UNDERSTANDING IGE RECEPTOR SIGNALING THROUGH COMPUTATIONAL MODELING AND QUANTITATIVE EXPERIMENTS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Lily Ann Chylek

May 29th, 2016

UNDERSTANDING IGE RECEPTOR SIGNALING THROUGH COMPUTATIONAL MODELING AND QUANTITATIVE EXPERIMENTS

Lily Ann Chylek, PhD
Cornell University, 2016

FcεRI is a multi-subunit receptor found on the surface of mast cells and basophils and binds immunoglobulin E (IgE) with high affinity. Stimulation of this receptor, typically via antigen-mediated crosslinking of IgE, can lead to release of histamine and other mediators that are involved in the allergic immune response. Thus, understanding the workings of FcεRI-mediated signaling brings us closer to understanding one of the most widespread health conditions in the developed world. FcεRI-mediated signaling processes are inherently complex, involving highly interconnected systems in which quantitative factors can play a decisive role. Such systems can be difficult to decipher using intuition alone, but computational models can extend our reasoning abilities and enable us to develop nontrivial hypotheses that generate experimentally testable predictions.

To develop a systems-level understanding of how combinations of non-covalent interactions and post-translational modifications are regulated to impact cellular decision-making, we developed a computational interaction library. The library consists of executable rules for protein-protein and protein-lipid interactions. The library is visualized to facilitate understanding of network circuitry and identification of network motifs. Using this library, we investigated branching pathways from the adaptor protein Lat, which influence production of the phospholipid phosphatidylinositol (3,4,5)-trisphosphate (PIP_3) at the plasma membrane and the soluble second messenger inositol trisphosphate (IP_3). We found that inclusion of a positive feedback loop gives rise to a

bistable switch, which may ensure robust responses to stimulation above a threshold level. Such robustness has been observed experimentally for some readouts. We also developed a model that proposes an explanation for experimentally observed oscillations in Ca^{2+} concentration, which is an important outcome of FcεRI stimulation.

To investigate FcεRI signaling from another angle, we modeled interactions between the receptor and a structurally defined ligand for IgE. We parameterized the model for consistency with kinetic fluorescence data as well as super-resolution imaging measurements of antigen-induced receptor aggregation. To facilitate this study, we developed a specialized tool for fitting biochemical models to experimental data. Finally, we examined how patterns of exposure to stimulatory and non-stimulatory ligands affect mast cells' secretory responses. Through iterative modeling and experimental tests, we learned that a tug-of-war between positive signals from the tyrosine kinase Syk and negative signals from the lipid phosphatase Ship-1 govern the magnitude of responses, with the adaptor protein Shc1 influencing how the balance of positive and negative signals changes with time.

BIOGRAPHICAL SKETCH

Lily was born in Norman, Oklahoma and spent the subsequent years across various parts of North America. Along the way she developed a love of books, dogs, and science. After completing her undergraduate degree at Indiana University in 2009, she spent a year working in the Theoretical Biology Group at Los Alamos National Laboratory. In 2010, she joined Cornell University for her graduate work in the laboratory of Dr. Barbara Baird.

ACKNOWLEDGMENTS

My deepest thanks goes to my advisor, Barbara Baird, for supporting, guiding, and believing in me throughout my time in graduate school; to David Holowka, for being insightful, patient, and a true role model; and Bill Hlavacek for being constantly encouraging and opening so many doors. I also thank my committee members, Richard Cerione and Holger Sondermann, for always being helpful and for asking thought-provoking questions.

This work would not have been possible without the friendship and cooperation of many Baird-Holowka lab members and collaborators, past and present, particularly: Kari, Deepti, Norah, Sarah, Devin, Marcus, Kirsten, Amit, Eshan, Kate, Chris, Josh, Jordan, Ben, and Alice.

Thanks to my parents for all their support. And finally, I am grateful for the loyal and inspiring friends that I made in Ithaca as well as in Los Alamos, especially Jeanne, Clio, and Elena.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

BCR: B-cell antigen receptor

BNGL: BioNetGen Language

BSA: bovine serum albumin

Btk: Bruton's tyrosine kinase

Csk: C-Src kinase

DAG: 2,3-diacylglycerol

DNP: 2,4-dinitrophenol

ER: Endoplasmic reticulum

Fab: Fragment antigen-binding

Fc: Fragment crystallizable

FITC: fluorescein isothiocyanate

Gab2: Grb2-associated binding protein 2

Grb2: Growth factor receptor-bound protein 2

IgE: immunoglobulin E

Inpp5d: inositol polyphosphate-5-phosphatase D (see also Ship1)

IP_3: inositol trisphosphate

ITAM: immunoreceptor tyrosine-based activation motif

KD: knockdown

KMC: Kinetic Monte Carlo

LAT: linker for activation of T cells

Lyn: Lck/Yes novel tyrosine kinase

MARCKS: Myristoylated alanine-rich C-kinase substrate

ODE: ordinary differential equation

Pag1: phosphoprotein associated with glycosphingolipid-enriched microdomains

PI3K: phosphatidylinositol-4,5-bisphosphate 3-kinase

PIP5K: phosphatidylinositol-4-phosphate 5- kinase

PIP_2: phosphatidylinositol (4,5)-bisphosphate

PIP_3: phosphatidylinositol (3,4,5)-trisphosphate

PKC: protein kinase C

PLC: phospholipase C

Ptpn6: tyrosine-protein phosphatase non-receptor type 6 (see also Shp1)

SFK: Src-family kinase

SH2: Src homology 2

SH3: Src homology 3

Shp1: src homology region 2 domain-containing phosphatase-1 (see also Ptpn6)

Ship1: SH2 domain-containing inositol-5-phosphatase 1 (see also Inpp5d)

SOCE: store-operated calcium entry

STORM: stochastic optical reconstruction microscopy

Syk: spleen tyrosine kinase

TIRF: total internal reflection microscopy

TCR: T-cell receptor

Y: tyrosine

Introduction: Quantitative modeling of mast cell signaling

The paths of cellular decision-making

Cells exist in environments that necessitate responses. Growth, death, movement, and other fundamental acts are all influenced by decisions made on the basis of environmental cues. Decision-making is mediated by cell signaling systems, which translate inputs into outputs through networks of interacting proteins, lipids, and other biomolecules.

A signaling process begins when a cell surface receptor interacts with an extracellular ligand and becomes activated by undergoing a change detectable within the cell: two common examples of such changes are oligomerization and conformational alterations. From the activated receptor, multitudinous biochemical events can emanate. These events, which include catalysis of post-translation modifications and formation of multi-protein complexes, can culminate in consequences such as migration, secretion, and changes in gene expression.

To understand how cells make decisions, and to potentially predict and manipulate these decisions, we need to understand the process by which a signaling system translates inputs to outputs. We can uncover the pieces of this puzzle through experimental measurements, such as western blots revealing dynamics of protein phosphorylation or a microscopy image revealing cellular localization of biomolecules. To integrate these puzzle pieces into a picture of the cell's inner workings, we need to assemble a model: a representation summarizing how we think the system works.

Models in cell biology

Models are ubiquitous in biology. They often take the form of maps that show proteins connected by arrows that represent various types of influences or interactions. The value of these diagrams is that they provide a visual overview that can be comprehended more readily than a textual description of the system's components and connections. However, our ability to enumerate and illustrate such connections tends to outstrip our ability to predict the outcomes that those interactions lead to, especially when trying to account for the quantitative factors that are inherent to signaling systems, such as the number of copies of a protein, the affinity of an interaction, or relative competition among binding partners. Another compounding factor is the intricacy of the "wiring" itself; cell signaling systems are rife with feedback and feed-forward loops (Alon, 2007), cross-talk (Guo and Wang, 2009) and redundancies (Sun and Bernards, 2014). These factors contribute to a variety of behaviors that seem at first perplexing: hysteresis (Das et al., 2009), oscillations (Wollman and Meyer, 2012), and desensitization (Weetall et al., 1993). It can be difficult to develop an intuition about these phenomena. A strategy to extend our reasoning capabilities about these complex and quantitative systems is to make our maps executable: to represent biomolecular interactions in a way that enables a computer to simulate them.

A spectrum of approaches has been used to model aspects of cell signaling, from molecular dynamics simulations of molecular movements on the sub-nanosecond timescale (Karplus and McCammon, 2002), to Boolean networks representing a coarse-grained view of the overall network (Albert and Thakar, 2014). Here, we will focus on models based on chemical kinetics (Aldridge et al., 2006) because they combine features that are especially relevant for studying cell signaling in a combined modeling-experimental framework: 1) these models take into account laws of physics and chemistry that govern molecular interactions in reality, 2) they

have the potential to encompass a large set of proteins, and 3) they can be used to simulate timescales from seconds to minutes to hours, which are timescales on which experimental measurements are commonly performed.

Quantitative analysis of IgE receptor signaling

One of the earliest signaling systems to be investigated through mathematical modeling and quantitative experiments is signaling via the high-affinity receptor for IgE, also known as FcεRI (Dembo and Goldstein, 1978). This multi-subunit receptor is found on the surface of mast cells and basophils. Stimulation of this receptor can lead to release of histamine and other mediators that are involved in the allergic immune response (Kraft and Kinet, 2007). Thus, understanding the workings of FcεRI signaling brings us closer to understanding one of the most widespread health problems in the developed world. Additionally, IgE receptor signaling is similar to signaling by other antigen recognition receptors, so studying this system has relevance for general understanding of immunity, beyond allergies.

FcεRI forms a stable, one-to-one complex with an IgE antibody (McDonnell et al., 2001), which is specific to a particular antigen. IgE is bivalent for antigen binding, meaning that its two Fab arms can each bind to a separate antigen site. Antigens that are multivalent are capable of cross-linking multiple receptors and forming receptor aggregates. These aggregates can initiate signaling by allowing kinases co-localized with clustered receptors to phosphorylate neighboring receptors. The resulting signaling cascade, including influx of extracellular Ca^{2+}, leads to release of pre-formed inflammatory mediators in a process called degranulation (Fig. 1). Aggregation is the mode of signal initiation for multiple immunoreceptors, as well as a process found at various

stages of diverse signaling systems. The parallels between FcεRI and other signaling systems make it a useful model system.

The mast cell system offers an unusual level of experimental tractability. One reason is that the antigen the cells will respond to is determined by the specificity of the IgE antibodies that are bound to cell-surface FcεRI. Exposing cells to different IgEs, separately or in combination, results in cells that are responsive to different (combinations of) stimuli.

Additionally, a variety of synthetic antigens can be constructed by attaching small molecules corresponding to antigenic sites, called haptens, to polymer scaffolds; these ligands can then bind to hapten-specific IgE. This feature has been leveraged to study a wide range of structurally distinct ligands, which give rise to a variety of responses. One of the most commonly used ligands is DNP-BSA, which is a bovine serum albumin (BSA) molecule conjugated to multiple 2,4-dinitrophenyl (DNP) haptens. This ligand stimulates robust signaling and degranulation, but a drawback is that it is chemically heterogeneous; different DNP-BSA molecules may bear different numbers of DNP groups, and two molecules with the same number of DNP groups may have them conjugated to different lysine residues of the protein. Thus, the mechanisms of ligand-receptor interactions are difficult to deconvolute. More controlled ligands have been constructed and used to shed light on what qualities of a ligand influence the magnitude of a signaling response.

Valency has been found to be an important factor. Monovalent ligands, such as DNP-lysine and DNP-aminocaproyl-L-tyrosine (DCT) do not stimulate signaling because they do not enable formation of receptor aggregates. However, monovalent ligands are a useful tool for studying inhibition of signals triggered by multivalent ligands, because an excess of monovalent

ligand is capable of breaking up aggregates (Subramanian et al., 1996), leading to downregulation of signaling.

Most bivalent ligands generate minimal degranulation (Paar et al., 2002), suggesting that the kind of aggregates that they form do not stimulate sufficient signaling within the cell. However, it is worth noting that some bivalent antibodies that bind to FcεRI directly stimulate signaling, suggesting that the steric constraints and/or binding properties of different ligands may have a deciding effect.

Trivalent ligands tend to stimulate a more robust range of degranulation responses. A set of ligands built on a rigid double-stranded DNA scaffold revealed that the longer the length of the scaffold, the weaker the resulting degranulation and phosphorylation of most, but not all, signaling proteins, including the receptor itself as well as the adaptor protein LAT (Sil et al., 2007). Degranulation decreased with ligand length, as did stimulated Ca^{2+} mobilization, but in a less drastic manner. These results support the concept of transphosphorylation by receptor-associated kinases being important to early signaling, with larger distances between receptors reducing the efficiency of transphosphorylation. These results also point to the existence of diverging pathways that enable some signaling components to be more independent of receptor phosphorylation than others.

Other trivalent DNP ligands include one built on a synthetic polymer (Posner et al., 2007), and one consisting of a trimeric fibritin foldon domain bound to DNP (Mahajan et al., 2014). The range of ligands available for manipulating clustering of IgE-FcεRI complexes has enabled a controlled analysis of the intracellular information-processing system that connects inputs to outputs, and many of these analyses have included modeling.

