but also provided insight into limitations on system concentrations and architecture. This strategy afforded a biologically intuitive framework for building large-scale kinetic models of steady-state cellular systems and their dynamics.

6.2 Extensions and Future Work

The insights derived from these studies (which have employed a relatively modest collection of signaling pathways and reaction steps) indicate that increased biological knowledge may be gained by incorporating additional cellular processes to form larger and more complex models. In this section, we outline some additional projects that extend both the numerical methods and biological scope of the work presented in this dissertation.

6.2.1 *Extension: Modeling the Platelet Response to Combinatorial Agonists*

As we have seen, the platelet activation response requires a precise program of molecular signaling events, including the activation of multiple platelet surface receptors (G protein-coupled receptors, receptor tyrosine kinases, integrins), phospholipid hydrolysis, release of intracellular Ca^{2+} , cytoskeletal rearrangement leading to shape change, secretion of storage granules, and paracrine signaling between platelets [17]. When one or more of these signaling mechanisms is perturbed by genetic or exogenous factors [118], pathological thrombus formation can result [57].

It is important to note that—in the environment of a forming clot—platelets experience all of these signaling cues at the same time. For example, the potent platelet activator thrombin activates the G protein-coupled receptors (GPCRs) PAR1 and PAR4 on the platelet surface; exposed collagen on subendothelial tissue binds to the tyrosine kinase GPVI and integrin $\alpha_2\beta_1$; ADP and thromboxane A2 (TxA2), which are both released from the storage granules of activated platelets, activate the P2Y family of GPCRs or the TP receptors, respectively. Although it is well appreciated that these signaling events occur simultaneously *in vivo*, most of our understanding of platelet biology is based on studies that examine a single agonist in isolation. For example, Yee et al. [182] have used aggregometry to phenotype normal donors and classify some as hyperreactive based on their individual responsiveness to ADP, epinephrine, collagen, collagen-related peptide and ristocetin. While these studies have been useful for characterizing the effect of individual agonists and their receptors, they fall short of capturing the highly heterogeneous and dynamic platelet environment in flowing blood. A better understanding of platelet function *in vivo* will require the simultaneous measurement of multiple interacting signaling pathways.

To broach this challenge, we developed a high-throughput experimental platform that measures the human platelet response to arbitrary combinations of 6 major agonists (Figure 6.1A,B). Agonists tested were: convulxin (GPIV activator), ADP, thromboxane analog (U46619), PAR1 agonist peptide, PAR4 agonist peptide, and PGE2. The method yields high-resolution time-series measurements of intracellular calcium levels over the first few minutes of platelet activation. Intracellular calcium concentration is a logical choice for monitoring the platelet activation response because it is the common result of most signaling pathways and is critical to many of the processes that control activation, such as actin mobilization and integrin activation [150]. We first tested the platelet response to all pair-wise combinations of 6 agonists at 3 doses (0.1, 1, $10 \times \text{EC}_{50}$). These 135 calcium traces trained a 2-layer neural network (NN) (Figure 6.1C) to successfully predict the platelet response to sequential addition of agonists and all ternary combinations of convulxin, TRAP and ADP (Figure 6.2A,D). The NN model also identified 45 conditions of 4, 5, and 6 agonists predicted

to display a range of synergistic signaling (Figure 6.2B,C). These predictions were confirmed experimentally ($R = 0.88$) and revealed 4 highly synergistic conditions of high U46619/PGE₂ ratio (Figure 6.2E), consistent with the known thrombotic risk of COX-2 therapy. Finally, we identified unique patterns of pairwise agonist synergy among 4 healthy donors, providing a functional fingerprint of each donor’s individual platelet phenotype. Pairwise Agonist Scanning (PAS) provides a sensitive, patient-specific experimental/computational platform for understanding how an individual’s cells integrate multiple molecular signals.

6.2.2 *Extension: Quantifying the role of store-operated calcium entry in platelet signaling*

The platelet model presented in Chapter 4 did not contain a description for extracellular calcium, which plays an important role in sustaining the intracellular calcium levels during activation. Store-operated calcium entry (SOCE) is the process by which the depletion of Ca²⁺ from the endoplasmic reticulum activates Ca²⁺ entry through channels in the plasma membrane. SOCE has recently emerged as an important mechanism for many cellular processes including growth, motility, secretion, and gene expression [99]. This mechanism has also been detected in human platelets and shown to be essential for pathological thrombus formation [18]. However, little is known about the relative importance of SOCE in platelet Ca²⁺ signaling in response to soluble agonists. Platelets provide an ideal test system to study SOCE in vivo since they are easily obtained from healthy donors and may be readily assayed for intracellular Ca²⁺ using fluorogenic dyes. A full description of platelet Ca²⁺ regulation is necessary to understand the observed correlations between platelet Ca²⁺ handling and thrombus formation [18], diabetes [166] and hypertension [47].

Extending our work in Chapter 4, a kinetic model of SOCE, calibrated to measurements of Ca²⁺ release from internal and external platelet stores, could be used to predict dysfunctional Ca²⁺ flux in thrombosis. Experimentally, the relative contribution of extracellular Ca²⁺ to the platelet activation response may be quantified using a small panel of platelet agonists. The concentration of free extracellular Ca²⁺ may be manipulated by adjusting the concentration of EDTA in the medium. Time-course measurements of intracellular Ca²⁺ release will be generated for 3 major physiological agonists: thrombin, ADP, and collagen-related peptide. For each agonist, differences among these traces will show how intracellular calcium levels depend on the amount of extracellular calcium that enters the platelet via SOCE. One may then construct an ODE-based model that quantifies the role of SOCE in platelet calcium signaling (Figure 6.3). Building upon the existing platelet model, additional mechanisms for thrombin and collagen receptors as well as SOCE may be added. Sufficient kinetic data exists to construct the model. For example, store-operated channel activation

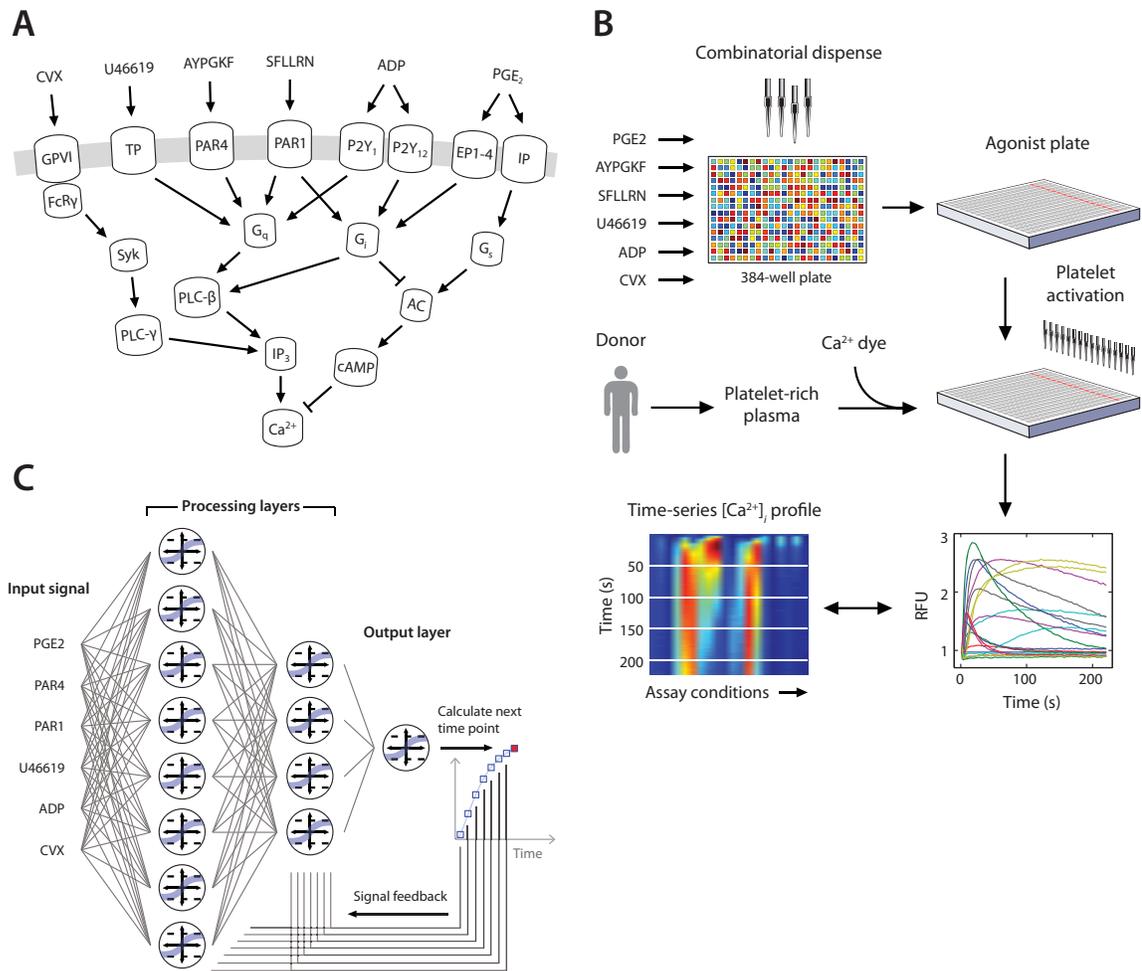


Figure 6.1: Pairwise agonist scanning to predict global calcium response in human platelets. (a) Simplified schematic of signaling pathways examined in this study that converge on intracellular calcium release in human platelets. (b) High-throughput combinatorial platelet activation procedure. An agonist plate containing combinations of signaling compounds is transferred to 384-well plate containing dye-loaded platelet-rich plasma. Time-course fluorescence changes were measured to quantify intracellular calcium concentrations. $[Ca^{2+}]_i$ transients may be represented as overlapping plots (*right*) or parallel heat maps (*left*). (c) Dynamic NN model used to train platelet response to combinatorial agonist activation. A sequence of input signals representing agonist concentrations is introduced to the network at each time point. Processing layers integrate input values with feedback signals to predict the next time point.