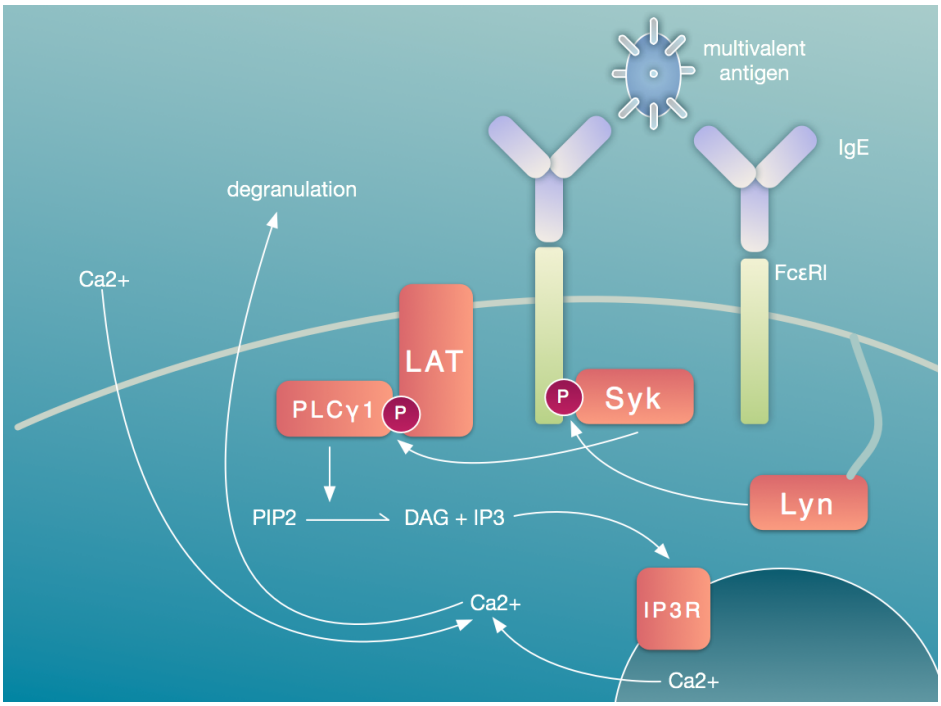


Fig. 0.1. An illustration of molecules involved in initiation of IgE receptor signaling. A one-to-one complex is formed between FcεRI and an IgE antibody. Each of the IgE's two Fab arms is capable of interacting with an antigen. A multivalent antigen is able to cross-link (aggregate) two or more IgE-receptor complexes. These receptor aggregates can then undergo phosphorylation by the kinase Lyn, which is tethered to the membrane via palmitoylation and myristolation. Lyn is co-localized to receptors via association with its intrinsically disordered unique domain and/or through association with lipid rafts (Holowka et al., 2005, Young et al., 2003). Lyn phosphorylates the receptor at several residues in its cytoplasmic subunits, each of which contains an immunoreceptor tyrosine-based activation motif (ITAM). An ITAM contains two tyrosine residues that can be phosphorylated. The beta subunit contains an atypical three tyrosines. The N-terminal one, when phosphorylated, can associate with the SH2 domain of Lyn. The pair of disulfide-linked gamma subunits each contain one ITAM that, when doubly phosphorylated, can associate with the tandem Src homology 2 (SH2) domains of the kinase Syk. Syk can then proceed to phosphorylate an array of downstream targets, including the adaptor LAT, which can then bind to PLCγ1. PLCγ1 cleaves the phospholipid PIP2, yielding diacylglycerol (DAG) and inositol trisphosphate (IP3). Interactions of IP3 with the IP3 receptor on the endoplasmic reticulum leads to calcium mobilization. Store-operated calcium entry (SOCE) enables influx of extracellular calcium into the cytoplasm. These events enable release of inflammatory mediators through degranulation.

Chemical kinetic models of IgE receptor signaling

Most early modeling studies of this system focused on the prerequisite for receptor-mediated signaling, which is ligand-receptor binding. Models for the kinetics of the system took the form

of ordinary differential equations (ODEs) that represented the changing concentrations of chemical species in the model, which were controlled by the processes of ligand capture, receptor crosslinking, and cyclization to form rings (Posner et al., 1991, Posner et al., 2005, Subramanian et al., 1996, Xu et al., 1998, Das et al., 2008). Equilibrium models in the form of algebraic expressions were also developed. Some of these models were fit to binding data, which was usually in the form of fluorescence measurements (Erickson et al., 1986). The fluorescence of IgE labeled with fluorescein isothiocyanate (FITC) is reduced in the presence of the hapten DNP. Thus, measurements of FITC fluorescence quenching can be translated to a quantity of IgE sites bound. Additionally, in some studies ligands were also fluorescently labeled and the amount of ligand bound to IgE was quantified, providing additional data for models to be fit to (Xu et al., 1998). These combined modeling and experimental studies were used to postulate mechanisms and estimate rate constants involved in ligand-receptor binding.

These studies of events on the cell surface provided a starting point for investigations into what occurs inside the cell. In 1997, IgE receptor signaling became the subject of one of the first modeling studies of intracellular signaling in a mammalian system (Wofsy et al., 1997). Going beyond ligand-receptor interactions, this model was used to investigate the relationship between receptor aggregation, receptor phosphorylation, and association of aggregates with the kinase Lyn. The ligand considered in this model, and the associated experiments, was a covalently cross-linked dimer of IgE that can bind to a pair of receptors. Phosphorylation sites in the receptor were simplified to a single site. Parameters for the model were obtained through a combination of published values and values obtained through fitting to experimental data. The model was used to make predictions about how Lyn is redistributed as receptor aggregates form, concluding that late-forming aggregates are less likely to contain Lyn than early-forming

aggregates. However, the model also predicted that under some conditions, late-forming aggregates can recruit Lyn away from aggregates formed earlier. This prediction was confirmed experimentally, an example of how a model can make a prediction that would be difficult to arrive at otherwise (Wofsy et al., 1997).

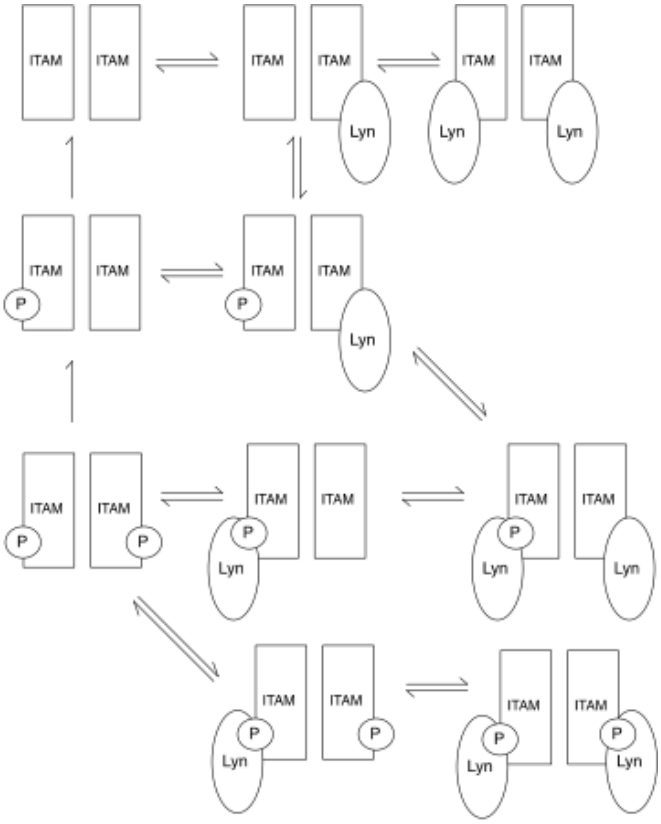


Fig. 0.2. A subset of the reactions included in the model of Wofsy et al. (1997). These reactions represent interactions between a receptor dimer, crosslinked by a covalently-linked dimer of IgE, and Lyn kinase. Lyn can interact with an unphosphorylated receptor via its unique domain. Lyn can interact with a phosphorylated receptor via its SH2 domain. In this model, receptor phosphorylation sites are lumped together as one single phosphorylation site. This model does not explicitly include phosphatases. Phosphatase activity is implied in the rapid dephosphorylation of receptors that are not bound to Lyn. In addition to the reactions shown here, the model also contains reactions for binding between receptor and ligand (the IgE dimer).

The next step in model development was addition of another kinase, Syk. However, this step required a rethinking of methodology. Recall that previous kinetic models were built by writing ODEs, which means that an ODE must be written for every chemical species that is

potentially populated in the system being modeled. As the number of interactions grows, the number of potentially populated species can explode. A model that included recruitment and activation of Syk entailed 354 distinct species (Faeder et al., 2003). Despite its large size from the perspective of equations, from the perspective of molecules the model only contained four players (a ligand, the receptor, Lyn, and Syk) out of the many that are known to be involved in mast cell signaling. It was apparent that continued progress in model development would necessitate a new approach to how we think about models.

The advent of rule-based modeling

So far, models have been built on the notion of reactions, where each reaction describes the behavior of a very specific set of chemical species. The number of reactions can be reduced by making them less specific (more general), and thus less numerous. A way to do this is to describe a generalized reaction, called a rule, and to task a computer with enumerating the specific reactions. A rule describes the necessary and sufficient conditions required for an interaction to occur. For example, a rule may state that for Lyn to bind a phosphorylated site on the receptor, Lyn must have an unbound SH2 domain and the phosphorylated site in the receptor must also be unbound. Any other molecular details that are assumed or known to not influence the interaction, such as phosphorylation and occupancy of other receptor sites, are omitted from the rule. A rule is a generalized representation of a reaction that could occur in many different instances that vary in their nonessential molecular details.

The specific, potentially numerous reactions implied by a set of rules can be enumerated by a program that processes the rules. From this point, ODEs can be derived from the list of reactions for numerical integration, or the reactions can be used as event generators in a

stochastic simulation based on Gillespie's method (Gillespie, 1976). In this way, a signaling network can be represented, by the human modeler, using a relatively compact set of rules and the drudgery of translating those rules into reactions is performed by a computer program (Faeder et al., 2009).

Several languages and software packages have been developed to aid rule-based modeling. One of the most commonly used is BioNetGen. We will focus on BioNetGen, which enables both stochastic and deterministic simulations. In BioNetGen's modeling language, called BNGL, the connectivity of molecules of interest are represented, essentially as graphs, along with their relevant internal components, which can include domains, motifs, and residues. If desired, components can have multiple possible states representing component properties such as location, post-translational modification status, conformation, or other variables.

For example, the text "`Syk(tSH2,Y519~0~P)`" is a string-based encoding of a graphical representation of the kinase Syk, which is taken here to contain a tandem pair of SH2 domains called tSH2 and a tyrosine called Y519. The text indicates that this tyrosine can either be unphosphorylated (`0`) or phosphorylated (`P`). An example of a rule is:

```
Syk(tSH2) + FcR(gITAM~PP) -> Syk(tSH2!1).FcR(gITAM~PP!1) k1
```

This rule states that the tSH2 domain in Syk can bind to a (doubly) phosphorylated ITAM in the gamma subunit of the IgE receptor. The bonded components are labeled with "!1". We are assuming that the phosphorylation state of the tyrosine in Syk, Y519, has no bearing on this interaction, and thus it is omitted from the rule. Reactions enumerated from this rule would include a reaction where Y519 in Syk is phosphorylated, and one where it is not phosphorylated. Both reactions would proceed with the same rate constant, `k1`. If we had reason to expect that the phosphorylation status of Y519 in Syk would affect binding of its tSH2 domain, which is true

of some tyrosine residues in this kinase (Grädler et al., 2013), the rule could be modified to include its phosphorylation state.

The 2003 model for IgE receptor signaling that was developed using rule-based methods revealed the effects of kinase copy numbers, dephosphorylation kinetics, and how the beta subunit of the receptor can act as either an amplifier or inhibitor (Faeder et al., 2003).

Representing a simple reaction network with rules, reactions, and equations

To illustrate how ODE-based and rule-based models represent the same chemical kinetics, we will go over a simple example toy model for IgE receptor signaling. We will assume that the IgE receptor has two binding sites, located in its beta and gamma subunits, for associating with the kinases Lyn and Syk, respectively, and that interactions with each kinase occurs independently of the other. We will consider the process by which the receptor becomes associated with both kinases.

The receptor (R) may first bind to Lyn (L), forming complex C_L, or it may first bind to Syk (S) forming complex C_S. From there, a final complex is formed when C_L binds to Syk or when C_S binds to Lyn. These complexes are illustrated in Fig. 3A. This set of interactions implies four reactions, illustrated in Fig. 3B. Each reaction is associated with a forward and reverse rate constant. We assume that the rate constant of Lyn or Syk binding is identical for when the other kinase is already bound to the receptor. Fig. 3C organizes these reactions into a network that illustrates the receptor transitioning from fully unbound to fully bound.

From the reactions of Fig. 3C, we can obtain the ODEs presented in Fig. 3D. The variables x_1, x_2, x_3, x_4, x_5, and x_6 represent the concentrations of the species R, L, S, C_L, C_R, and C_{LS}, respectively. The equations describe the change in concentration of each of these species over time. The equations are found by considering which reactions either contribute to the

formation or destruction of the species in question. For example, C_L is formed by binding of Lyn to the receptor, and consumed by dissociation of Lyn from the receptor (leaving L and R), as well as by binding of Syk (forming the next species, C_{LS}). We obtain terms in the equation by multiplying the rate constant by the concentration(s) of species involved in the reaction. The sum of these terms makes up the complete ODE. There are six ODEs in total.

In panel E, we show the same process represented by rules. In each rule, we list the reaction centers in each molecule, that is, the part of the molecule that is involved in the interaction. For binding of Lyn to the receptor, the reaction centers are the phosphorylated beta subunit of the receptor (b~P) and the Lyn SH2 domain. For binding of Syk to the receptor, the reaction centers are the phosphorylated gamma subunit of the receptor (g~P) and the Syk SH2 domain. Because we assume that the other kinase does not affect the rate of the interaction, it is omitted from the rule. The first rule encompasses reactions 1 and 3 in panel B, whereas the second rule encompasses reactions 2 and 4 in panel B.

It is important to emphasize that this toy model is simplistic in multiple respects. One simplification is that the receptor is taken to be permanently in a phosphorylated state. To illustrate how the two modeling approaches fare when model extension is pursued, we will consider what happens when we include the receptor switching from an unphosphorylated to phosphorylated state.

In the reaction network, we would need to introduce an entirely new set of species to account for a receptor that is completely unphosphorylated, phosphorylated only at the beta chain, phosphorylated only at the gamma chain, and phosphorylated at both. (Each chain contains multiple tyrosine residues, but as a simplification we are lumping all of the sites in a given chain as a single site of phosphorylation.) We would then have to consider two versions of

each of the receptor-kinase complexes, differing based on receptor phosphorylation state. We would also have to account for phosphorylation of a receptor site when the other site is phosphorylated and bound. The 12 reactions are shown in Fig. 4A. There are a total of 10 distinct chemical species, and an ODE would need to be written for each one.

In contrast, in the rule-based model we would only need to add two more rules, each describing the independent phosphorylation of one receptor site (Fig. 4B). These rules can be combined with the rules of Fig. 3E, without having to edit the original rules. From this example it can be seen that expanding the rule-based model to account for more details will be a faster (and less error-prone) task than writing the reactions and equations by hand.

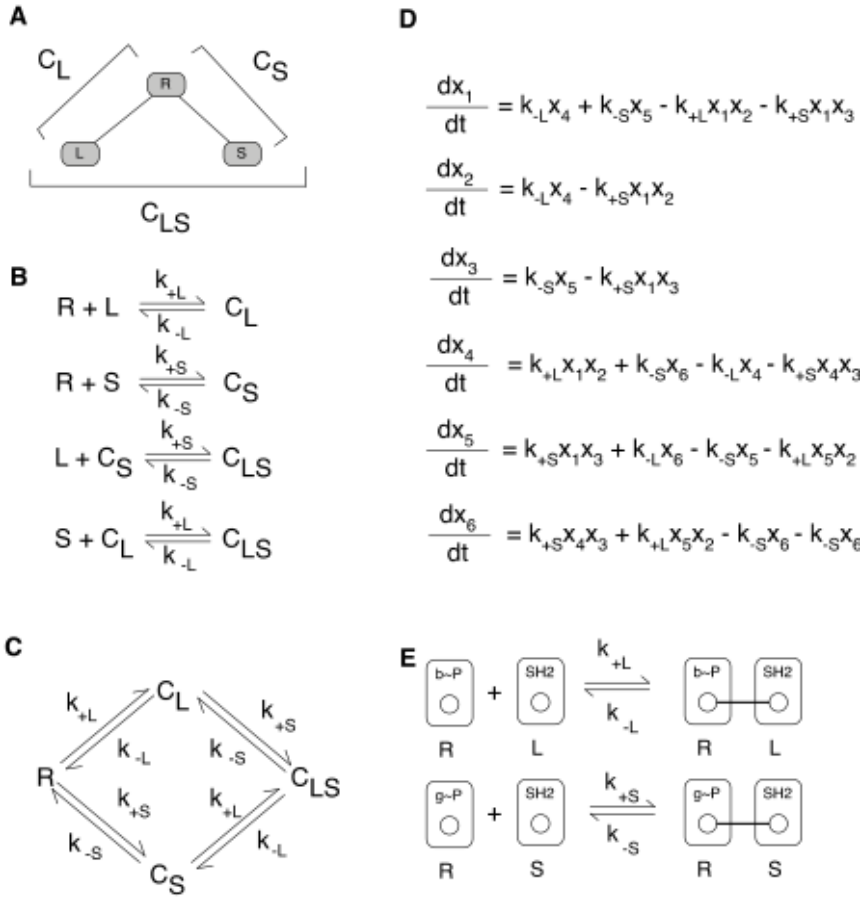


Fig. 0.3. A comparison of representations of a chemical reaction network. **A**. The receptor (R) can bind to Lyn (L) and to Syk (S). A complex of the receptor with Lyn only, the receptor with

Syk only, and the receptor with both Lyn and Syk are named R_L, R_S, and R_{LS} respectively. **B**. The list of reactions describing the binding processes of panel A. C_L is a complex of the receptor and Lyn only. C_S is a complex of the receptor and Syk only. C_{LS} is a complex of the receptor, Lyn, and Syk. **C**. The reactions in this system form a reaction network. In the reaction scheme shown, vertices are receptor complexes, and the arrows between them represent the chemical reactions that account for formation or destruction of each species. Each reaction is associated with a rate constant. **D**. The ordinary differential equations (ODEs) derived from the chemical reaction network. The variables x_1, x_2, x_3, x_4, x_5, and x_6 represent the concentrations of the species R, L, S, C_L, C_S, and C_{LS} respectively. These equations describe the changes of each species' concentration, as determined by reaction rates, over time. **E**. The rules representing an equivalent set of reactions. These rules can be processed by rule-based modeling software to yield the reactions and equations shown in other panels of this figure, or used directly to advance the state of a system if one is using network-free simulation techniques (described later).

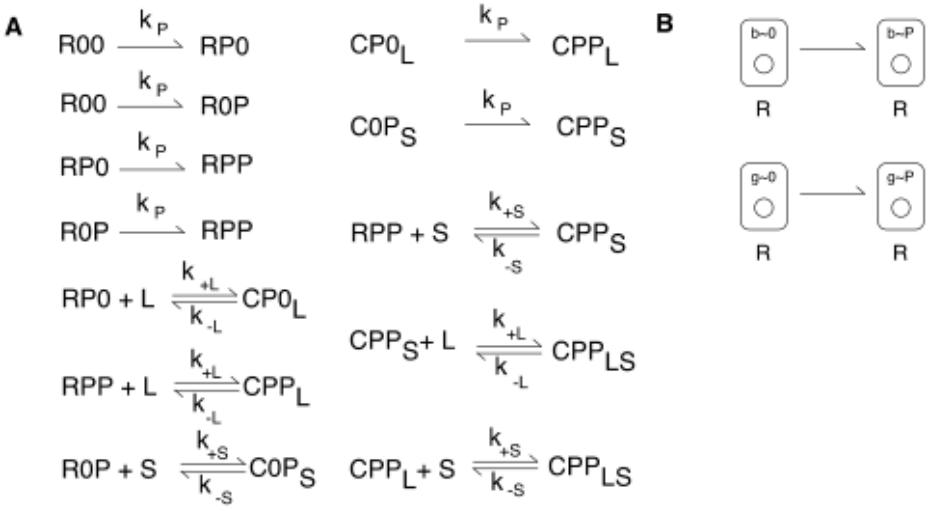


Fig. 0.4. An extension to the model of Fig. 0.2. **A**. This panel shows the chemical reactions that would need to be written if one were to add a simple mechanism whereby receptors are phosphorylated via a first-order reaction. The species R00, RP0, RP0, and RPP represent, respectively, the receptor in an unphosphorylated form, phosphorylated at beta only, phosphorylated at gamma only, and phosphorylated at both sites. The same nomenclature extends to the complex species (C0P, CP0, and CPP). To turn these reactions into a system of ODEs, the equations of Fig. 2D would have to be modified, and new equations would need to be written to account for the new species. **B**. In contrast, only two additional rules would need to be added to the rules of Fig. 2E for the model to be expanded. The original rules would not need to be modified.

Towards comprehensiveness and modularity

As our knowledge of signaling proteins grows, we want to develop models that encompass more molecular players, especially when we have ever-increasing ways to probe and quantify cellular behaviors. For example, mass spectrometry (MS) has made it possible to track phosphorylation states of hundreds to thousands of specific protein sites simultaneously (Cox and Mann, 2011). Super-resolution microscopy has revealed detailed pictures of how molecules spatially organize themselves within the cell (Shelby et al., 2013). Microfludic devices make it possible to implement complex time-varying inputs and to monitor what cells do in response (Cheong et al., 2009). To remain relevant, modeling techniques have needed to advance to keep up with the array of directions in which experimental data are expanding.

One of these advances has been the development of new simulation techniques. Although generating reactions from a rule set is feasible for many models, for some models the number of possible reactions is too large for even this process to be manageable. In some cases, it is the result of the model including many different proteins in the signaling network. In other cases, there may be only a few different types of proteins that can, however, combine with each other in seemingly endless configurations. An example of such a situation is the problem of a bivalent receptor (e.g., the FcεRI-IgE complex) and a trivalent ligand (e.g., one of the several trivalent ligands described above). The branching aggregates that may arise in this scenario are too numerous to be enumerated (Monine et al, 2010), but as described above, it is a biologically relevant process and thus something that we would like to be able to simulate.

A solution to this problem came in the mid 2000's with the development of network-free simulation methods (Yang et al., 2008, Danos et al., 2007). As the name suggests, this method bypasses the step of generating a chemical reaction network, by using the rules themselves as event generators in a stochastic simulation protocol based on Gillespie's method (Gillespie,

1976). In short, the rates of rules are calculated based on the properties of individually tracked sites in a model, and a waiting time is determined. A rule is selected for execution, reactive sites are selected, and the system state is updated (through site state changes consistent with the selected rule), and time is incremented. Several variants of this algorithm exist, and are encoded in the general-purpose software tools DYNSTOC (Colvin et al., 2009), RuleMonkey (Colvin et al., 2010), NFsim (Sneddon et al., 2011), and KaSim (http://www.kappalanguage.org/).

Taking advantage of network-free simulation, large models have been built and analyzed. An example is a large model for T-cell receptor (TCR) signaling that, like FcεRI, is a multi-subunit immunoreceptor. The model was fit to mass spectrometry-based time-resolved measurements of phosphorylation at specific protein sites that characterize signaling during the first 60 s of TCR stimulation, yielding an unprecedented number of matched simulation and experimental time courses, and the model was also found to be predictive (Chylek et al., 2014). Another example of a large model is one for growth factor signaling (Stites et al., 2015), which was also used to analyze proteomic datasets characterizing protein copy numbers in an array of cell lines and affinities of protein-protein interactions. Recently developed fitting software (Thomas et al., 2015) will facilitate further efforts to optimize parameter values of large models so that these models are consistent with high-throughput experimental measurements.

In parallel with construction of these models, rigorous methods for model visualization and annotation were developed to handle the issue of how we can represent what a model contains and where the information comes from (Chylek et al., 2011). These methods include visualization of models as extended contact maps and their annotation with model guides. Model guides may take the form of a wiki (Creamer et al. 2012), pointing the way towards community-driven modeling efforts where new models can be built upon the old.

Recently, we used these visualization, simulation, and annotation techniques to develop a large model for IgE receptor signaling that also serves as a modular library from which more specialized models can be obtained (Chylek et al., 2014c). This model incorporates many features of IgE receptor signaling that had previously been part of separate signaling models, combined with additional proteins and interactions that had not been considered in models before. As described in Chapter 2, subset of components from this library was used to develop a model addressing a question raised in one of the differential ligand response studies discussed above (Sil et al., 2007): namely, why do some responses depend heavily on upstream phosphorylation whereas others do not? The model predicted that bistability in lipid synthesis, resulting from positive feedback loops, may be a factor (Chapter 2). A model lacking positive feedback predicted that when phosphorylation of the adaptor LAT was reduced, all downstream outputs would be reduced to an equal or greater extent. In contrast, a model including a positive feedback loop involving the adaptor protein Gab2 and the lipid kinase Phosphoinositide 3 kinase (PI3K) predicted that under some conditions, a subset of downstream outputs would be buffered from reductions in upstream phosphorylation. These results provide a possible explanation for why ligands that induce different levels of receptor phosphorylation can induce similar levels of some downstream phosphorylation targets: as long as upstream phosphorylation exceeds a minimum level, positive feedback can sustain robust activation of certain downstream pathways. These predictions are experimentally testable by, for example, disrupting the feedback loop through an siRNA knockdown of one of its components.

As described in Chapter 3, another recent example of an IgE receptor signaling model was used to investigate the mechanisms underlying mast cells' responses to complex waveform inputs. Pulse-chase experiments revealed that the response of a mast cell to repeated stimulation

depends on the frequency of stimulations, with a high frequency leading to attenuated responses and a low frequency leading to enhanced responses. A model was developed that combines known features of the signaling system with hypothesized new links to explain the origins of this non-intuitive response pattern. The predictions of this model were tested experimentally.

Considerations on how a model is built

A rule essentially states three pieces of information: 1) which molecules are interacting, 2) which sites are directly affected by the interaction (by becoming bound, unbound, or post-translationally modified), and 3) which sites are not directly affected by the interaction, but that exert an influence on the interaction.

To explore what is meant by each piece of information, let us return to the example of Syk interacting with FcεRI, and model it in more detail by considering the two SH2 domains separately. We can write a rule stating that the N-terminal SH2 domain of Syk binds to the C-terminal tyrosine in the gamma ITAM of FcεRI, and that this interaction proceeds with a forward rate constant, *k*.

```
Syk(nSH2) + Rec(g_Y2) -> Syk(nSH2!1).Rec(g_Y2!1) k
```

We can also specify that the kinetics of the interaction are affected by whether the other SH2 domain of Syk is already bound to the receptor:

```
Syk(nSH2,cSH2) + Rec(g_Y2) -> Syk(nSH2!1,cSH2).Rec(g_Y2!1) k1
Syk(nSH2,cSH2!1) + Rec(g_Y1!1,g_Y2) -> Syk(nSH2!2,cSH2!1).Rec(g_Y1!1,g_Y2!2)
k2
```

The first rule states that the C-terminal SH2 domain of Syk is unbound. The second rule states that this C-terminal SH2 domain is already bound to the receptor. The forward rate constant for this second rule, k2, will be larger than k1 because Syk is already tethered to the receptor, increasing the likelihood that the N-terminal SH2 domain will encounter its binding partner.

The primary literature is rich in information about biomolecular interactions. Many models are based on rules gleaned from potentially hundreds of individual studies of investigators who characterized signaling systems, one interaction at a time. These types of studies have been carried out using biochemical techniques such as immunoprecipitation (Markam et al., 2007), or more biophysical techniques like Förster resonance energy transfer (FRET) (Piston and Kremers, 2007). Other studies have used more high-throughput techniques, such as mass spectrometry or protein microarrays, to identify a large number of interaction partners simultaneously (Hause et al., 2012). One consideration to keep in mind is to what extent different studies were performed under different conditions, and whether this affects the compatibility of different data sets.

The specific sites involved in interactions have been uncovered by implementing mutations (e.g., mutating a particular tyrosine to a phenylalanine, and determining whether a certain interaction is affected by the change). Some experiments have also shed light on what protein components, beyond the ones that are directly interacting, can influence the interaction. When this type of contextual information is available, it is usually a good idea to make a rule more specific by including it.