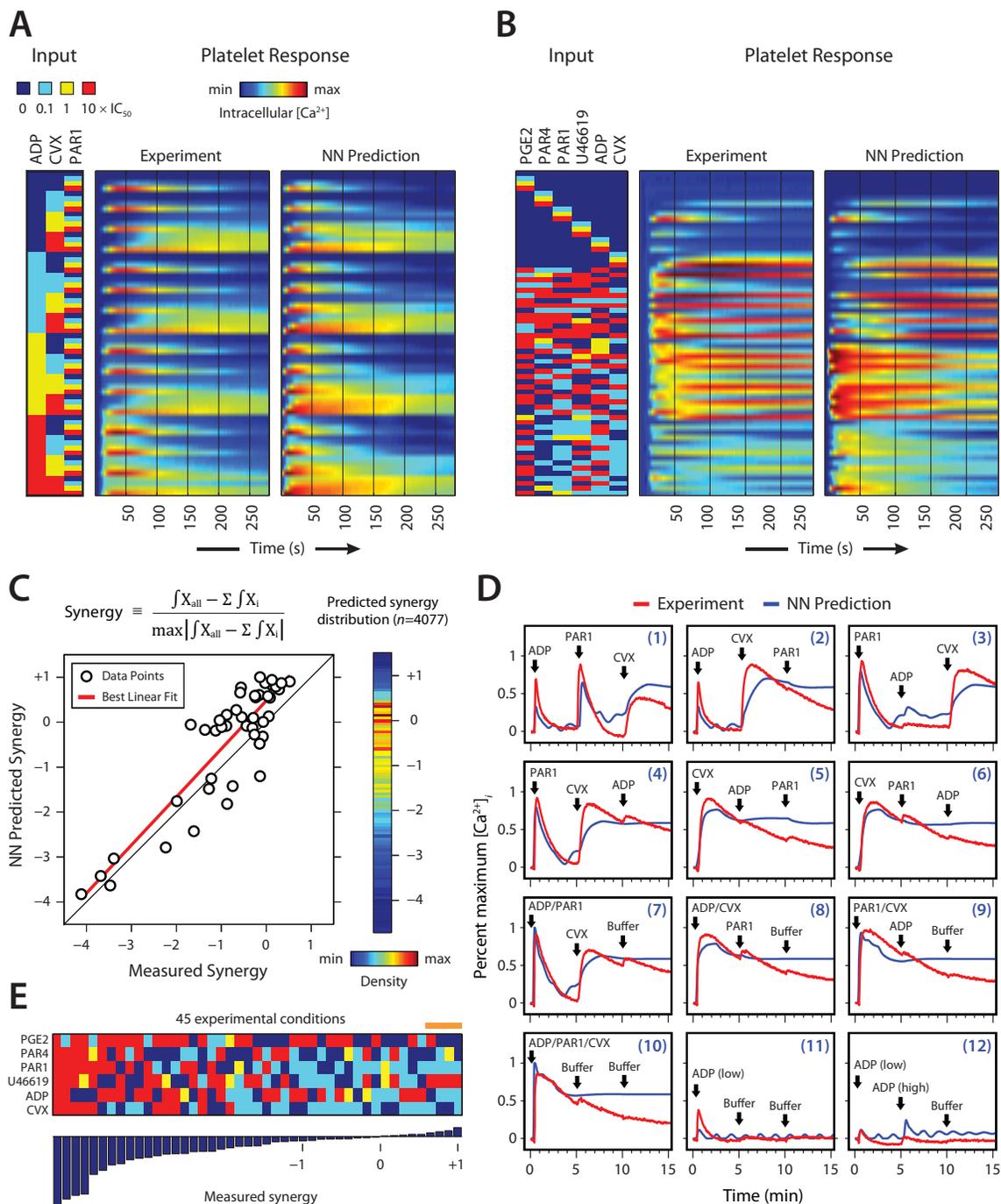


Figure 6.2: Experimental confirmation of global 6-dimensional platelet calcium response to combinatorial agonists. (A) All 64 ternary combinations of the agonists ADP, SFLLRN and CVX at 0, 0.1, 1 and $10 \times EC_{50}$ were measured and compared to the pairwise-trained NN model. *De novo* predictions of (B) time-course measurements and (C) synergistic signaling for 45 of 4077 possible combinations of 6 agonists were experimentally verified ($R = 0.88$). (D) Measured versus simulated responses to sequential additions of ADP, SFLLRN, and CVX. (E) The 4 highest synergies were observed for agonist mixtures that contained high concentrations of U46619, pointing to the potentiating effect of secreted TxA2 in an *in vivo* thrombotic setting.

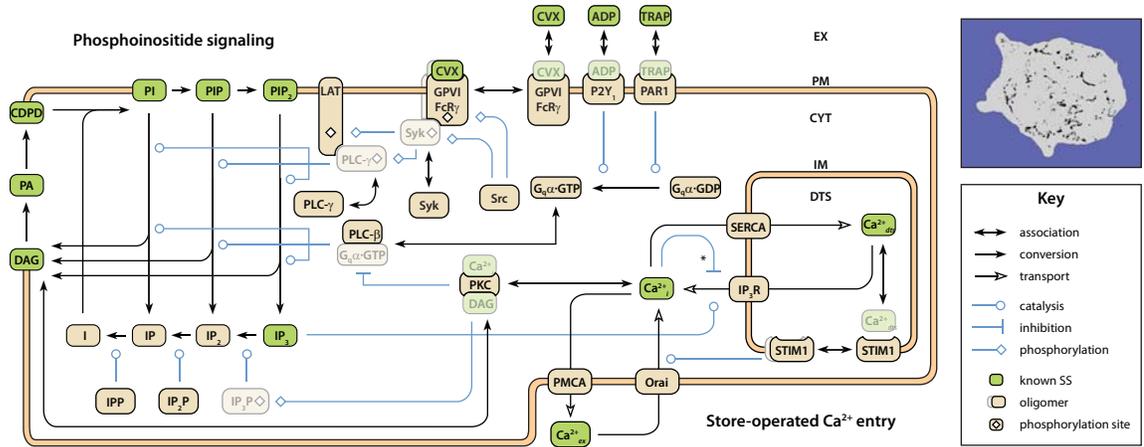


Figure 6.3: Schematic of kinetic model for quantifying platelet calcium regulation with mechanisms for store-operated calcium entry. (*Inset*) We will measure agonist-induced Ca^{2+} release in the presence of varying concentrations of extracellular Ca^{2+} (*blue area*).

was recently related to luminal Ca^{2+} concentration in human Jurkat leukaemic T cells [136]. These kinetic measurements will be used to supply an inward calcium current into the cytosol of the platelet model as calcium in the platelet internal stores is depleted.

6.2.3 *Future Work:* A signal processing model to predict apoptotic response

Cellular signaling networks may be probed experimentally by changing molecular “inputs” (e.g., the concentration of an activating ligand) and recording molecular outputs (e.g., a change in gene expression). The obvious challenge is to determine the processing “logic” [20] that occurs in between these experimental endpoints. Accomplishing this goal in a laboratory would require the spatial and temporal measurement of all of the molecular participants in a signaling response—a task that is clearly intractable using current methods of experimental analysis. An alternative strategy is to develop a computer simulation of the signaling system of interest, hoping that the simulation is accurate enough to reasonably approximate the true behavior of the system. Using an accurate simulation of signaling, one may harness the powerful analytical capability of numerical methods to slow down, speed up, perturb, calculate and analyze the quantitative elements in these systems, affording extended knowledge about how individual molecules interact to produce a signal. Various modeling strategies have been used to predict cellular responses to molecular signaling. Ordinary differential equation models [38] provide excellent mechanistic detail but require extensive knowledge of network topology and kinetic rate parameters. “Data-driven” approaches [74] such