Once a rule is formulated, parameters must be supplied, which can come from many different sources. The parameters for some specific interactions have been quantified through experimental studies. Sometimes these studies identify both forward and reverse rate constants, whereas others only identify the ratio of the two (in the form of an affinity or dissociation constant); this latter type of measurement is still useful because it constrains our estimates of rate constants.

Other interactions have not been studied at the level of affinities and rate constants, but measurements for analogous interactions can sometimes suggest a typical value. In cases where a specific value has not been determined, fitting (i.e., the optimization of parameter values to achieve stimulated behavior matching experimental data) is useful. A general-purpose fitting tool for rule-based modeling has recently been developed (Thomas et al., 2015). This tool implements a genetic algorithm that makes it possible to optimize multiple parameters simultaneously while also leveraging parallel computing resources.

As a modeler learns more about protein interactions and writes more rules, the model grows, and the eventual scope of a model must be defined. By scope, we mean how much of a network's proteins, protein components, and interactions should be included in a model. The answer is determined by the question that the model is being formulated to address. For example, if the goal is to develop a model that reproduces a defined set of experimental time courses of phosphorylation, the proteins that were measured should be the focus of the model, along with any players needed to fill out the connections between them. In other cases, the scope might not be so clearly defined, at first. For example, one might need to develop a model with the goal of reproducing a certain overall cellular behavior (e.g., desensitization) that could result from the confluence of multiple biochemical pathways (e.g., various positive and negative signals). There is no clear-cut procedure for deciding what to include in such a model, but the trial-and-error process of discovering which components of a network are needed to produce a specific behavior can be an informative and surprising undertaking in itself.

Although the literature contains a great deal of information about interaction networks, it doesn't spell out the answer to every question. Sometimes, possibilities that aren't emphasized or even explored in individual studies can actually be significant when the system is regarded as a

whole. The modeler's starting point is to take information that is already known, turn it into an executable form, and to determine whether it forms a coherent picture or whether there are gaps that need to be filled.

How do we know when gaps exist? We begin by trying to ground a model in reality by comparing it to relevant experimental data and working towards agreement between the two (by including various known network components and exploring reasonable parameter values). The presence of gaps is suggested when a model seems like it should be able to match an experimental data set, but it can't. Discrepancies between model and data are often the start of something interesting, because it indicates that some factor (in the model, experiments, or both) has not been considered.

Sometimes, to fill these gaps we need to propose new interactions that haven't been reported before. At other times, we need to look at available information in a different way; it's possible that the necessary factors have already been found, but that their combined effect hasn't been appreciated. In either case, the model should then be used to make experimentally testable predictions that could be used to support or disprove the model.

For example, in the TCR modeling study discussed above it was found that the conventionally accepted pathway for recruitment of the protein WASP to the T-cell receptor was insufficient to account for observed rapid dynamics of WASP phosphorylation. An alternative, shortcut pathway was needed. The shortcut's individual components had been described before, but hadn't been regarded as forming a pathway for fast recruitment of WASP. The model also pointed to previously uncharacterized roles for the tyrosine phosphatase SHP1 (PTPN6). Both aspects of the model were supported by experiments prompted for the purpose of testing the model (Chylek et al., 2014a) as well as in later studies (Paensuwan et al., 2015).

Model visualization is almost always helpful, especially if a model is large. In addition to making a model more understandable, it can also serve as a basis for identification of network motifs, such as feedback loops. These motifs have been identified by hand (Chylek et al., 2014c), followed by tools that automate this type of analysis.

The parameter values of a model can be explored in several ways. If there is reason to think that the model could exhibit bistability (the existence of two steady states, which can result from, for example, positive feedback), bifurcation analysis is a possibility. This type of analysis is performed by incrementing a parameter from a low value to a high value and plotting the steady-state value of the resulting outputs for each value of that parameter. Then, the parameter is incremented from high to low, and outputs are plotted again. If there is a range of parameter values where a given output has a different value in the two cases (the low-to-high case, and the high-to-low case), then that is a region of bistability. Bifurcation analysis has been employed in recent studies of B-cell antigen receptor signaling (Barua et al., 2012) and FcεRI signaling (Chylek et al., 2014c).

Another common type of analysis that can be applied to any model is sensitivity analysis, whereby a parameter is altered by a certain percentage, and then the difference in output values are quantified. The ratio of the two is a sensitivity coefficient that informs us of which parameters in the model have the greatest bearing on specific model behaviors (Barua et al., 2012).

In a similar vein, one can explore what a model does under the conditions of various "virtual experiments." Knockdowns, overexpressions, mutations, and the like can all be achieved by (usually straightforward) modifications of the model. For example, a knockdown is achieved by lowering the copy number of a protein, an overexpression by increasing copy number, and a

mutation by modifying the functionality of a component in a protein. This type of exploration may reveal interesting directions for future studies.

At each step in the history of quantitative investigation of FcεRI signaling, the content of models has been determined by multiple factors, including the state of knowledge in the field, the type of experimental data available for comparison, and the status of modeling software itself. The models of FcεRI signaling form an example of how different considerations influence the content of a cell signaling model, and how the very concept of what a model is has grown over time, from a picture of interactions, to a system of equations, to an executable program of rules.

In Chapter 1, we present a rule-based interaction library for FcεRI signaling. In Chapter 2, we use this library to develop models for specific aspects of intracellular signaling, and discuss experimentally testable predictions of these models. In Chapter 3, we present an integrated experimental and modeling study that elucidates the frequency response properties of mast cells. Chapter 4 discusses a tool for fitting rule-based models to experimental data. This tool is utilized in Appendix 1, where we present a model for ligand-receptor binding with a preliminary fit to super-resolution imaging data, and outline further steps for software development that will tighten the link between experiments and modeling. Finally, Appendix 2 details model visualization methods that are used throughout this work.

References

Albert, R., Thakar, J. Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. *Wiley Interdiscip Rev Syst Biol Med*. **6**, 353-369. (2014)

Aldridge, B.B. Burke, J.M., Lauffenburger, D.A., Sorger, P.K. Physicochemical modelling of cell signalling pathways. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol*. **8**, 1195-203. (2006)

Alon, U. Network motifs: theory and experimental approaches. *Nat Rev Genet*. **8**, 450-4561. (2007)

Barua, D., Hlavacek, W.S., Lipniacki, T. A computational model for early events in B cell antigen receptor signaling: analysis of the roles of Lyn and Fyn. *J Immunol*. **189**, 646-658. (2012)

Cheong, R., Wang, C.J., Levchenko, A. Using a microfluidic device for high-content analysis of cell signaling. *Sci Signal*. **2**, pl2. (2009)

Chylek, L.A. Decoding the language of phosphorylation site dynamics. *Sci Signal*. **6**, jc2. (2013)

Chylek, L.A., Akimov, V., Dengjel, J., Rigbolt, K.T., Hu, B., Hlavacek, W.S., Blagoev, B. Phosphorylation site dynamics of early T-cell receptor signaling. *PLoS One*. **9**, e104240. (2014a)

Chylek, L.A., Harris, L.A., Tung, C.S., Faeder, J.R., Lopez, C.F., Hlavacek, W.S. Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. *Wiley Interdiscip Rev Syst Biol Med*. **6**, 13-36. (2014b)

Chylek, L.A., Holowka, D.A., Baird, B.A., Hlavacek, W.S. An Interaction Library for the FcεRI Signaling Network. *Front Immunol*. **5**, 172. (2014c)

Chylek, L.A., Hu, B., Blinov, M.L., Emonet, T., Faeder, J.R., Goldstein, B., Gutenkunst, R.N., Haugh, J.M., Lipniacki, T., Posner, R.G., Yang, J., Hlavacek, W.S. Guidelines for visualizing and annotating rule-based models. *Mol Biosyst*. **7**, 2779-2795. (2011)

Colvin, J., Monine, M.I., Faeder, J.R., Hlavacek, W.S., Von Hoff, D.D., Posner, R.G. Simulation of large-scale rule-based models. *Bioinformatics*. **25**, 910-917. (2009)

Colvin, J., Monine, M.I., Gutenkunst, R.N., Hlavacek, W.S., Von Hoff, D.D., Posner, R.G. RuleMonkey: software for stochastic simulation of rule-based models. *BMC Bioinformatics*. **11**, 404. (2010)

Cox, J., Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem*. **80**, 273-299. (2011)

Creamer, M.S., Stites, E.C., Aziz, M., Cahill, J.A., Tan, C.W., Berens, M.E., Han, H., Bussey, K.J., Von Hoff, D.D., Hlavacek, W.S., Posner, R.G. Specification, annotation, visualization and simulation of a large rule-based model for ERBB receptor signaling. *BMC Syst Biol*. **6**, 107. (2012)

Danos, V, Feret, J., Fontana, W., Harmer, R., Krivine, J. Rule-Based Modeling of Cellular Signaling. *Lecture Notes in Computer Science*, **4703**, 17-41 (2007)

Das, J., Ho, M., Zikherman, J., Govern, C., Yang, M., Weiss, A., Chakraborty, A.K., Roose, J.P. Digital signaling and hysteresis characterize ras activation in lymphoid cells. *Cell*. **136**, 337-351. (2009)

Das, R., Baird, E., Allen, S., Baird, B., Holowka, D., Goldstein, B. Binding mechanisms of PEGylated ligands reveal multiple effects of the PEG scaffold. *Biochemistry*, **47**, 1017-1030. (2008)

Dembo, M., Goldstein, B. Theory of equilibrium binding of symmetric bivalent haptens to cell surface antibody: application to histamine release from basophils. *J Immunol*. **121**, 345-53. (1978)

Erickson, J., Kane, P., Goldstein, B., Holowka, D., Baird, B. Cross-linking of IgE-receptor complexes at the cell surface: a fluorescence method for studying the binding of monovalent and bivalent haptens to IgE. *Mol Immunol*. **23**, 769-781. (1986)

Faeder, J.R, Hlavacek, W.S., Reischl, I., Blinov, M.L., Metzger, H., Redondo, A., Wofsy, C., Goldstein, B. Investigation of early events in Fc epsilon RI-mediated signaling using a detailed mathematical model. *J Immunol*. **170**, 3769-3781. (2003)

Faeder, J.R., Blinov, M.L., Hlavacek, W.S. Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol Biol*. **500**, 113-167. (2009)

Gillespie, D, T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J Chem Phys* **81**, 2340–2361. (1977)

Grädler, U, Schwarz, D., Dresing, V., Musil, D., Bomke, J., Frech, M., Greiner, H., Jäkel, S., Rysiok, T., Müller-Pompalla, D., Wegener, A. Structural and biophysical characterization of the Syk activation switch. *J Mol Biol*. **425**, 309-333. (2013)

Guo, X., Wang, X.F. Signaling cross-talk between TGF-beta/BMP and other pathways. *Cell Res*. **19**, 71-88. (2009)

Hause, R.J. Jr, Leung, K.K., Barkinge, J.L., Ciaccio, M.F., Chuu, C.P., Jones, R.B. Comprehensive binary interaction mapping of SH2 domains via fluorescence polarization reveals novel functional diversification of ErbB receptors. *PLoS One*. **7**, e44471. (2012)

Holowka, D., Gosse, J.A., Hammond, A.T., Han, X., Sengupta, P., Smith, N.L., Wagenknecht-Wiesner, A., Wu, M., Young, R.M., Baird, B. Lipid segregation and IgE receptor signaling: a decade of progress. *Biochim Biophys Acta*. **1746**, 252-259 (2005)

Karplus, M., McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*. **9**, 646-652. (2002)

Kraft, S., Kinet, J.P. New developments in FcepsilonRI regulation, function and inhibition. *Nat Rev Immunol*. **7**, 365-378. (2007)

Mahajan, A., Barua, D., Cutler, P., Lidke, D.S., Espinoza, F.A., Pehlke, C., Grattan, R., Kawakami, Y., Tung, C.S., Bradbury, A.R., Hlavacek, W.S., Wilson, B.S. Optimal aggregation of FcεRI with a structurally defined trivalent ligand overrides negative regulation driven by phosphatases. *ACS Chem Biol*. **9**, 1508-1519. (2014)

Markham, K., Bai, Y., Schmitt-Ulms, G. Co-immunoprecipitations revisited: an update on experimental concepts and their implementation for sensitive interactome investigations of endogenous proteins. *Anal Bioanal Chem*. **389**, 461-73 (2007)

McDonnell, J.M., Calvert, R., Beavil, R.L., Beavil, A.J., Henry, A.J., Sutton, B.J., Gould, H.J., Cowburn, D. The structure of the IgE Cepsilon2 domain and its role in stabilizing the complex with its high-affinity receptor FcεRIα. *Nat Struct Biol*. 8, 437-441. (2001)

Monine, M.I., Posner, R.G., Savage, P.B., Faeder, J.R., Hlavacek, W.S. Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. *Biophys J*. **98**, 48-56. (2010)

Paar, J.M., Harris, N.T., Holowka, D., Baird, B. Bivalent ligands with rigid double-stranded DNA spacers reveal structural constraints on signaling by FcεRI. *J Immunol*. **169**, 856-864. (2002)

Paensuwan, P., Ngoenkam, J., Khamsri, B., Preechanukul, K., Sanguansermsri, D., Pongcharoen, S. Evidence for inducible recruitment of Wiskott-Aldrich syndrome protein to T cell receptor-CD3 complex in Jurkat T cells. *Asian Pac J Allergy Immunol*. **33**, 189-195. (2015)

Piston, D.W., Kremers, G.J. Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem Sci*. **32**, 407-14. (2007)

Posner, R.G., Erickson, J.W., Holowka, D., Baird, B., Goldstein, B. Dissociation kinetics of bivalent ligand-immunoglobulin E aggregates in solution. *Biochemistry*. **30**, 2348-2356. (1991)

Posner, R.G., Geng, D., Haymore, S., Bogert, J., Pecht, I., Licht, A., Savage, P.B. Trivalent antigens for degranulation of mast cells. *Org Lett*. **9**, 3551-3554. (2007)

Posner, R.G., Subramanian, K., Goldstein, B., Thomas, J., Feder, T., Holowka, D., Baird, B. Simultaneous cross-linking by two nontriggering bivalent ligands causes synergistic signaling of IgE FcεRI complexes. *J Immunol*. **155**, 3601-369. (1995)