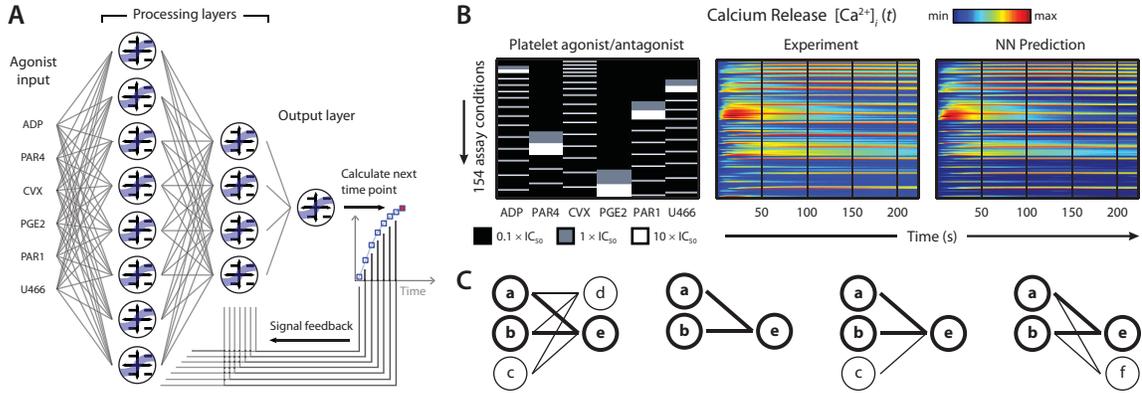


Figure 6.4: (A) NN trained to predict (B) Ca^{2+} response to all pairwise combinations of 6 agonists at 3 doses (unpublished data). (C) Common motifs among high-scoring NNs suggest functional interactions. Here, all subnetworks agree on nodes a, b, and e (*bold*).

as clustering, principal component analysis, and partial least squares regression can derive useful biological insights from large data sets [71] but do not incorporate information about network topology. Alternatively, artificial neural networks (NNs) offer a flexible framework for modeling cellular signal networks and contain both topological and time-dependent features. We recently trained a recurrent NN model to learn the platelet calcium response to all pair-wise combinations of 6 platelet agonists at 3 doses (Figure 6.4A-B).

To determine whether a NN model can predict cytokine-induced apoptotic response and identify important functional interactions within the signaling network, we can train a collection of recurrent NN models to predict existing time-course measurements [72] of 19 apoptotic proteins in response to 3 molecular stimuli (tumor necrosis factor, epidermal growth factor, and insulin). By generating a large number of random network topologies, we will search for common features that appear in most of the high-scoring network structures (Figure 6.4C). A similar technique, Bayesian network inference [122], was used to reconstruct causal influences in human primary T cell signaling [136]. Our choice of using a recurrent NN, however, is novel and offers two important advantages over Bayesian methods in that it allows cyclical structures (e.g., negative feedback) and time-dependence in the model. This aim will provide a time-dependent model of cell survival responses (e.g., nuclear fragmentation) to 3 cytokines and identify gross features of the network that processes these molecular signals. To quantify the network response to perturbation, the interaction of pharmacological agents with the NN will be modeled by performing a sensitivity analysis on the “consensus” NN structures identified in the training step. For example, by eliminating a single node or edge in the apoptotic network, one can examine the network’s survival response and

thereby quantify the impact of each perturbation to the NN. This analysis will identify specific components of the network that may be targeted for rational drug design.

6.2.4 *Future Work: Predicting global regulators of transcription in “oncogene-addicted” cell lines*

Mutations that affect cell signaling can cause differences in basal gene expression [14]. For example, mutations in the tyrosine kinase domain of EGFR (e.g., EGFR^{L834R}) can cause distinct patterns of gene expression [26] and are associated with sensitivity to tyrosine kinase inhibitors (TKIs) [35]. Statistical methods such as principal component analysis have been used to identify disrupted signaling pathways based solely on expression signatures [173]. While this approach is useful for placing mutant cells into previously known categories, there are currently no methods for predicting the signaling mechanisms that regulate transcription based on entirely new gene expression profiles. Such a method could be used to identify the upstream events that cause these differential expression signatures and predict novel drug targets.

Here, statistical and graph-theoretical methods can be used to predict upstream signaling events that cause changes in global gene expression for TKI-sensitive cell lines (e.g., H3255 with EGFR^{L834R}). To identify programs of transcription factor (TF) regulation, we can extend a simple approach [167] in which the abundances of known binding motifs in the upstream sequence of differentially expressed genes are used to determine a probable set of transcription factors responsible for generating the observed expression profile (Figure 6.5A). Focusing on a set of TKI-sensitive cell lines [35], we will incorporate a confidence score for each differentially expressed gene to weight the estimation of enriched TF binding sites. This Bayesian approach will afford a more powerful prediction of TF activation by incorporating all of the information obtained from microarray analysis. Specifically, we will calculate for each TF the probability that it is overrepresented in the data set:

$$Prob(H_i|D) = \frac{(\pi_i Prob(D|H_i))}{\sum_j \pi_j Prob(D|H_j)} \quad (6.1)$$

The probability that TF_{*i*} binding sites are overrepresented (H_i) in the data is weighted by π_i , a confidence value for each gene that contains binding sites for TF_{*i*}. Predictions are validated by TF protein arrays (Figure 6.5B). This aim will provide a ranked list of TFs that are similarly regulated in a pool of TKI-sensitive cell lines.

To predict upstream transcriptional regulators, publically available databases of protein-protein interactions (e.g., IntAct [84]) will be used to construct an *in silico* signaling network (a directed graph) that also includes the TFs identified in the previous step. Subnetwork structures within the

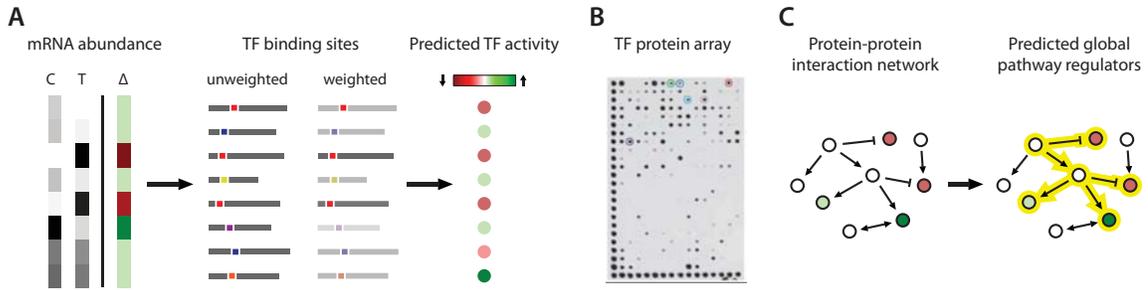


Figure 6.5: Information flow for predicting transcription regulatory networks. **(A)** Confidence scores for measured differences in gene expression (Δ) are used to calculate the abundance of TF binding sites in the DNA sequence surrounding each gene. **(B)** TF protein arrays [20] are used to validate predictions. **(C)** Relative TF activities are used to search protein-protein interaction networks for upstream signaling pathways that cause the predicted changes in TF activity.

full graph are identified that most likely explain the calculated changes TF activities (Figure 6.5C). Beginning with the TF nodes, a recursive algorithm performs a breadth-first search along neighboring edges and assigns a score to each node based on the consistency between the known interaction (activating or inhibiting) and the TF activity. Predictions are validated by RNAi [20].

6.3 Future Outlook

Systems biology—the discipline of integrating complex data sets to understand and discover the molecular basis of emergent properties of biological systems—is nearly a decade old but still in a nascent state. To truly advance our understanding of complex cellular systems and derive high-impact clinical applications using this approach, the field requires a new suite of computational tools and creative experimental strategies. The single greatest asset of all computational methods is the ability to perform simple arithmetic in a very short period of time. Thus, to the extent that signaling networks can be represented as a series of elementary calculations that describe the dynamics of the system (e.g., ODEs that describe catalytic steps in an enzyme pathway), numerical methods offer a technological benefit that is not directly accessible to the unaided human brain. The grand challenge for systems biologists, however, will be to more clearly define which problems to solve, how they may be cast in mathematical terms, and—most importantly—how to interpret the solutions. The work presented in this dissertation illustrates how novel computational strategies can be used to derive meaningful insight from two mammalian signaling systems with important clinical implications. At the very least, these analyses may serve as a starting point for future investigation of signaling systems from a computational systems perspective.

Bibliography

- [1] J. G. Albeck, G. MacBeath, F. M. White, P. K. Sorger, D. A. Lauffenburger, and S. Gaudet. Collecting and organizing systematic sets of protein data. *Nat Rev Mol Cell Biol*, 7(11):803–12, 2006.
- [2] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. *Molecular Biology of the Cell*. Garland Publishing, Inc., New York, 1994.
- [3] J. Amann, S. Kalyankrishna, P. P. Massion, J. E. Ohm, L. Girard, H. Shigematsu, M. Peyton, D. Juroske, Y. Huang, J. Stuart Salmon, Y. H. Kim, J. R. Pollack, K. Yanagisawa, A. Gazdar, J. D. Minna, J. M. Kurie, and D. P. Carbone. Aberrant epidermal growth factor receptor signaling and enhanced sensitivity to egfr inhibitors in lung cancer. *Cancer Res*, 65(1):226–35, 2005. CA084971/CA/NCI NIH HHS/United States P50CA070907/CA/NCI NIH HHS/United States P50CA90949/CA/NCI NIH HHS/United States R01CA105155/CA/NCI NIH HHS/United States Journal Article Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, P.H.S. United States.
- [4] S. S. Andrews and A. P. Arkin. Simulating cell biology. *Curr Biol*, 16(14):R523–7, 2006.
- [5] C. L. Arteaga. Epidermal growth factor receptor dependence in human tumors: more than just expression? *Oncologist*, 7 Suppl 4:31–9, 2002.
- [6] C. L. Arteaga and C. I. Truica. Challenges in the development of anti-epidermal growth factor receptor therapies in breast cancer. *Semin Oncol*, 31(1 Suppl 3):3–8, 2004.
- [7] A. R. Asthagiri and D. A. Lauffenburger. Bioengineering models of cell signaling. *Annu Rev Biomed Eng*, 2:31–53, 2000.
- [8] I. Aursnes, K. Gjesdal, and U. Abildgaard. Platelet aggregation induced by adp from un-sheared erythrocytes at physiological ca⁺⁺-concentration. *Br J Haematol*, 47(1):149–52, 1981.