Shelby, S.A., Holowka, D., Baird, B., Veatch, S.L. Distinct stages of stimulated FcεRI receptor clustering and immobilization are identified through superresolution imaging. *Biophys J*. **105**, 2343-2354. (2013)

Sil, D., Lee, J.B., Luo, D., Holowka, D., Baird, B. Trivalent ligands with rigid DNA spacers reveal structural requirements for IgE receptor signaling in RBL mast cells. *ACS Chem Biol*. **2**, 674-284. (2007)

Sneddon, M.W., Faeder, J.R., Emonet, T. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat Methods*. **8**, 177-183. (2011)

Stites, E.C., Aziz, M., Creamer, M.S., Von Hoff, D.D., Posner, R.G., Hlavacek, W.S. Use of mechanistic models to integrate and analyze multiple proteomic datasets. *Biophys J*. **108**, 1819-1829. (2015)

Subramanian, K., Holowka, D., Baird, B., Goldstein, B. The Fc segment of IgE influences the kinetics of dissociation of a symmetrical bivalent ligand from cyclic dimeric complexes. *Biochemistry* **35**, 5518-5527.(1996)

Sun, C., Bernards, R. Feedback and redundancy in receptor tyrosine kinase signaling: relevance to cancer therapies. *Trends Biochem Sci*. **39**, 465-474. (2014)

Thomas, B.R., Chylek, L.A., Colvin, J., Sirimulla, S., Clayton, A.H., Hlavacek, W.S., Posner, R.G. BioNetFit: a fitting tool compatible with BioNetGen, NFsim, and distributed computing environments. *Bioinformatics* **pii**, btv655 (2015)

Wollman, R., Meyer, T. Coordinated oscillations in cortical actin and Ca2+ correlate with cycles of vesicle secretion. *Nat Cell Biol*. **14**, 1261-1269. (2012)

Weetall, M., Holowka, D., Baird, B. Heterologous desensitization of the high affinity receptor for IgE (FcεR1) on RBL cells. *J Immunol*. **150**, 4072-4083. (1993)

Young, R.M., Holowka, D., Baird, B. A lipid raft environment enhances Lyn kinase activity by protecting the active site tyrosine from dephosphorylation. *J Biol Chem*. **278**, 20746-20752. (2003)

Wofsy, C., Torigoe, C., Kent, U.M., Metzger, H., Goldstein, B. Exploiting the difference between intrinsic and extrinsic kinases: implications for regulation of signaling by immunoreceptors. *J Immunol*. **159**, 5984-5992. (1997)

Xu, K., Goldstein, B., Holowka, D., Baird, B. Kinetics of multivalent antigen DNP-BSA binding to IgE-Fc epsilon RI in relationship to the stimulated tyrosine phosphorylation of FcεRI. *J Immunol*. **160**, 3225-35. (1998)

Yang, J., Monine, M.I., Faeder, J.R., Hlavacek, W.S. Kinetic Monte Carlo method for rule-based modeling of biochemical networks. *Phys Rev E Stat Nonlin Soft Matter Phys*. **78**, 031910. (2008)

Chapter 1: An interaction library for FcεRI signaling[1]

Abstract

Antigen receptors play a central role in adaptive immune responses. Although the molecular networks associated with these receptors have been extensively studied, we currently lack a systems-level understanding of how combinations of non-covalent interactions and post-translational modifications are regulated during signaling to impact cellular decision-making. To fill this knowledge gap, it will be necessary to formalize and piece together information about individual molecular mechanisms to form large-scale computational models of signaling networks. To this end, we have developed an interaction library for signaling by the high-affinity IgE receptor, FcεRI. The library consists of executable rules for protein-protein and protein-lipid interactions. This library extends earlier models for FcεRI signaling and introduces new interactions that have not previously been considered in a model. Thus, this interaction library is a toolkit with which existing models can be expanded and from which new models can be built.

Introduction

To further our systems-level understanding of immunoreceptor signaling, we have developed a map and a rule library for early signaling mediated by FcεRI, which shares features with other related immunoreceptors. The FcεRI signaling system has a special feature of experimental tractability because the receptor can be stimulated using structurally defined antigens (Holowka et al., 2007, Paar et al., 2002, Sil et al., 2007), making it a valuable model system for

[1]This work was previously published in: Chylek LA, Holowka DA, Baird BA, Hlavacek WS. (2014) An interaction library for the FcεRI signaling network. *Frontiers in Immunology* **5**, 172.

understanding how signaling is initiated. Furthermore, FcεRI has been the subject of several past modeling studies that have elucidated early events following receptor crosslinking (Goldstein et al., 2002, Faeder et al., 2003), the flow of information during signaling (Faeder et al., 2005), aggregation of adaptor proteins (Nag et al., 2009, Nag et al., 2012), and the impact of ligand dose and binding kinetics on kinase activation (Nag et al., 2010a, Nag et al., 2010b). Aspects of the models used in these studies form a foundation for the rule library presented here. The library extends previous work by adding rules for interactions not previously included in models for FcεRI signaling. Thus, the library serves as a bridge between past studies of relatively small scope, and potential future studies that integrate information about more network elements to, for example, analyze multiplexed signaling data (Cao et al., 2007). As a first example of library use, we present simulations of recruitment of signaling proteins to the adaptor Lat, which is phosphorylated in response to FcεRI stimulation (Gilfillan et al., 2006).

Methods

We developed a library of rules based on known protein–protein and protein–lipid interactions, which were identified through a survey of the FcεRI literature. The rules can be assembled into different sets to form different models that capture the chemical kinetics of FcεRI signaling with site-specific resolution (Faeder et al., 2009). Here, the term "site" is used to refer to a generic functional site in a biomolecule, which in the case of a protein may be a domain, linear motif, or amino acid residue subject to post-translational modification. In a rule-based model, rules capture knowledge about biomolecular interactions. The rules in a model specify what interactions can occur in a system and under what conditions these interactions occur. A rule

provides necessary and sufficient conditions for testing its applicability, a definition of the consequences of an interaction, and a rate law. Rules, in combination with parameters and initial conditions, can be processed to simulate the behavior of a signaling system over time, including the time-dependent formation of protein complexes and post-translational modifications of proteins at specific sites. A benefit of a rule-based approach is that it enables concise specification and efficient simulation of models that include multivalent interactions and multi-site phosphorylation, which are two inherent characteristics of immunoreceptor signaling systems that are otherwise difficult or impossible to fully capture in a physicochemical model. We specified our library using a domain-specific language for rule-based modeling, the BioNetGen language (BNGL) (Faeder et al., 2009), which is compatible with several software tools for simulation and analysis.

Results and Discussion

In this section, we present a collection of rules, which can be viewed as a single model or as an assemblage of multiple models. Our main purpose is not to simulate the full set of interactions represented by these rules, but to formalize available knowledge about the FcεRI system to facilitate future modeling studies aimed at addressing specific questions.

Rule-based models are compositional, meaning that rules can be specified somewhat independently of each other, enabling construction of new models from components of existing models. We have taken advantage of this feature to build on three previously reported models: one for ligand–receptor interactions and two for intracellular signaling. Below, we briefly review these models and the processes that they capture. A visual overview of the intracellular processes captured in the library is provided in Figure 1.

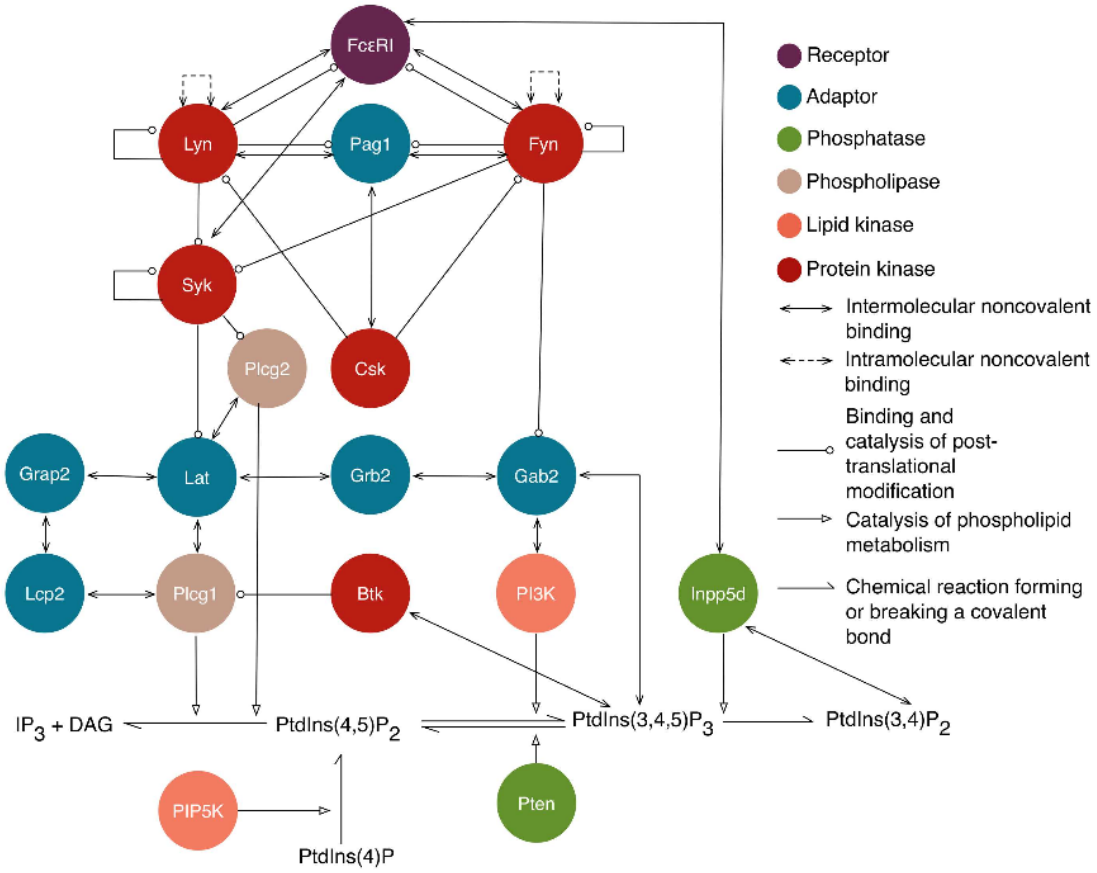


Figure 1.1. An overview of intracellular signaling interactions included in the model/library for FcεRI signaling. Rules are included in the library for the interactions depicted here. Proteins are represented as circles that are color-coded according to their function, as indicated in the legend. Standard UniProt names are used, and we note that Grap2 is commonly known as Gads, Lcp2 is commonly known as Slp76, and Inpp5d is commonly known as Ship1. The legend also indicates the arrows that are used to represent different types of interactions and influences. Reactions of lipid species are illustrated at the bottom. Arrows from proteins that point to lipid reactions indicate that the reaction is catalyzed by that protein. Arrows from protein to lipid species indicate that the protein binds that lipid. Not shown are implicit phosphatase reactions that cause dephosphorylation of all sites that can be phosphorylated. Ligand–receptor interactions are shown in Figure 2. A subset of interactions is illustrated with site-specific detail in Figure 3.

Initiation of signaling by FcεRI requires aggregation of receptors, which can be induced by reagents such as haptenated proteins and polymers, as well as by anti-receptor antibodies (Holowka et al., 2007). Several models have been developed to investigate the interactions that

lead to receptor aggregation. The model that we consider here is that of Xu et al. (Xu et al., 1998) for interactions of IgE-FcεRI with DNP–BSA, a multivalent antigen (haptenated protein). We chose this model because DNP–BSA is commonly used for stimulation of mast cells sensitized with anti-DNP IgE, and receptor aggregation induced by this antigen has been studied in detail (Shelby et al., 2013). In this model, the effective valence of the ligand was taken to be two. The model includes transient hapten exposure, initial binding of a ligand to a receptor, crosslinking of neighboring receptors, and dissociation of ligand–receptor bonds. In this model, it was assumed that receptor sites (antigen-combining sites in cell-surface IgE) are equivalent and that the single-site dissociation rate constant is the same for both ligand sites, regardless of whether the second site is bound or free. Cyclic aggregates are not considered. For use in this study, the model was translated from its original form to rules, which was also done in another recent study (Liu et al., 2013). The model of Xu et al. is illustrated in Figure 2.

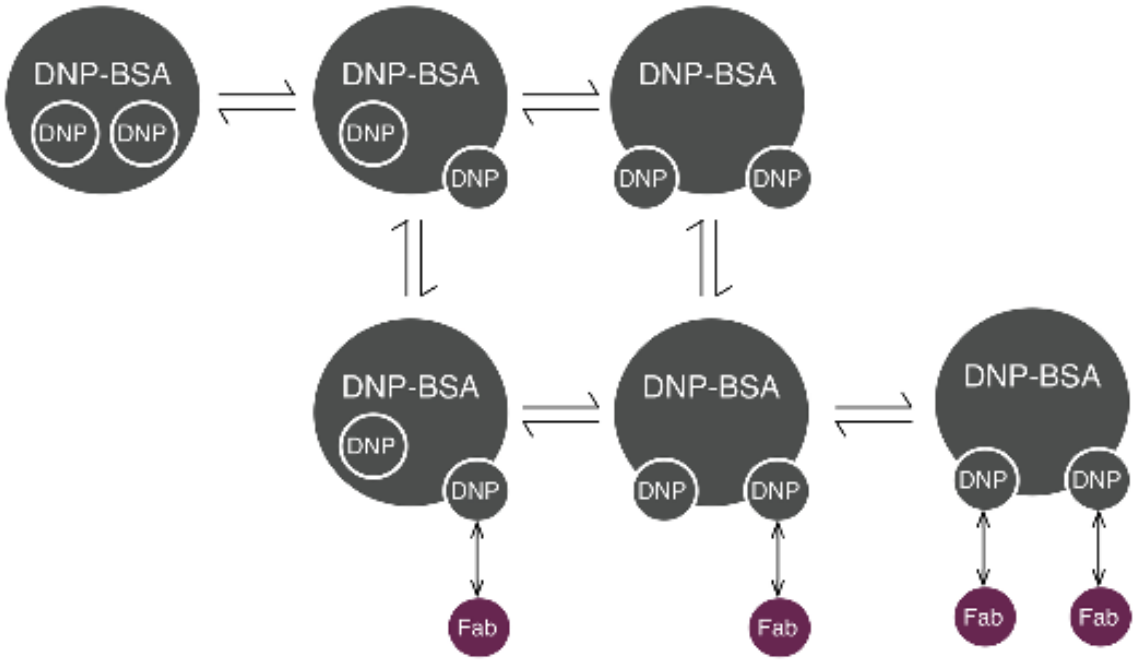


Figure 1.2. Reaction scheme for DNP–BSA interactions with cell- surface IgE. BSA (bovine serum albumin) is haptenated with multiple DNP groups, which are assumed to transition between two states: inaccessible (represented as being inside the molecule) and accessible (represented as being on the edge of the molecule). Accessible DNP can bind Fab arms of IgE. Each IgE antibody has two Fab arms, and is thus bivalent.