- [9] N. S. Baliga. Systems biology. the scale of prediction. *Science*, 320(5881):1297–8, 2008. Comment Journal Article United States.
- [10] J. Barthelme, C. Ebeling, A. Chang, I. Schomburg, and D. Schomburg. Brenda, amenda and frenda: the enzyme information system in 2007. *Nucleic Acids Res*, 35(Database issue):D511–4, 2007.
- [11] A. Baurand, A. Eckly, B. Hechler, G. Kauffenstein, J. L. Galzi, J. P. Cazenave, C. Leon, and C. Gachet. Differential regulation and relocalization of the platelet p2y receptors after activation: a way to avoid loss of hemostatic properties? *Mol Pharmacol*, 67(3):721–33, 2005.
- [12] A. Baurand, P. Raboisson, M. Freund, C. Leon, J. P. Cazenave, J. J. Bourguignon, and C. Gachet. Inhibition of platelet function by administration of mrs2179, a p2y1 receptor antagonist. *Eur J Pharmacol*, 412(3):213–21, 2001.
- [13] U. S. Bhalla and R. Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283(5400):381–7, 1999. GM-54508/GM/United States NIGMS Journal Article Research Support, U.S. Gov’t, P.H.S. United states.
- [14] A. H. Bild, G. Yao, J. T. Chang, Q. Wang, A. Potti, D. Chasse, M. B. Joshi, D. Harpole, J. M. Lancaster, A. Berchuck, Jr. Olson, J. A., J. R. Marks, H. K. Dressman, M. West, and J. R. Nevins. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature*, 439(7074):353–7, 2006. Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov’t England.
- [15] S. J. Bornheimer, M. R. Maurya, M. G. Farquhar, and S. Subramaniam. Computational modeling reveals how interplay between components of a gtpase-cycle module regulates signal transduction. *Proc Natl Acad Sci U S A*, 101(45):15899–904, 2004.
- [16] L. F. Brass and S. K. Joseph. A role for inositol triphosphate in intracellular ca²⁺ mobilization and granule secretion in platelets. *J Biol Chem*, 260(28):15172–9, 1985.
- [17] L. F. Brass, T. J. Stalker, L. Zhu, and D. S. Woulfe. Signal transduction during platelet plug formation. In A. D. Michelson, editor, *Platelets*, pages 319–346. Elsevier Science, second edition, 2007.
- [18] A. Braun, D. Varga-Szabo, C. Kleinschnitz, I. Pleines, M. Bender, M. Austinat, M. Bosl, G. Stoll, and B. Nieswandt. Orai1 (cracm1) is the platelet soc channel and essential for pathological thrombus formation. *Blood*, 2008. Journal article.

- [19] P. S. Brignola, K. Lackey, S. H. Kadwell, C. Hoffman, E. Horne, H. L. Carter, J. D. Stuart, K. Blackburn, M. B. Moyer, K. J. Alligood, W. B. Knight, and E. R. Wood. Comparison of the biochemical and kinetic properties of the type 1 receptor tyrosine kinase intracellular domains. demonstration of differential sensitivity to kinase inhibitors. *J Biol Chem*, 277(2):1576–85, 2002.
- [20] K. D. Bromberg, A. Ma'ayan, S. R. Neves, and R. Iyengar. Design logic of a cannabinoid receptor signaling network that triggers neurite outgrowth. *Science*, 320(5878):903–9, 2008. 1 S10 RR0 9145-01/RR/United States NCCR 5R24 CA095823-04/CA/United States NCI GM072853/GM/United States NIGMS GM54508/GM/United States NIGMS P50-071558/United States PHS T32 CA88796/CA/United States NCI Journal Article Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S. United States.
- [21] K. S. Brown, C. C. Hill, G. A. Calero, C. R. Myers, K. H. Lee, J. P. Sethna, and R. A. Cerione. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys Biol*, 1(3-4):184–95, 2004.
- [22] A. W. Burgess, H. S. Cho, C. Eigenbrot, K. M. Ferguson, T. P. Garrett, D. J. Leahy, M. A. Lemmon, M. X. Sliwkowski, C. W. Ward, and S. Yokoyama. An open-and-shut case? recent insights into the activation of egf/erbB receptors. *Mol Cell*, 12(3):541–52, 2003.
- [23] K. D. Carey, A. J. Garton, M. S. Romero, J. Kahler, S. Thomson, S. Ross, F. Park, J. D. Haley, N. Gibson, and M. X. Sliwkowski. Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. *Cancer Res*, 66(16):8163–71, 2006.
- [24] C. Chaabane, S. Dally, E. Corvazier, R. Bredoux, R. Bobe, B. Ftouhi, A. Raies, and J. Enouf. Platelet pmca- and serca-type ca²⁺-atpase expression in diabetes: a novel signature of abnormal megakaryocytopoiesis. *J Thromb Haemost*, 5(10):2127–35, 2007. Journal Article England Jth.
- [25] Y. R. Chen, Y. N. Fu, C. H. Lin, S. T. Yang, S. F. Hu, Y. T. Chen, S. F. Tsai, and S. F. Huang. Distinctive activation patterns in constitutively active and gefitinib-sensitive egfr mutants. *Oncogene*, 25(8):1205–15, 2006.
- [26] S. H. Choi, J. M. Mendrola, and M. A. Lemmon. EGF-independent activation of cell-surface egf receptors harboring mutations found in gefitinib-sensitive lung cancer. *Oncogene*, 26(11):1567–76, 2007.

- [27] T. Y. Chou, C. H. Chiu, L. H. Li, C. Y. Hsiao, C. Y. Tzen, K. T. Chang, Y. M. Chen, R. P. Perng, S. F. Tsai, and C. M. Tsai. Mutation in the tyrosine kinase domain of epidermal growth factor receptor is a predictive and prognostic factor for gefitinib treatment in patients with non-small cell lung cancer. *Clin Cancer Res*, 11(10):3750–7, 2005.
- [28] F. Ciardiello. Epidermal growth factor receptor tyrosine kinase inhibitors as anticancer agents. *Drugs*, 60 Suppl 1:25–32; discussion 41–2, 2000.
- [29] F. Ciardiello, R. Caputo, R. Bianco, V. Damiano, G. Fontanini, S. Cuccato, S. De Placido, A. R. Bianco, and G. Tortora. Inhibition of growth factor production and angiogenesis in human cancer cells by zd1839 (iressa), a selective epidermal growth factor receptor tyrosine kinase inhibitor. *Clin Cancer Res*, 7(5):1459–65, 2001.
- [30] F. Ciardiello, R. Caputo, G. Borriello, D. Del Bufalo, A. Biroccio, G. Zupi, A. R. Bianco, and G. Tortora. Zd1839 (iressa), an egfr-selective tyrosine kinase inhibitor, enhances taxane activity in bcl-2 overexpressing, multidrug-resistant mcf-7 adr human breast cancer cells. *Int J Cancer*, 98(3):463–9, 2002.
- [31] F. Ciardiello, F. De Vita, M. Orditura, and G. Tortora. The role of egfr inhibitors in nonsmall cell lung cancer. *Curr Opin Oncol*, 16(2):130–5, 2004. Journal Article Research Support, Non-U.S. Gov’t Review United States.
- [32] F. Ciardiello and G. Tortora. A novel approach in the treatment of cancer: targeting the epidermal growth factor receptor. *Clin Cancer Res*, 7(10):2958–70, 2001.
- [33] A. Citri and Y. Yarden. Egf-erbB signalling: towards the systems level. *Nat Rev Mol Cell Biol*, 7(7):505–16, 2006. Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov’t Review England.
- [34] S. D. Cohen and A. C. Hindmarsh. Cvode, a stiff/nonstiff ode solver in c. *Computers in Physics*, 10(2):138–143, 1996.
- [35] C. D. Coldren, B. A. Helfrich, S. E. Witt, M. Sugita, R. Lapadat, C. Zeng, A. Baron, W. A. Franklin, F. R. Hirsch, M. W. Geraci, and Jr. Bunn, P. A. Baseline gene expression predicts sensitivity to gefitinib in non-small cell lung cancer cell lines. *Mol Cancer Res*, 4(8):521–8, 2006. P30 CA046934/CA/United States NCI P50 CA058187/CA/United States NCI P50CA70907/CA/United States NCI Comparative Study Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov’t United States Mcr.