Receptor aggregation initiates signaling by bringing receptors into proximity with the Src-family kinase (SFK) Lyn. Lyn's association with receptors may be facilitated by several complementary mechanisms, including regulation by the membrane lipid environ- ment (Young et al., 2003) and constitutive direct binding to FcεRI via Lyn's unique N-terminal domain (Jouvin et al., 1994). For simplicity, we explicitly model the latter mechanism because it allows the plasma membrane to be treated as well-mixed and has been formalized in past modeling studies (Goldstein et al., 2002, Faeder et al., 2003). Lyn mediates phosphorylation of other receptors in an aggregate, thereby generating binding sites for the SH2 domain of Lyn. In this model, FcεRI constitutively associates with the unique N-terminal domain of Lyn. Crosslinking of receptors enables Lyn to transphosphorylate a second receptor at sites in the receptor's cytoplasmic subunits. These subunits, a β chain (Ms4a2) and a homodimer of two γ chains (Fcer1g), each contain an immunoreceptor tyrosine-based activation motif (ITAM). Each ITAM contains two (canonical) tyrosine residues that can be phosphorylated. The β chain contains an additional, non-canonical tyrosine in the middle of the ITAM sequence. In the original model, the tyrosines in the β chain were treated as a single-site, as were tyrosines in the γ chains. Here, we consider the β chain's N-terminal (canonical) and middle (non-canonical) tyrosines separately because they are capable of recruiting distinct binding partners. The phosphorylated N-terminal tyrosine recruits Lyn to aggregated receptors via SH2 domain binding, and enhances Lyn's catalytic activity by disruption of an inhibitory intramolecular bond, forming a positive

feedback loop. The non-canonical phosphotyrosine binds the lipid phosphatase Inpp5d (Ship1), which we will discuss below. The dually phosphorylated γ ITAM binds the tandem SH2 domains of the kinase Syk. Tyrosine residues in the linker region of Syk are phosphorylated by Lyn. Syk trans phosphorylates the activation loop in a second Syk molecule that is co-localized by being bound to cross-linked receptor, which constitutes positive feedback.

Rules for additional interactions among signaling proteins, which include mediators of negative regulation, were adapted from a model for BCR signaling (Barua et al., 2012). Lyn and a second SFK, Fyn, bind the transmembrane adaptor protein Pag1. Pag1 can then be phosphorylated by these kinases, generating additional binding sites for Lyn and Fyn, as well as for the kinase Csk. When co-localized on Pag1, Csk can phosphorylate Lyn and Fyn at an inhibitory C-terminal tyrosine. In this model, it was assumed that phosphorylation occurs in *cis*, meaning that Csk mediates phosphorylation of an SFK only when both are bound to the same Pag1 molecule. The C-terminal phosphotyrosine of an SFK forms an intramolecular bond with the SFK's SH2 domain, resulting in autoinhibition of the SFK's kinase domain.

The new rules of our library join the proximal signaling events described above to downstream processes that have not previously been considered in mechanistic models of FcεRI signaling. New rules are discussed in the sections that follow and are illustrated in Figure 3. The nomenclature and residue numbers used are consistent with UniProt conventions for rat proteins (http://www.uniprot.org/), because rat cells are commonly used in experimental studies of FcεRI signaling. If we view the rules of our library as constituting a single model, then the terminal output of the model is production of IP3, which is a second messenger. Binding of IP3 to its receptor on the endoplasmic reticulum leads to release of Ca2+ ions from intracellular stores, which is a key step for several processes in mast cell function, including degranulation and

chemotaxis (Holowka et al., 2012). Finally, we note that the interactions included in this library are not all unique to FcεRI signaling and are shared by pathways operative in TCR and BCR signaling. Thus, to facilitate identification of rules applicable to multiple pathways/cell types, in Table 1, we list protein–protein interactions included in the FcεRI library and whether each interaction is part of TCR and BCR signaling according to the NetPath database (Kandasamy et al., 2010).

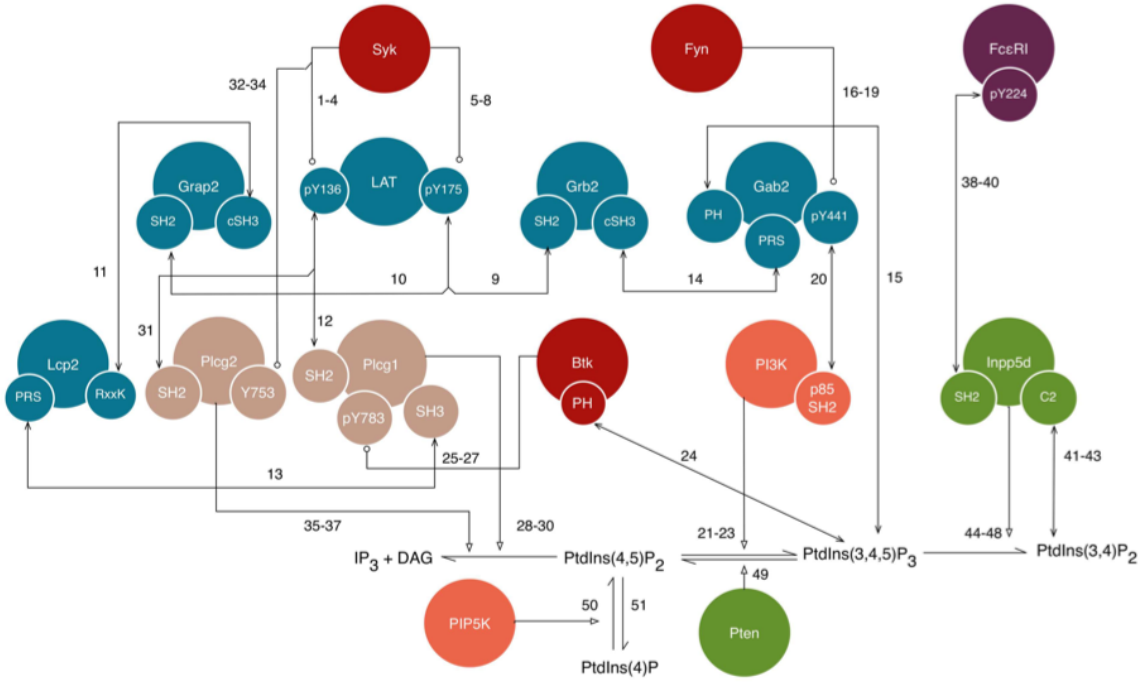


Figure 1.3. A detailed illustration of new interactions included in the model for FcεRI signaling. This diagram shows a subset of the interactions shown in Figure 1, but with illustration of additional details, namely the sites responsible for interactions. Conventions for color-coding and arrow symbols are the same as in Figure 1. Large circles represent proteins. Small circles, overlapping the edges of large circles, represent protein sites/components, such as domains, motifs, and amino acid residues. Standard UniProt names are used, and we note that Grap2 is commonly known as Gads, Lcp2 is commonly known as Slp76, and Inpp5d is commonly known as Ship1. Arrows represent interactions and are numbered to correspond to the numbering of rules given in the text. Phosphatase activity is considered implicitly as dephosphorylation reactions that apply to all sites, and is not illustrated in this figure.

Table 1.1: Shared components between FcεRI signaling and other immunoreceptor signaling sys- tems. Protein-protein interactions were classified as being involved (Y) or not involved (N) in BCR and TCR signaling on the basis of interactions listed for each system in the Net- Path database (http://www.netpath.org/). We use NetPath as an external standard. We note that "N" merely indicates that the interaction is not currently included in NetPath. This designation does not exclude the possibility of the interaction being part of BCR or TCR signaling.

FcεRI	BCR	TCR
FcεRI binds Lyn	N	N
FcεRI binds Fyn	N	N
FcεRI binds Syk	N	N
FcεRI binds Inpp5d	N	N
Lyn phosphorylates Lyn	N	N
Fyn phosphorylates Fyn	N	N
Syk phosphorylates Syk	Y	N
Fyn binds Pag1	N	Y
Lyn binds Pag1	N	N
Fyn phosphorylates Pag1	N	N
Lyn phosphorylates Pag1	N	N
Pag1 binds Csk	N	N
Csk phosphorylates Lyn	N	N
Csk phosphorylates Fyn	N	N
Syk phosphorylates Lat	N	N
Grb2 binds Lat	N	Y
Grap2 binds Lat	N	Y
Grap2 binds Lcp2	N	N
Plcg1 binds Lat	N	Y
Plcg1 binds Lcp2	Y	Y
Gab2 binds Grb2	N	N
Fyn phosphorylates Gab2	N	N
Gab2 binds PI3K	Y	N
Btk phosphorylates Plcg1	Y	N
Syk phosphorylates Plcg2	Y	N

Phosphorylation of Lat

Lat is a transmembrane, palmitoylated adaptor protein (Zhang et al., 1998a) that is involved in many signaling processes in both T cells and mast cells (Zhang et al., 1998b, Saitoh et al., 2000).

Syk phosphorylates Lat at multiple tyrosine residues (Zhang et al., 1998b), of which we focus on two: Y136 and Y175, which are better known as Y132 and Y191 in human Lat. Recent imaging studies suggest that Lat and the receptor become co-clustered after antigen-mediated receptor aggregation (Das et al., 2008, Veatch et al., 2012). However, it is not clear if Syk-mediated phosphorylation of Lat takes place within the context of a signaling complex that co-localizes Syk and Lat, or if instead, Syk-mediated phosphorylation of Lat takes place through random collisions between Syk's kinase domain and tyrosine substrates in Lat that generate short-lived enzyme–substrate complexes, as in a Michaelis–Menten mechanism. It has previously been assumed that the latter mechanism holds (Barua et al., 2012) and we follow this approach, using rules capturing enzyme–substrate binding, dissociation, and catalysis. For example, the rules listed below, which are written using the conventions of BNGL (Faeder et al., 2009), represent Syk-catalyzed phosphorylation of Y136 in Lat. Mass action kinetics are assumed. Bond indices are prefixed with the "!" symbol and internal state labels are prefixed with the "~" symbol. Here, internal state labels indicate whether a tyrosine residue is phosphorylated ("P") or unphosphorylated ("0").

```
(1) Syk(tSH2!+,PTK)+Lat(Y136~0) ->
Syk(tSH2! + ,PTK!1).Lat(Y136~0!1) kfSykLat
(2) Syk(PTK!1).Lat(Y136~0!1) ->
Syk(PTK)+ Lat(Y136~0) krSykLat
(3) Syk(PTK!1,Y519_Y520~P).Lat(Y136~0!1) ->
Syk(PTK,Y519_Y520~P) + Lat(Y136~P) kpSykLat136_1
(4) Syk(PTK!1,Y519_Y520~0).Lat(Y136~0!1) ->
Syk(PTK,Y519_Y520~0) + Lat(Y136~P) kpSykLat136_2
```

The first rule represents binding of Syk to Lat. In general, for rules in our library, protein components' names are consistent with terminology used in the biological literature. Here, the PTK component of Syk represents the protein tyrosine kinase domain of the protein. We assume that the interaction represented by Rule 1 only occurs when Syk is recruited to the plasma membrane, through binding of its tandem SH2 domains (tSH2) to phosphorylated FcεRI. Thus, the rule specifies that the tSH2 component must be bound for the reaction to occur (indicated by "!+"). The second rule represents the reverse reaction, which occurs independently of the binding state of Syk. Thus, the tSH2 component of Syk is not included in this rule. Rules 3 and 4 represent phosphorylation of Lat Y136 by Syk. These two rules differ in whether Syk is phosphorylated at its activation loop tyrosine residues Y519 and Y520, which are treated as a single site for simplicity. Phosphorylation of the activation loop enhances the catalytic activity of Syk (Zhang et al., 2000). Rate constants consistent with this regulatory mechanism are given after each rule, and are assigned values in the "parameters" block of the model specification (Listing 1). A similar set of rules is used to capture phosphorylation of Y175 in Lat.

```
(5) Syk(tSH2!+,PTK)+Lat(Y175~0) ->
Syk(tSH2! + ,PTK!1).Lat(Y175~0!1) kfSykLat
(6) Syk(PTK!1).Lat(Y175~0!1) ->
Syk(PTK)+ Lat(Y175~0) krSykLat
(7) Syk(PTK!1,Y519_Y520~P).Lat(Y175~0!1) ->
Syk(PTK,Y519_Y520~P) + Lat(Y175~P) kpSykLat175_2
(8) Syk(PTK!1,Y519_Y520~0).Lat(Y175~0!1) ->
Syk(PTK,Y519_Y520~0) + Lat(Y175~P) kpSykLat175_1
```

Interactions among Lat and its binding partners

Phosphorylated Y136 and Y175 have preferences for distinct binding partners, although crosstalk occurs between the pathways that branch from each site. Phosphorylated Y175 binds Grb2 and Grap2 (commonly known as Gads) (Houtman et al., 2004), which are two related cytosolic adaptor proteins that each contain an SH2 domain flanked by two SH3 domains (Jang et al., 2009). These adaptors are also able to bind other sites in Lat, with Grb2 being more promiscuous (Cho et al., 2004), but for simplicity we focus on Y175. The interactions of Lat pY175 with Grb2 and Grap2, which are taken to be mutually exclusive, are modeled as follows:

```
(9) Lat(Y175~P)+Grb2(SH2)<->
Lat(Y175~P!1).Grb2(SH2!1) kfLatGrb2, krLatGrb2
(10) Lat(Y175~P)+Grap2(SH2)<->
Lat(Y175~P!1).Grap2(SH2!1) kfLatGrap2, krLatGrap2
```

These rules are nearly as general as possible, in that minimal molecular context is included on the left-hand side of either of these rules (i.e., the only requirements for a bond to form is availability of the cognate binding sites in each molecule). For this reason, a large number of distinct reactions are implicitly defined by each rule. This feature is a generic aspect of rules and what allows for concise model specification.

Grap2 binds Lcp2, which is also known as Slp76. This high-affinity interaction occurs through the SH3 domain of Grap2 and an unconventional RxxK motif in Lcp2 (Seet et al., 2007).

```
(11) Grap2(SH3)+Lcp2(RxxK)<->
Grap2(SH3!1).Lcp2(RxxK!1) kfGrap2Lcp, krGrap2Lcp
```

Phosphorylated Y136 in Lat binds phospholipase Cγ1 (Plcg1) with high specificity (Houtman et al., 2004), and the interaction is modeled with the following rule:

```
(12) Lat(Y136~P) + Plcg1(SH2) <->
Lat(Y136~P!1).Plcg1(SH2!1) kfLatPlcg, krLatPlcg
```

Both of the tandem SH2 domains of Plcg1 contribute to colocalization of this enzyme with FcεRI upon stimulation (Staffer et al., 1997), and there is evidence both SH2 domains are capable of binding Lat (Samelson et al., 2002). However, for simplicity, we only consider a single SH2 domain in this model.

Plcg1 also interacts with Lcp2, via the SH3 domain of Plcg1 (Yablonski et al., 2001).

```
(13) Lcp2(PRS)+Plcg1(SH3)<->
Lcp2(PRS!1).Plcg1(SH3!1) kfLcp2Plcg1,krLcp2Plcg1
```

The final adaptor protein that we consider is Gab2. A linear motif in Gab2 can bind to the C-terminal SH3 domain of Grb2. We designate this motif as a proline-rich sequence (PRS), although its sequence differs from conventional SH3 binding motifs (Lewitsky et al., 2001).

```
(14)  Grb2(cSH3) + Gab2(PRS) <->
Grb2(cSH3!1). Gab2(PRS!1) kfGrb2Gab2,krGrb2Gab2
```

In addition, Gab2 can be recruited by binding of its PH domain to phosphatidylinositol 3,4,5-trisphosphate [PtdIns(3,4,5)P3], also abbreviated as PIP_3, in the plasma membrane (57).

```
(15) PI345P3(headgroup)+Gab2(PH)<->
PI345P3(headgroup!1).Gab2(PH!1) kfGab2Pip3,krGab2Pip3
```

The "headgroup" component in these rules represents the headgroup of the lipid, which is responsible for interactions with proteins.

Recruitment of PI3K to Gab2

PI3K association with Gab2 is dependent on Gab2 phosphorylation. Gab2 is phosphorylated by Fyn (Parravicini et al., 2002), which we assume catalyzes phosphorylation through a Michaelis–Menten mechanism.

```
(16) Fyn(U! + ,SH2,PTK) + Lat(Y175~P!1).
Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~0) ->
Fyn(U! + ,SH2,PTK!3).Lat(Y175~P!1). Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~0!3)
kfFynGab2
(17) Rec(b_Y210~P!4).Fyn(U,SH2!4,PTK) + Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).
Gab2(PRS!2,Y441~0)      -> Rec(b_Y210~P!4). Fyn(U,SH2!4,PTK!3).Lat(Y175~P!1).
Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~0!3) kfFynGab2
(18) Fyn(PTK!1).Gab2(Y441~0!1) ->
Fyn(PTK) + Gab2(Y441~0) krFynGab2
(19) Fyn(PTK!1).Gab2(Y441~0!1) ->
Fyn(PTK) + Gab2(Y441~P) kpFynGab2
```

The first two rules differ with respect to the mechanism by which Fyn is bound to a receptor. In the first rule, Fyn is taken to be bound by its unique domain (U). In the second rule, Fyn is taken to be bound by its SH2 domain.

Phosphorylated Gab2 binds the SH2 domain in the p85 subunit of PI3K (p85_SH2). Y441 of Gab2 lies in a consensus sequence for p85 binding (Gu et al., 2000).

```
(20) Gab2(Y441~P) + Pi3k(p85_SH2) <->
Gab2(Y441~P!1).Pi3k(p85_SH2!1) kfGab2Pi3k,krGab2Pi3k
```

PI3K Activity

Once recruited, PI3K phosphorylates the 3rd position in the inositol ring of phosphatidylinositol 4,5-bisphosphate [PtdIns(4,5)P2], also abbreviated as PIP_2, generating PIP_3.

```
(21) Lat(Y175~P!1).Grb2(SH2!1,cSH3!2). Gab2(PRS!2,Y441~P!3).Pi3k(p85_SH2!3,
PI3Kc)+PI45P2(headgroup) -> Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).
Gab2(PRS!2,Y441~P!3).Pi3k(p85_SH2!3, PI3Kc!4).PI45P2(headgroup!4) kfPi3kPip2
(22) Pi3k(PI3Kc!1).PI45P2(headgroup!1) ->
Pi3k(PI3Kc) + PI45P2(headgroup) krPi3kPip2
```

```
(23) Pi3k(PI3Kc!1).PI45P2(headgroup!1) ->
Pi3k(PI3Kc) + PI345P3(headgroup) kpPi3k DeleteMolecules
```

In these rules, lipid phosphorylation is treated as consumption and production of different lipid species. For this reason, the BNGL keyword "DeleteMolecules" is used to indicate removal of reactant molecules (Faeder et al., 2009).

Btk-mediated activation of Plcg1

PtdIns(3,4,5)P3 is a binding partner for multiple proteins, including the Tec-family kinase Btk, which is involved in activating Plcg1. The PH domain of Btk mediates this interaction.

```
(24) Btk(PH)+PI345P3(headgroup)<->
Btk(PH!1).PI345P3(headgroup!1) kfBtkPip3,krBtkPip3
```

Recruited Btk can phosphorylate Plcg1 at sites that are associated with enhancement of phospholipase activity (Qiu et al., 2000). In this way, pathways that branch from the two Lat phosphosites, Y136 and Y175, converge in contributing to IP3 production.

```
(25) Btk(PH! + ,PTK) + Plcg1(SH2! + ,Y783~0) -> Btk(PH! + ,PTK!1).Plcg1(SH2!
+ ,Y783~0!1) kfBtkPlcg
(26) Btk(PTK!1).Plcg1(Y783~0!1) Plcg1(Y783~0) krBtkPlcg (27)
Btk(PTK!1).Plcg1(Y783~0!1) Plcg1(Y783~P) kpBtkPlcg
```

Plcg1 activity

Plcg1 cleaves PtdIns(4,5)P_2 to generate the second messengers diacyl glycerol (DAG) and inositol 1,4,5-trisphosphate (IP3) (Oh-hora et al., 2008). The cleavage reaction is taken to occur through a Michaelis–Menten mechanism:

```
(28) Plcg1(SH2! + ,PLC) + PI45P2(headgroup) ->
Plcg1(SH2! + ,PLC!1).PI45P2(headgroup!1) kfPlcgPip2
(29) Plcg1(PLC!1).PI45P2(headgroup!1) - >
Plcg1(PLC) + PI45P2(headgroup) krPlcgPip2
```

```
(30) Plcg1(PLC!1,Y783~P).PI45P2(headgroup!1) ->
Plcg1(PLC,Y783~P) + IP3() + DAG() kcPlcg DeleteMolecules
```

Recruitment and activity of Plcg2

In addition to Plcg1, we also include Plcg2 in the library because isoform-specific differences between these two proteins have been found in FcεRI signaling. It has been observed that phosphorylation and activation of Plcg2 is less sensitive to PI3K inhibi- tion than Plcg1 (Barker et al., 1998). Thus, we include a mechanism by which Plcg2 is activated by Syk rather than by Btk. However, we note that other studies have found phosphorylation of Plcg2 to be reduced in the absence of Btk (Kawakami et al., 2000), suggesting that Btk may act on Plcg2.

```
(31) Lat(Y136~P) + Plcg2(SH2) < - >
Lat(Y136~P!1).Plcg2(SH2!1) kfLatPlcg, krLatPlcg
(32) Syk(tSH2! + ,PTK) + Plcg2(SH2! + ,Y753~0) - >
Syk(tSH2! + ,PTK!1).Plcg2(SH2! + , Y753~0!1) kfSykPlcg
(33) Syk(PTK!1).Plcg2(Y753~0!1) ->
Syk(PTK) + Plcg2(Y753~0) krSykPlcg
(34) Syk(PTK!1).Plcg2(Y753~0!1) -> Syk(PTK) + Plcg2(Y753~P) kpSykPlcg
(35) Plcg2(SH2! + ,PLC) + PI45P2(headgroup)- >
Plcg2(SH2!+,PLC!1).PI45P2(headgroup!1) kfPlcgPip2
(36) Plcg2(PLC!1).PI45P2(headgroup!1)- >
Plcg2(PLC) + PI45P2(headgroup)      krPlcgPip2
(37) Plcg2(PLC!1,Y753~P).PI45P2(headgroup!1) - >
Plcg2(PLC,Y753~P) + IP3() + DAG() kcPlcg DeleteMolecules
```

Rule 31 represents binding to Lat. Rules 32–34 represent phosphorylation of Plcg2 through a Michaelis–Menten mechanism. Rule 35–37 represent catalyzed hydrolysis of PIP_2.

Activation of Inpp5d

The final regulator of lipid signaling explicitly considered in our model is Inpp5d, also known as Ship1, a phosphatase that can be recruited to FcεRI by binding a non-canonical ITAM tyro- sine in the β subunit of the receptor (Kimura et al., 1997, On et al., 2004). Although Inpp5d and Lyn both bind the β subunit, they have preferences for different phosphotyrosines and thus we treat these interactions as non-competitive. Inpp5d dephosphorylates the 5th position of the inositol ring of PtdIns(3,4,5)P3 to form PtdIns(3,4)P2. This product of Inpp5d activity can in turn bind the Inpp5d C2 domain (Conde et al., 2011), forming a positive feedback loop that has an overall negative impact on FcεRI-mediated degranulation. The following rules are used to model binding of Inpp5d to the receptor:

```
(38) Inpp5d(SH2,C2) + Rec(b_Y224~P) ->
Inpp5d(SH2!1,C2).Rec(b_Y224~P!1) kfShipRec
(39) Inpp5d(IPP,C2! +) + Rec(b_Y224~P) ->
Inpp5d(IPP!1,C2! +).Rec(b_Y224~P!1) 100*kfShipRec
(40) Inpp5d(SH2!1).Rec(b_Y224~P!1)  ->
Inpp5d(SH2) + Rec(b_Y224~P)   krShipRec
```

In the first rule, Inpp5d is cytosolic, because its SH2 and C2 domains are both free and, in the model, these are the only domains that mediate membrane recruitment. In the second rule, Inpp5d is already membrane associated through binding of its C2 domain to PtdIns(3,4)P2. For this reason, receptor binding occurs more quickly (we assume a 100-fold enhancement). The third rule represents dissociation of Inpp5d from the receptor.

Binding of Inpp5d to PtdIns(3,4)P2 is modeled similarly, with different rules for membrane-recruited and cytosolic Inpp5d:

```
(41) Inpp5d(SH2,C2)+PI34P2(headgroup) -> Inpp5d(SH2,C2!1).PI34P2(headgroup!1)
kfShipPip2
(42) Inpp5d(SH2! + ,C2) + PI34P2(headgroup) - >
Inpp5d(SH2!+,C2!1).PI34P2(headgroup!1) 100*kfShipPip2
(43) Inpp5d(C2!1).PI34P2(headgroup!1) - >
Inpp5d(C2) + PI34P2(headgroup) krShipPip2
```

In the first rule, Inpp5d is cytosolic, whereas in the second rule, it is localized to the membrane through binding of its SH2 domain to the receptor. As above, a 100-fold enhancement is assumed. The third rule represents dissociation.

Inpp5d activity

The following rules capture the catalytic activity of Inpp5d:

```
(44) Inpp5d(SH2! + ,C2,IPP) + PI345P3 (headgroup) - >
Inpp5d(SH2! + ,C2,IPP!1). PI345P3(headgroup!1) kfShipPip3
(45) Inpp5d(SH2,C2! + ,IPP) + PI345P3 (headgroup) - >
Inpp5d(SH2,C2! + ,IPP!1). PI345P3(headgroup!1) kfShipPip3
(46) Inpp5d(SH2!+,C2!+,IPP)+PI345P3 (headgroup) - >
Inpp5d(SH2! + ,C2! + , IPP!1).PI345P3(headgroup!1) kfShipPip3
(47) Inpp5d(IPP!1).PI345P3(headgroup!1) ->
Inpp5d(IPP) + PI345P3(headgroup) krShipPip3
(48) Inpp5d(IPP!1).PI345P3(headgroup!1) -> Inpp5d(IPP) + PI34P2(headgroup)
kdpShipPip3 DeleteMolecules
```

In the rules above, "IPP" represents the catalytic domain of Inpp5d.