- [36] B. S. Coller. A brief history of ideas about platelets in health and disease. In A. D. Michelson, editor, *Platelets*, pages xxiii–xlii. Elsevier Science, second edition, 2007.
- [37] T. M. Connolly, T. E. Bross, and P. W. Majerus. Isolation of a phosphomonoesterase from human platelets that specifically hydrolyzes the 5-phosphate of inositol 1,4,5-trisphosphate. *J Biol Chem*, 260(13):7868–74, 1985.
- [38] E. D. Conrad and J. J. Tyson. Modeling molecular interaction networks with nonlinear ordinary differential equations. In Z. Szallasi, J. Stelling, and V. Periwal, editors, *System Modeling in Cellular Biology*, pages 97–123. The MIT Press, Cambridge, 2006.
- [39] P. A. Corning. The re-emergence of "emergence": A venerable concept in search of a theory. *Complexity*, 7(6):18–30, 2002.
- [40] J. Couzin. Cancer drugs. smart weapons prove tough to design. *Science*, 298(5593):522–5, 2002. News United States.
- [41] J. Dancey. Epidermal growth factor receptor inhibitors in clinical development. *Int J Radiat Oncol Biol Phys*, 58(3):1003–7, 2004.
- [42] J. E. Dancey. Predictive factors for epidermal growth factor receptor inhibitors—the bull’s-eye hits the arrow. *Cancer Cell*, 5(5):411–5, 2004. Journal Article Review United States.
- [43] J. L. Daniel, C. Dangelmaier, J. Jin, B. Ashby, J. B. Smith, and S. P. Kunapuli. Molecular basis for adp-induced platelet activation. i. evidence for three distinct adp receptors on human platelets. *J Biol Chem*, 273(4):2024–9, 1998.
- [44] K. A. Dill. *Molecular Driving Forces*. Garland Science, New York, 2003.
- [45] L. Dode, B. Vilsen, K. Van Baelen, F. Wuytack, J. D. Clausen, and J. P. Andersen. Dissection of the functional differences between sarco(endo)plasmic reticulum ca^{2+} -atpase (serca) 1 and 3 isoforms by steady-state and transient kinetic analyses. *J Biol Chem*, 277(47):45579–91, 2002.
- [46] L. Ebbeling, C. Robertson, A. McNicol, and J. M. Gerrard. Rapid ultrastructural changes in the dense tubular system following platelet activation. *Blood*, 80(3):718–23, 1992.
- [47] P. Erne, P. Bolli, E. Burgisser, and F. R. Buhler. Correlation of platelet calcium with blood pressure. effect of antihypertensive therapy. *N Engl J Med*, 310(17):1084–8, 1984.

- [48] M. A. Fabian, 3rd Biggs, W. H., D. K. Treiber, C. E. Atteridge, M. D. Azimioara, M. G. Benedetti, T. A. Carter, P. Ciceri, P. T. Edeen, M. Floyd, J. M. Ford, M. Galvin, J. L. Gerlach, R. M. Grotzfeld, S. Herrgard, D. E. Insko, M. A. Insko, A. G. Lai, J. M. Lelias, S. A. Mehta, Z. V. Milanov, A. M. Velasco, L. M. Wodicka, H. K. Patel, P. P. Zarrinkar, and D. J. Lockhart. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotechnol*, 23(3):329–36, 2005. Evaluation Studies Journal Article United States.
- [49] I. Famili, R. Mahadevan, and B. O. Palsson. k-cone analysis: determining all candidate values for kinetic parameters on a network scale. *Biophys J*, 88(3):1616–25, 2005. R01 GM68837/GM/United States NIGMS Comparative Study Evaluation Studies Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S. Validation Studies United States.
- [50] I. Famili and B. O. Palsson. Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices. *J Theor Biol*, 224(1):87–96, 2003. GM57089/GM/United States NIGMS Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S. England.
- [51] Y. X. Fan, L. Wong, T. B. Deb, and G. R. Johnson. Ligand regulates epidermal growth factor receptor kinase specificity: activation increases preference for gab1 and shc versus autophosphorylation sites. *J Biol Chem*, 279(37):38143–50, 2004.
- [52] K. Farsad, N. Ringstad, K. Takei, S. R. Floyd, K. Rose, and P. De Camilli. Generation of high curvature membranes mediated by direct endophilin bilayer interactions. *J Cell Biol*, 155(2):193–200, 2001.
- [53] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. O. Palsson. A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Mol Syst Biol*, 3:121, 2007. GM057089/GM/United States NIGMS R01 GM057089/GM/United States NIGMS Comparative Study Journal Article Research Support, N.I.H., Extramural England.
- [54] K. M. Ferguson, M. B. Berger, J. M. Mendrola, H. S. Cho, D. J. Leahy, and M. A. Lemmon. Egf activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol Cell*, 11(2):507–17, 2003.

- [55] T. M. Filtz, M. L. Cunningham, K. J. Stanig, A. Paterson, and T. K. Harden. Phosphorylation by protein kinase c decreases catalytic activity of avian phospholipase c-beta. *Biochem J*, 338 (Pt 2):257–64, 1999.
- [56] M. G. Ford, I. G. Mills, B. J. Peter, Y. Vallis, G. J. Praefcke, P. R. Evans, and H. T. McMahon. Curvature of clathrin-coated pits driven by epsin. *Nature*, 419(6905):361–6, 2002.
- [57] B. Furie and B. C. Furie. Mechanisms of thrombus formation. *N Engl J Med*, 359(9):938–49, 2008. Journal Article Review United States.
- [58] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi. *GNU Scientific Library Reference Manual*. Network Theory Limited, Bristol, second edition, 2006.
- [59] D. T. Gillespie. Exact stochastic simulation of coupled chemical-reactions. *Abstracts of Papers of the American Chemical Society*, 173(MAR20):128–128, 1977.
- [60] N. Gotoh, A. Tojo, M. Hino, Y. Yazaki, and M. Shibuya. A highly conserved tyrosine residue at codon 845 within the kinase domain is not required for the transforming activity of human epidermal growth factor receptor. *Biochem Biophys Res Commun*, 186(2):768–74, 1992.
- [61] J.D. Griffin and T. G Kolda. Asynchronous parallel generating set search for linearly-constrained optimization. Technical report, Sandia National Laboratories, Albuquerque, NM and Livermore, CA, July 2006.
- [62] E. Guadagnoli and W. F. Velicer. Relation of sample size to the stability of component patterns. *Psychol Bull*, 103(2):265–75, 1988.
- [63] J. W. Heemskerk, J. Hoyland, W. T. Mason, and S. O. Sage. Spiking in cytosolic calcium concentration in single fibrinogen-bound fura-2-loaded human platelets. *Biochem J*, 283 (Pt 2):379–83, 1992.
- [64] J. W. Heemskerk, G. M. Willems, M. B. Rook, and S. O. Sage. Ragged spiking of free calcium in adp-stimulated human platelets: regulation of puff-like calcium signals in vitro and ex vivo. *J Physiol*, 535(Pt 3):625–35, 2001.
- [65] B. S. Hendriks, G. J. Griffiths, R. Benson, D. Kenyon, M. Lazzara, J. Swinton, S. Beck, M. Hickinson, J. M. Beusmans, D. Lauffenburger, and D. de Graaf. Decreased internalisation of erbb1 mutants in lung cancer is linked with a mechanism conferring sensitivity to gefitinib. *Syst Biol (Stevenage)*, 153(6):457–66, 2006. Journal Article Research Support, Non-U.S. Gov't England.

- [66] Matthew F. Hockin, Kenneth C. Jones, Stephen J. Everse, and Kenneth G. Mann. A model for the stoichiometric regulation of blood coagulation. *J. Biol. Chem.*, 277(21):18322–18333, 2002.
- [67] S. R. Hubbard and J. H. Till. Protein tyrosine kinase structure and function. *Annu Rev Biochem*, 69:373–98, 2000.
- [68] B. P. Ingalls, T. Yi, and P. A. Iglesias. Using control theory to study biology. In Z. Szallasi, J. Stelling, and V. Periwal, editors, *System Modeling in Cellular Biology*, pages 243–267. The MIT Press, Cambridge, 2006.
- [69] J. H. Italiano, J. E. Hartwig. Megakaryocyte development and platelet formation. In A. D. Michelson, editor, *Platelets*, pages 23–44. Elsevier Science, second edition, 2007.
- [70] N. Jamshidi and B. O. Palsson. Formulating genome-scale kinetic models in the post-genome era. *Mol Syst Biol*, 4:171, 2008. Journal Article England.
- [71] K. A. Janes, J. G. Albeck, S. Gaudet, P. K. Sorger, D. A. Lauffenburger, and M. B. Yaffe. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science*, 310(5754):1646–53, 2005.
- [72] K. A. Janes, S. Gaudet, J. G. Albeck, U. B. Nielsen, D. A. Lauffenburger, and P. K. Sorger. The response of human epithelial cells to tnf involves an inducible autocrine cascade. *Cell*, 124(6):1225–39, 2006. P50-GM68762/GM/NIGMS NIH HHS/United States Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov’t United States.
- [73] K. A. Janes and D. A. Lauffenburger. A biological approach to computational models of proteomic networks. *Curr Opin Chem Biol*, 10(1):73–80, 2006.
- [74] K. A. Janes and M. B. Yaffe. Data-driven modelling of signal-transduction networks. *Nat Rev Mol Cell Biol*, 7(11):820–8, 2006.
- [75] M. L. Janmaat and G. Giaccone. Small-molecule epidermal growth factor receptor tyrosine kinase inhibitors. *Oncologist*, 8(6):576–86, 2003. Historical Article Journal Article Review United States.
- [76] K. Jaqaman and G. Danuser. Linking data to models: data regression. *Nat Rev Mol Cell Biol*, 7(11):813–9, 2006.
- [77] L. M. Jarvis. Battling breast cancer. *Chem. Eng. News*, 84(32):21–27, 2006.

- [78] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, and M. J. Thun. Cancer statistics, 2008. *CA Cancer J Clin*, 58(2):71–96, 2008. Journal Article United States.
- [79] M. Jirouskova, A. S. Shet, and G. J. Johnson. A guide to murine platelet structure, function, assays, and genetic alterations. *J Thromb Haemost*, 5(4):661–9, 2007.
- [80] R. N. Jorissen, F. Walker, N. Pouliot, T. P. Garrett, C. W. Ward, and A. W. Burgess. Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp Cell Res*, 284(1):31–53, 2003.
- [81] A. R. Joyce and B. O. Palsson. The model organism as a system: integrating ‘omics’ data sets. *Nat Rev Mol Cell Biol*, 7(3):198–210, 2006. Journal Article Review England.
- [82] B. N. Kahner, H. Shankar, S. Murugappan, G. L. Prasad, and S. P. Kunapuli. Nucleotide receptor signaling in platelets. *J Thromb Haemost*, 4(11):2317–26, 2006.
- [83] P. D. Karp, M. Riley, M. Saier, I. T. Paulsen, J. Collado-Vides, S. M. Paley, A. Pellegrini-Toole, C. Bonavides, and S. Gama-Castro. The ecocyc database. *Nucleic Acids Res*, 30(1):56–8, 2002. 1-R01-RR07861-01/RR/United States NCRR Journal Article Research Support, U.S. Gov’t, P.H.S. England.
- [84] S. Kerrien, Y. Alam-Faruque, B. Aranda, I. Bancarz, A. Bridge, C. Derow, E. Dimmer, M. Feuermann, A. Friedrichsen, R. Huntley, C. Kohler, J. Khadake, C. Leroy, A. Liban, C. Lieftink, L. Montecchi-Palazzi, S. Orchard, J. Risse, K. Robbe, B. Roechert, D. Thorncroft, Y. Zhang, R. Apweiler, and H. Hermjakob. Intact–open source resource for molecular interaction data. *Nucleic Acids Res*, 35(Database issue):D561–5, 2007. Journal Article Research Support, Non-U.S. Gov’t England.
- [85] B. N. Kholodenko. Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol*, 7(3):165–76, 2006. Journal Article Research Support, N.I.H., Extramural Review England.
- [86] B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek. Quantification of short term signaling by the epidermal growth factor receptor. *J Biol Chem*, 274(42):30169–81, 1999.
- [87] T. L. Kinzer-Ursem and J. J. Linderman. Both ligand- and cell-specific parameters control ligand agonism in a kinetic model of g protein-coupled receptor signaling. *PLoS Comput Biol*, 3(1):e6, 2007.
- [88] S. Klamt and J. Stelling. Stoichiometric and constraint-based modeling. In Z. Szallasi, J. Stelling, and V. Periwal, editors, *System Modeling in Cellular Biology*. The MIT Press, 2006.

- [89] S. Kobayashi, T. J. Boggon, T. Dayaram, P. A. Janne, O. Kocher, M. Meyerson, B. E. Johnson, M. J. Eck, D. G. Tenen, and B. Halmos. Egfr mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med*, 352(8):786–92, 2005.
- [90] G. Krauss. *Biochemistry of Signal Transduction and Regulation*. Wiley-VCH, Weinheim, 3rd edition, 2003.
- [91] A. L. Kuharsky and A. L. Fogelson. Surface-mediated control of blood coagulation: the role of binding site densities and platelet deposition. *Biophys J*, 80(3):1050–74, 2001.
- [92] N. Kumar, B. S. Hendriks, K. A. Janes, D. de Graaf, and D. A. Lauffenburger. Applying computational modeling to drug discovery and development. *Drug Discov Today*, 11(17-18):806–11, 2006. Journal Article Research Support, Non-U.S. Gov’t Review England.
- [93] C. Lacabaratz-Porret, S. Launay, E. Corvazier, R. Bredoux, B. Papp, and J. Enouf. Biogenesis of endoplasmic reticulum proteins involved in ca^{2+} signalling during megakaryocytic differentiation: an in vitro study. *Biochem J*, 350 Pt 3:723–34, 2000.
- [94] G. Lahav. Oscillations by the p53-mdm2 feedback loop. *Adv Exp Med Biol*, 641:28–38, 2008. Journal Article Research Support, Non-U.S. Gov’t Review United States.
- [95] M. J. Lazzara and D. A. Lauffenburger. Quantitative modeling perspectives on the erbb system of cell regulatory processes. *Exp Cell Res*, 315(4):717–25, 2009. R01-CA096504/CA/NCI NIH HHS/United States U54-CA112967/CA/NCI NIH HHS/United States Journal Article Research Support, N.I.H., Extramural United States.
- [96] J. A. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction*. Springer, New York, 2007.
- [97] J. W. Lee, Y. H. Soung, S. Y. Kim, W. S. Park, S. W. Nam, S. H. Kim, J. Y. Lee, N. J. Yoo, and S. H. Lee. Erbb2 kinase domain mutation in a gastric cancer metastasis. *Apmis*, 113(10):683–7, 2005.
- [98] J. W. Lee, Y. H. Soung, S. H. Seo, S. Y. Kim, C. H. Park, Y. P. Wang, K. Park, S. W. Nam, W. S. Park, S. H. Kim, J. Y. Lee, N. J. Yoo, and S. H. Lee. Somatic mutations of erbb2 kinase domain in gastric, colorectal, and breast carcinomas. *Clin Cancer Res*, 12(1):57–61, 2006.
- [99] R. S. Lewis. The molecular choreography of a store-operated calcium channel. *Nature*, 446(7133):284–7, 2007. Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov’t Review England.

- [100] J. Lin, C. M. Gan, X. Zhang, S. Jones, T. Sjoblom, L. D. Wood, D. W. Parsons, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, G. Parmigiani, and V. E. Velculescu. A multidimensional analysis of genes mutated in breast and colorectal cancers. *Genome Res*, 17(9):1304–18, 2007. CA 121113/CA/NCI NIH HHS/United States CA 43460/CA/NCI NIH HHS/United States CA 57345/CA/NCI NIH HHS/United States CA105090-03/CA/NCI NIH HHS/United States CA62924/CA/NCI NIH HHS/United States HHSN261200433002C/PHS HHS/United States Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. United States.
- [101] J. J. Linderman. Modeling of g-protein coupled receptor signaling pathways. *J Biol Chem*, 2008. Journal article.
- [102] Y. Liu, J. Purvis, A. Shih, J. Weinstein, N. Agrawal, and R. Radhakrishnan. A multiscale computational approach to dissect early events in the erb family receptor mediated activation, differential signaling, and relevance to oncogenic transformations. *Ann Biomed Eng*, 35(6):1012–25, 2007. Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't United States.
- [103] D. Lloyd-Jones, R. Adams, M. Carnethon, G. De Simone, T. B. Ferguson, K. Flegal, E. Ford, K. Furie, A. Go, K. Greenlund, N. Haase, S. Hailpern, M. Ho, V. Howard, B. Kissela, S. Kittner, D. Lackland, L. Lisabeth, A. Marelli, M. McDermott, J. Meigs, D. Mozaffarian, G. Nichol, C. O'Donnell, V. Roger, W. Rosamond, R. Sacco, P. Sorlie, R. Stafford, J. Steinberger, T. Thom, S. Wasserthiel-Smoller, N. Wong, J. Wylie-Rosett, and Y. Hong. Heart disease and stroke statistics–2009 update: a report from the american heart association statistics committee and stroke statistics subcommittee. *Circulation*, 119(3):e21–181, 2009. American Heart Association Statistics Committee and Stroke Statistics Subcommittee Journal Article United States.
- [104] J. J. Lopez, P. C. Redondo, G. M. Salido, J. A. Pariente, and J. A. Rosado. Two distinct ca^{2+} compartments show differential sensitivity to thrombin, adp and vasopressin in human platelets. *Cell Signal*, 18(3):373–81, 2006.
- [105] D. Luan, M. Zai, and J. D. Varner. Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies. *PLoS Comput Biol*, 3(7):e142, 2007.

- [106] T. J. Lynch, D. W. Bell, R. Sordella, S. Gurubhagavatula, R. A. Okimoto, B. W. Brannigan, P. L. Harris, S. M. Haserlat, J. G. Supko, F. G. Haluska, D. N. Louis, D. C. Christiani, J. Settleman, and D. A. Haber. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med*, 350(21):2129–39, 2004. P01 95281/PHS HHS/United States P30 CA0516/CA/NCI NIH HHS/United States P50 CA 090578/CA/NCI NIH HHS/United States R01 CA 092824/CA/NCI NIH HHS/United States Journal Article Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, P.H.S. United States.
- [107] K. M. Marks and G. P. Nolan. Chemical labeling strategies for cell biology. *Nat Methods*, 3(8):591–6, 2006.
- [108] L. Matthews, G. Gopinath, M. Gillespie, M. Caudy, D. Croft, B. de Bono, P. Garapati, J. Hemish, H. Hermjakob, B. Jassal, A. Kanapin, S. Lewis, S. Mahajan, B. May, E. Schmidt, I. Vastrik, G. Wu, E. Birney, L. Stein, and P. D’Eustachio. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res*, 2008. Journal article.
- [109] J. Mendelsohn and J. Baselga. The egf receptor family as targets for cancer therapy. *Oncogene*, 19(56):6550–65, 2000.
- [110] A. S. Mildvan. Mechanisms of signaling and related enzymes. *Proteins*, 29(4):401–16, 1997. DK28616/DK/NIDDK NIH HHS/United States Journal Article Research Support, U.S. Gov’t, P.H.S. Review United states.
- [111] C. G. Moles, P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res*, 13(11):2467–74, 2003.
- [112] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14):1639–1662, 1998.
- [113] S. L. Moulder, F. M. Yakes, S. K. Muthuswamy, R. Bianco, J. F. Simpson, and C. L. Arteaga. Epidermal growth factor receptor (her1) tyrosine kinase inhibitor zd1839 (iressa) inhibits her2/neu (erbb2)-overexpressing breast cancer cells in vitro and in vivo. *Cancer Res*, 61(24):8887–95, 2001.
- [114] S. Mukhopadhyay and E. M. Ross. Rapid gtp binding and hydrolysis by g(q) promoted by receptor and gtpase-activating proteins. *Proc Natl Acad Sci U S A*, 96(17):9539–44, 1999.

- [115] B. Nolen, S. Taylor, and G. Ghosh. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell*, 15(5):661–75, 2004.
- [116] R. Noro, A. Gemma, S. Kosaihira, Y. Kokubo, M. Chen, M. Seike, K. Kataoka, K. Matsuda, T. Okano, Y. Minegishi, A. Yoshimura, and S. Kudoh. Gefitinib (iressa) sensitive lung cancer cell lines show phosphorylation of akt without ligand stimulation. *BMC Cancer*, 6:277, 2006. Comparative Study Journal Article England.
- [117] A. T. Nurden and P. Nurden. Inherited disorders of platelets: an update. *Curr Opin Hematol*, 13(3):157–62, 2006.
- [118] P. Nurden and A. T. Nurden. Congenital disorders associated with platelet dysfunctions. *Thromb Haemost*, 99(2):253–63, 2008.
- [119] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol*, 1:2005 0010, 2005. Journal Article Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, Non-P.H.S. England.
- [120] J. G. Paez, P. A. Janne, J. C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F. J. Kaye, N. Lindeman, T. J. Boggon, K. Naoki, H. Sasaki, Y. Fujii, M. J. Eck, W. R. Sellers, B. E. Johnson, and M. Meyerson. Egfr mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, 304(5676):1497–500, 2004. Journal Article Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, P.H.S. United States.
- [121] W. Pao and V. A. Miller. Epidermal growth factor receptor mutations, small-molecule kinase inhibitors, and non-small-cell lung cancer: current knowledge and future directions. *J Clin Oncol*, 23(11):2556–68, 2005. Journal Article Research Support, Non-U.S. Gov’t Review United States official journal of the American Society of Clinical Oncology.
- [122] D. Pe’er. Bayesian network analysis of signaling networks: a primer. *Sci STKE*, 2005(281):pl4, 2005. Journal Article Review United States signal transduction knowledge environment.
- [123] A. Petrelli, G. F. Gilestro, S. Lanzardo, P. M. Comoglio, N. Migone, and S. Giordano. The endophilin-cin85-cbl complex mediates ligand-dependent downregulation of c-met. *Nature*, 416(6877):187–90, 2002.
- [124] V. Petrov, E. Nikolova, and O. Wolkenhauer. Reduction of nonlinear dynamic systems with an application to signal transduction pathways. *IET Syst Biol*, 1(1):2–9, 2007.

- [125] John Robinson Pierce. *An introduction to information theory : symbols, signals & noise*. Dover Publications, New York, 2nd, rev. edition, 1980. 80066678 John R. Pierce. ill. ; 22 cm. Previously published as: Symbols, signals, and noise. 1961. Includes bibliographical references and index.
- [126] J. Preiss, C. R. Loomis, W. R. Bishop, R. Stein, J. E. Niedel, and R. M. Bell. Quantitative measurement of sn-1,2-diacylglycerols present in platelets, hepatocytes, and ras- and sis-transformed normal rat kidney cells. *J Biol Chem*, 261(19):8597–600, 1986.
- [127] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 3rd edition, 2007.
- [128] J. Purvis, V. Ilango, and R. Radhakrishnan. Role of network branching in eliciting differential short-term signaling responses in the hypersensitive epidermal growth factor receptor mutants implicated in lung cancer. *Biotechnol Prog*, 24(3):540–53, 2008. Journal Article United States.
- [129] J. E. Purvis, M. S. Chatterjee, L. F. Brass, and S. L. Diamond. A molecular signaling model of platelet phosphoinositide and calcium regulation during homeostasis and p2y1 activation. *Blood*, 112(10):4069–79, 2008. R01-HL-56 621/HL/United States NHLBI R33-HL-87 317/HL/United States NHLBI Journal Article Research Support, N.I.H., Extramural United States.
- [130] James W. Putney. *Calcium signaling*. Methods in signal transduction. CRC Press, Boca Raton, Fla, 2000.
- [131] A. K. Rao. Inherited defects in platelet signaling mechanisms. *J Thromb Haemost*, 1(4):671–81, 2003. R01 HL056724/HL/NHLBI NIH HHS/United States Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review England Jth.
- [132] S. E. Rittenhouse and J. P. Sasson. Mass changes in myoinositol trisphosphate in human platelets stimulated by thrombin. inhibitory effects of phorbol ester. *J Biol Chem*, 260(15):8657–60, 1985.
- [133] C. A. Ritter and C. L. Arteaga. The epidermal growth factor receptor-tyrosine kinase: a promising therapeutic target in solid tumors. *Semin Oncol*, 30(1 Suppl 1):3–11, 2003.
- [134] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–6, 2000. Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. United states.

- [135] S. H. Ryu, U. H. Kim, M. I. Wahl, A. B. Brown, G. Carpenter, K. P. Huang, and S. G. Rhee. Feedback regulation of phospholipase c-beta by protein kinase c. *J. Biol. Chem.*, 265(29):17941–17945, 1990.
- [136] Karen Sachs, Omar Perez, Dana Pe'er, Douglas A. Lauffenburger, and Garry P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.
- [137] U. J. Sachs and B. Nieswandt. In vivo thrombus formation in murine models. *Circ Res*, 100(7):979–91, 2007.
- [138] S. O. Sage, E. H. Yamoah, and J. W. Heemskerk. The roles of p(2x1) and p(2t ac) receptors in adp-evoked calcium signalling in human platelets. *Cell Calcium*, 28(2):119–26, 2000.
- [139] B. Sariyar, S. Perk, U. Akman, and A. Hortacsu. Monte carlo sampling and principal component analysis of flux distributions yield topological and modular information on metabolic networks. *J Theor Biol*, 242(2):389–400, 2006. Journal Article Netherlands.
- [140] K. Saso, G. Moehren, K. Higashi, and J. B. Hoek. Differential inhibition of epidermal growth factor signaling pathways in rat hepatocytes by long-term ethanol treatment. *Gastroenterology*, 112(6):2073–88, 1997.
- [141] J. Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 103(2):211–25, 2000.
- [142] J. Schlessinger. Common and distinct elements in cellular signaling via egf and fgf receptors. *Science*, 306(5701):1506–7, 2004.
- [143] S. L. Schmid. Clathrin-coated vesicle formation and protein sorting: an integrated process. *Annu Rev Biochem*, 66:511–48, 1997.
- [144] A. Schmidt, M. Wolde, C. Thiele, W. Fest, H. Kratzin, A. V. Podtelejnikov, W. Witke, W. B. Huttner, and H. D. Soling. Endophilin i mediates synaptic vesicle formation by transfer of arachidonate to lysophosphatidic acid. *Nature*, 401(6749):133–41, 1999.
- [145] H. Schmidt and E. W. Jacobsen. Linear systems approach to analysis of complex dynamic behaviours in biochemical networks. *Syst Biol (Stevenage)*, 1(1):149–58, 2004.
- [146] H. Schmidt and M. Jirstrand. Systems biology toolbox for matlab: a computational platform for research in systems biology. *Bioinformatics*, 22(4):514–5, 2006.

- [147] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Muller. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nat Biotechnol*, 20(4):370–5, 2002. Journal Article United States.
- [148] S. V. Sharma and J. Settleman. Oncogene addiction: setting the stage for molecularly targeted cancer therapy. *Genes Dev*, 21(24):3214–31, 2007. Journal Article Review United States.
- [149] L. K. Shawver, D. Slamon, and A. Ullrich. Smart drugs: tyrosine kinase inhibitors in cancer therapy. *Cancer Cell*, 1(2):117–23, 2002.
- [150] W. Siess. Molecular mechanisms of platelet activation. *Physiol Rev*, 69(1):58–178, 1989.
- [151] E. Sinn, W. Muller, P. Pattengale, I. Tepler, R. Wallace, and P. Leder. Coexpression of mmtv/v-ha-ras and mmtv/c-myc genes in transgenic mice: synergistic action of oncogenes in vivo. *Cell*, 49(4):465–75, 1987. 5 T32 GM07196/GM/NIGMS NIH HHS/United States CA-07968/CA/NCI NIH HHS/United States K12DK01401/DK/NIDDK NIH HHS/United States Journal Article Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, P.H.S. United states.
- [152] S. Sivakumaran, S. Hariharaputran, J. Mishra, and U. S. Bhalla. The database of quantitative cellular signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics*, 19(3):408–15, 2003.
- [153] K. Smallbone, E. Simeonidis, D. S. Broomhead, and D. B. Kell. Something from nothing: bridging the gap between constraint-based and kinetic modelling. *Febs J*, 274(21):5576–85, 2007.
- [154] J. Sneyd and J. F. Dufour. A dynamic model of the type-2 inositol trisphosphate receptor. *Proc Natl Acad Sci U S A*, 99(4):2398–403, 2002.
- [155] J. L. Snoep, F. Bruggeman, B. G. Olivier, and H. V. Westerhoff. Towards building the silicon cell: a modular approach. *Biosystems*, 83(2-3):207–16, 2006.
- [156] Z. Songyang, 3rd Carraway, K. L., M. J. Eck, S. C. Harrison, R. A. Feldman, M. Mohammadi, J. Schlessinger, S. R. Hubbard, D. P. Smith, C. Eng, and et al. Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. *Nature*, 373(6514):536–9, 1995.
- [157] R. Sordella, D. W. Bell, D. A. Haber, and J. Settleman. Gefitinib-sensitizing egfr mutations in lung cancer activate anti-apoptotic pathways. *Science*, 305(5687):1163–7, 2004. P01 95281/United States PHS Journal Article Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, P.H.S. United States.

- [158] P. Soubeyran, K. Kowanetz, I. Szymkiewicz, W. Y. Langdon, and I. Dikic. Cbl-cin85-endophilin complex mediates ligand-induced downregulation of egf receptors. *Nature*, 416(6877):183–7, 2002.
- [159] J. Stamos, M. X. Sliwkowski, and C. Eigenbrot. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J Biol Chem*, 277(48):46265–72, 2002.
- [160] T. A. Steitz, S. J. Smerdon, J. Jager, and C. M. Joyce. A unified polymerase mechanism for nonhomologous dna and rna polymerases. *Science*, 266(5193):2022–5, 1994.
- [161] A. P. Tarver, W. G. King, and S. E. Rittenhouse. Inositol 1,4,5-trisphosphate and inositol 1,2-cyclic 4,5-trisphosphate are minor components of total mass of inositol trisphosphate in thrombin-stimulated platelets. rapid formation of inositol 1,3,4-trisphosphate. *J Biol Chem*, 262(36):17268–71, 1987.
- [162] S. Tertyshnikova and A. Fein. $[ca^{2+}]_i$ oscillations and $[ca^{2+}]_i$ waves in rat megakaryocytes. *Cell Calcium*, 21(5):331–44, 1997.
- [163] S. Tracy, T. Mukohara, M. Hansen, M. Meyerson, B. E. Johnson, and P. A. Janne. Gefitinib induces apoptosis in the egfr1858r non-small-cell lung cancer cell line h3255. *Cancer Res*, 64(20):7241–4, 2004.
- [164] P. Traxler and P. Furet. Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. *Pharmacol Ther*, 82(2-3):195–206, 1999.
- [165] P. Traxler, J. Green, H. Mett, U. Sequin, and P. Furet. Use of a pharmacophore model for the design of egfr tyrosine kinase inhibitors: isoflavones and 3-phenyl-4(1h)-quinolones. *J Med Chem*, 42(6):1018–26, 1999.
- [166] D. Tschope, P. Rosen, and F. A. Gries. Increase in the cytosolic concentration of calcium in platelets of diabetics type ii. *Thromb Res*, 62(5):421–8, 1991. In Vitro Journal Article Research Support, Non-U.S. Gov’t United states.
- [167] R. Vadigepalli, P. Chakravarthula, D. E. Zak, J. S. Schwaber, and G. E. Gonye. Paint: a promoter analysis and interaction network generation tool for gene regulatory network identification. *Omics*, 7(3):235–52, 2003. Journal Article Research Support, U.S. Gov’t, Non-P.H.S. United States a journal of integrative biology.

- [168] J. D. Vickers, R. L. Kinlough-Rathbone, M. A. Packham, and J. F. Mustard. Inositol phospholipid metabolism in human platelets stimulated by adp. *Eur J Biochem*, 193(2):521–8, 1990.
- [169] G. L. Waldo and T. K. Harden. Agonist binding and gq-stimulating activities of the purified human p2y1 receptor. *Mol Pharmacol*, 65(2):426–36, 2004.
- [170] Y. Wang, R. I. Litvinov, X. Chen, T. L. Bach, L. Lian, B. G. Petrich, S. J. Monkley, D. R. Critchley, T. Sasaki, M. J. Birnbaum, J. W. Weisel, J. Hartwig, and C. S. Abrams. Loss of pip5kigamma, unlike other pip5ki isoforms, impairs the integrity of the membrane cytoskeleton in murine megakaryocytes. *J Clin Invest*, 118(2):812–9, 2008.
- [171] H. Waterman and Y. Yarden. Molecular mechanisms underlying endocytosis and sorting of erbb receptor tyrosine kinases. *FEBS Lett*, 490(3):142–52, 2001.
- [172] J. Weinstein and R. Radhakrishnan. kmc-tdgla coarse-grained methodology for simulating interfacial dynamics in complex fluids: application to protein-mediated membrane processes. *Molecular Physics*, 104(22):3653 – 3666, 2006.
- [173] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, Jr. Olson, J. A., J. R. Marks, and J. R. Nevins. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A*, 98(20):11462–7, 2001. CA 68438/CA/NCI NIH HHS/United States CA 84955/CA/NCI NIH HHS/United States Journal Article Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, P.H.S. United States.
- [174] James G. White. Platelet structure. In A. D. Michelson, editor, *Platelets*, pages 45–73. Elsevier Science, 2007.
- [175] S. J. Wiback, I. Famili, H. J. Greenberg, and B. O. Palsson. Monte carlo sampling can be used to determine the size and shape of the steady-state flux space. *J Theor Biol*, 228(4):437–47, 2004. Journal Article Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, Non-P.H.S. England.
- [176] D. B. Wilson, E. J. Neufeld, and P. W. Majerus. Phosphoinositide interconversion in thrombin-stimulated human platelets. *J Biol Chem*, 260(2):1046–51, 1985.
- [177] R. L. Winslow, A. Tanskanen, M. Chen, and J. L. Greenstein. Multiscale modeling of calcium signaling in the cardiac dyad. *Ann N Y Acad Sci*, 1080:362–75, 2006.

- [178] A. Wolf-Yadlin, N. Kumar, Y. Zhang, S. Hautaniemi, M. Zaman, H. D. Kim, V. Grantcharova, D. A. Lauffenburger, and F. M. White. Effects of her2 overexpression on cell signaling networks governing proliferation and migration. *Mol Syst Biol*, 2:54, 2006.
- [179] E. R. Wood, A. T. Truesdale, O. B. McDonald, D. Yuan, A. Hassell, S. H. Dickerson, B. Ellis, C. Pennisi, E. Horne, K. Lackey, K. J. Alligood, D. W. Rusnak, T. M. Gilmer, and L. Shewchuk. A unique structure for epidermal growth factor receptor bound to gw572016 (lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res*, 64(18):6652–9, 2004.
- [180] Y. Yarden and M. X. Sliwkowski. Untangling the erbb signalling network. *Nat Rev Mol Cell Biol*, 2(2):127–37, 2001.
- [181] Joel A. Ybe, Frances M. Brodsky, Kay Hofmann, Kai Lin, Shu-Hui Liu, Lin Chen, Thomas N. Earnest, Robert J. Fletterick, and Peter K. Hwang. Clathrin self-assembly is mediated by a tandemly repeated superhelix. *Nature*, 399(6734):371–375, 1999. 0028-0836 10.1038/20708 10.1038/20708.
- [182] D. L. Yee, C. W. Sun, A. L. Bergeron, J. F. Dong, and P. F. Bray. Aggregometry detects platelet hyperreactivity in healthy individuals. *Blood*, 106(8):2723–9, 2005. HL65229/HL/NHLBI NIH HHS/United States RR17665/RR/NCRR NIH HHS/United States Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. United States.
- [183] L. Zhang, P. Carroll, and E. Meggers. Ruthenium complexes as protein kinase inhibitors. *Org Lett*, 6(4):521–3, 2004.
- [184] X. Zhang, J. Gureasko, K. Shen, P. A. Cole, and J. Kuriyan. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell*, 125(6):1137–49, 2006.
- [185] Y. Zhang, A. Wolf-Yadlin, P. L. Ross, D. J. Pappin, J. Rush, D. A. Lauffenburger, and F. M. White. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics*, 4(9):1240–50, 2005. CA96504/CA/NCI NIH HHS/United States DK070172/DK/NIDDK NIH HHS/United States DK42816/DK/NIDDK NIH HHS/United States GM68762/GM/NIGMS NIH HHS/United States Comparative Study Journal Article Research Support, N.I.H., Extramural Research Support, U.S. Gov't, P.H.S. United States Mcp.

- [186] A. Zschauer, C. van Breemen, F. R. Buhler, and M. T. Nelson. Calcium channels in thrombin-activated human platelet membrane. *Nature*, 334(6184):703–5, 1988.