Additional lipid reactions

Conversion of PtdIns(3,4,5)P3 to PtdIns(4,5)P2 by Pten is considered implicitly as a first-order reaction. Conversions between PtdIns(4,5)P2 and PtdIns(4)P are modeled similarly.

```
(49) PI345P3(headgroup) - > PI45P2(headgroup) kPten DeleteMolecules
(50) PI4P(headgroup) -> PI45P2(headgroup) kfP5 DeleteMolecules
(51) PI45P2(headgroup) -> PI4P(headgroup) krP5 DeleteMolecules
```

Identification of network motifs

It has been hypothesized that relatively simple network motifs with specialized functions play important roles in cellular regulatory systems and that understanding the design principles of these motifs can help us better understand the complex systems in which they are embedded (Milo et al., 2002, Lim et al., 2013). Network motifs, such as feedback loops, have the potential to generate and/or regulate non-linear dynamical behavior (Tyson et al., 2010), which may, for example, enable precise encoding of information about a stimulus (Dolmetsch et al., 1998). We assessed the FcεRI signaling network for the presence of network motifs, and identified motifs from four classes: positive feedback loops, negative feedback loops, incoherent feed-forward loops, and coherent feed-forward loops. Several of the positive and negative feedbacks contribute to regulation of the SFKs Lyn and Fyn, as well as Syk. One positive feedback loop arises because SFKs phosphorylate tyrosine residues in FcεRI, which serve as binding sites that recruit additional Lyn and Fyn molecules. Furthermore, Lyn and Fyn can each transphosphorylate their own activation loop, which enhances catalytic activity. A similar mechanism also activates the kinase Syk. Negative feedback arises because Lyn and Fyn can phosphorylate the adaptor Pag1, which recruits Csk to negatively regulate SFK activity. This set of interactions has been predicted to lead to oscillations in BCR signaling (Barua et al., 2012).

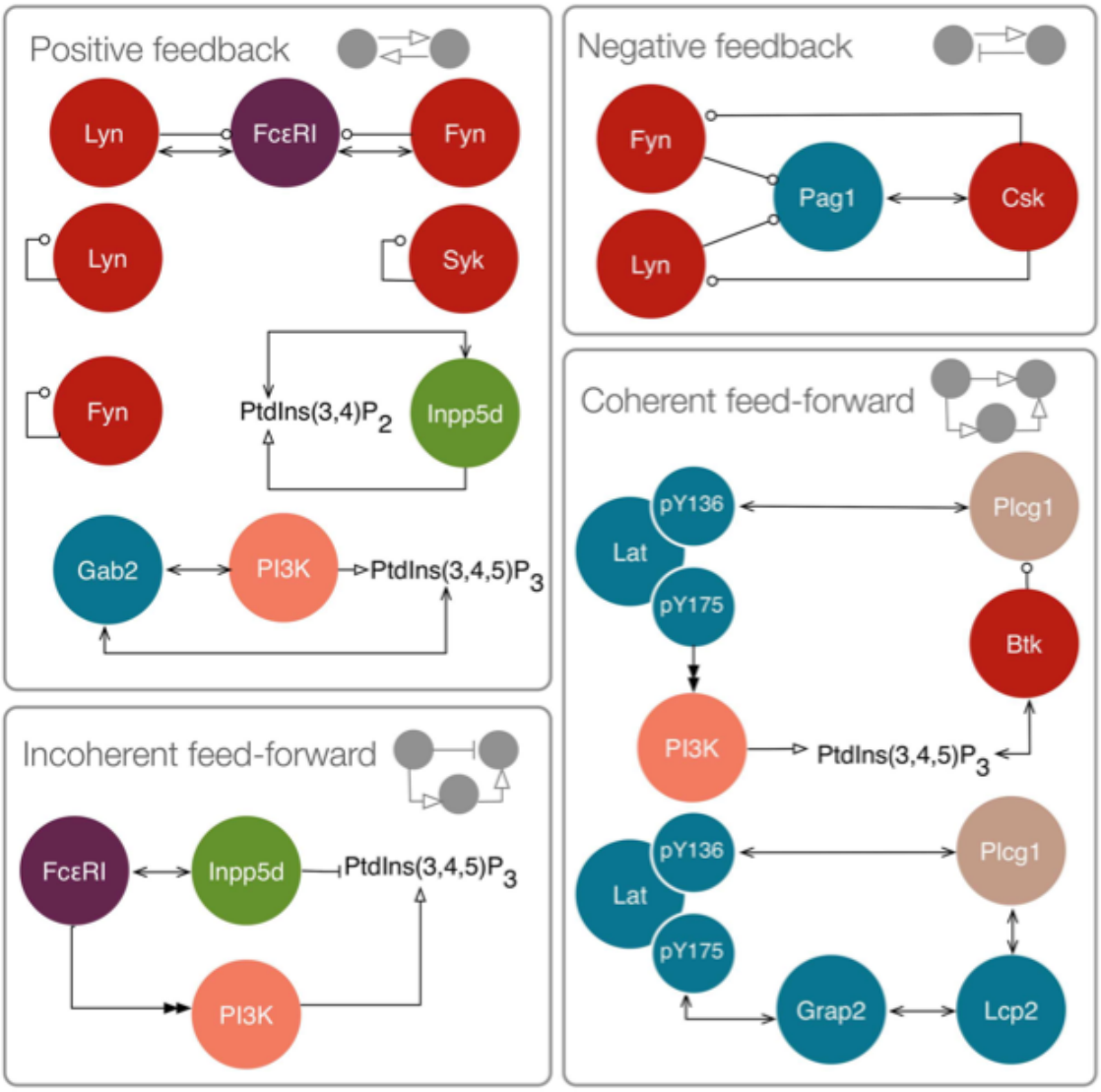


Figure 1.4. Motifs in the FcεRI signaling network. Positive feedbacks include interactions between FcεRI and Lyn and Fyn, because Lyn and Fyn catalyze phosphorylation of additional binding sites for these kinases. Lyn, Fyn, and Syk are subject to trans autophosphorylation at activating sites. Inpp5d binds its own product. Gab2 recruits PI3K, which generates PIP3, which can recruit additional Gab2. Negative feedback includes inhibition of Lyn and Fyn by Csk. Incoherent feed-forward includes FcεRI stimulation leading to activation of both PI3K and Inpp5d, which exert opposing influences on PIP3 level. Coherent feed-forwards include recruitment and activation of Plcg1, and recruitment of Plcg1 through two pathways.

Other positive feedback loops are involved in regulating lipid metabolism. PI3K generates PIP_3, which recruits Gab2. Gab2 can in turn recruit additional PI3K. An additional

positive feedback loop regulates Inpp5d, because it is capable of binding its own product. Inpp5d is also involved in an incoherent feed-forward loop, meaning a process in which two parallel mechanisms have opposite influences on an output. Here, the output is PIP3. Inpp5d is recruited to FcεRI and dephosphorylates PIP_3. Incoherence arises because FcεRI contributes to activation of PI3K, which generates PIP_3. In this way, opposing influences are exerted on the abundance of PIP_3 upon stimulation of FcεRI signaling. Such circuitry has been hypothesized to be involved in adaptation, the capacity of a system to respond to an input and then reset itself to a pre-stimulated state (Ma et al., 2009). Thus, PIP3 level may be raised and then lowered after a period of FcεRI stimulation, with Inpp5d-mediated positive feedback reinforcing negative regulation over time.

Finally, we identified a pair of coherent feed-forward loops stemming from the adaptor Lat. In a coherent feed-forward loop, two processes exert the same influence (either positive or negative) on an output. In each of the feed-forward loops of interest here, both processes in the network motif have a positive influence on Plcg1 activity. In the first feed-forward loop, Lat recruits Plcg1 via one of its phosphotyrosines. Other Lat phosphotyrosines are involved in assembly of a signaling complex that ultimately recruits PI3K. The product of PI3K, PIP_3, binds the kinase Btk, which phosphorylates Plcg1 at an activating site. In the second feed-forward loop, Lat contributes to Plcg1 recruitment through direct binding as well as through recruitment of another adaptor, Lcp2.

What function could be achieved by these (overlapping) feed-forward loops? In transcriptional regulatory networks, it has been found that feed-forward loops can act as sign-sensitive delay elements, meaning that they enable rapid responses to changes in an input in one direction, and slow responses to changes in the input in the opposite direction (Magan et al.,

2003a, Mangan et al., 2003b). Thus, the feed-forward loops initiated by Lat may influence the timing of Plcg1 activation and deactivation after increases or decreases in, for example, upstream receptor phosphorylation.

It is worth noting that Plcg1 and PI3K act on the same substrate, PIP_2. Thus, although PI3K can positively influence Plcg1, these two enzymes also compete with one another and could together deplete available PIP_2, assuming both access the same lipid pool. In this way, the feed-forward loop may be self-limiting. For example, if Plcg1 causes rapid conversion of PIP_2 to IP3, less PIP_2 would be available to PI3K and as a result, less PIP_3 would be generated and the impact of the feed-forward loop would be reduced. The strength of the feed-forward loop would also be influenced by the rate of production of PIP2 by specific lipid kinases and phosphatases. A caveat is that Plcg1 and PI3K may act on spatially distinct lipid pools, which PIP2 has been found to exist in (Calloway et al., 2011). These factors are not immediately evident from examination of isolated circuitry. This example highlights the importance of considering broader context and physical parameters (e.g., concentrations and binding affinities) in assessment of network motif functionality.

Listing 1 (below): BioNetGen Language encoding of the rule library

```
# References:
# Xu K, Goldstein B, Holowka D, Baird B. Kinetics of multivalent antigen DNP-
BSA binding to IgE-Fc epsilon RI in relationship to the stimulated tyrosine
phosphorylation of Fc epsilon RI. J Immunol. (1998) 160:3225-35.
# Faeder JR, Hlavacek WS, Reischl I, Blinov ML, Metzger H, Redondo A, Wofsy
C, Goldstein B. Investigation of early events in Fc epsilon RI-mediated
signaling using a detailed mathematical model. J Immunol. (2003) 170:3769-81.
# Barua D, Hlavacek WS, Lipniacki T. A computational model for early events
in B cell antigen receptor signaling: analysis of the roles of Lyn and Fyn. J
Immunol. (2012) 189:646-58.
# Lenoci L, Duvernay M, Satchell S, DiBenedetto E, Hamm HE. Mathematical
model of PAR1-mediated activation of human platelets. Mol Biosyst.7:1129-37.
(2011)


begin molecule types

# Residue numbers, where applicable, are given for rat proteins.

# The ligand is DNP-BSA and is taken to have two virtual haptens.
# Each hapten has two possible states, buried (b) or exposed (e).
Lig(hap~b~e,hap~b~e)

# The receptor is a complex of FceRI and an IgE antibody.
# The "fab" components represent the two fab arms of hapten-specific IgE.
# b_Y218 and b_Y224 are tyrosines in the ITAM of the beta subunit.
# b_Y224 is non-canonical, located between the two canonical tyrosines of the
ITAM.
# Y65 and Y76 are tyrosines in the ITAM of the gamma subunit and are lumped
as a single site.
# Only one (of the two) gamma chains is considered, as a simplification.
Rec(fab,fab,b_Y218~0~P,b_Y224~0~P,g_Y65_Y76~0~P)

# Lyn is an SFK with a unique domain (U), an SH2 domain, and an SH3 domain.
# Y397 is located in the kinase activation loop, and Y508 is the inhibitory
C-terminal tyrosine.
Lyn(U,SH3,SH2,Y397~0~P,Y508~0~P)

# Fyn is an SFK with a unique domain (U), an SH2 domain, and an SH3 domain.
# Y420 is located in the kinase activation loop, and Y531 is the inhibitory
C-terminal tyrosine.
Fyn(U,SH3,SH2,PTK,Y420~0~P,Y531~0~P)

# Syk is a kinase with tandem SH2 domains, which are lumped as a single site.
# Y346 is located in Interdomain B.
# Y519 and Y520 are located in the activation loop of the kinase domain.
These tyrosines are lumped as a single site.
Syk(tSH2,Y346~0~P,PTK,Y519_Y520~0~P)

# Pag1 is an adaptor protein with two proline-rich sequences (PRS1 and PRS2)
and multiple phosphosites.
```

```
# Y165 and Y183 are lumped as a single site, as are Y386 and Y409.
# PRS1 represents the PRIP motif starting at P136.
# PRS2 represents the PPVP motif starting at P255.
Pag1(PRS1,Y165_Y183~0~P,PRS2,Y317~0~P,Y386_Y409~0~P)

# Csk is a kinase with an SH2 domain.
Csk(SH2)

# Lat is an adaptor protein with multiple tyrosine residues that undergo
phosphorylation, of which we focus on Y136 and Y175
Lat(Y136~0~P,Y175~0~P)

# Phospholipase C gamma 1 (Plcg1)
Plcg1(SH2,SH3,PLC,Y783~0~P)

# Grb2 is an adaptor protein with an SH2 domain and two SH3 domains, of which
we consider the C-terminal domain.
Grb2(SH2,cSH3)

# Gab2 is an adaptor protein
Gab2(PRS,Y441~0~P)

# Grap2 is an adaptor protein
Grap2(SH2,SH3)

# Lcp2 is an adaptor protein
Lcp2(RxxK,PRS)

# PI 3-kinase is a heterodimeric lipid kinase that can phosphorylate the 3
position of an inositol ring.
# Its p85 subunit contains an SH2 domain.
Pi3k(p85_SH2,PI3Kc)

# Btk is a Tec family kinase
Btk(PH,PTK)

# Inpp5d is an inositol phosphatase. It is also known as Ship1.
# It contains an SH2 domain, a C2 domain that can bind lipid species, and an
inositol polyphosphate phosphatase catalytic domain (IPP)
Inpp5d(SH2,IPP,C2)

# Phosphatidylinositol 3,4-bisphosphate, or PtdIns(3,4)P2, is a phospholipid.
# The "bind" component is a virtual site for protein interactions.
PI34P2(bind)

# Phosphatidylinositol 4,5-bisphosphate, or PtdIns(4,5)P2, is a phospholipid.
# The "bind" component is a virtual site for protein interactions.
PI45P2(bind)

# Phosphatidylinositol 3,4,5-trisphosphate, or PtdIns(3,4,5)P3, is a
phospholipid.
# The "bind" component is a virtual site for protein interactions.
PI345P3(bind)

# Inositol 1,4,5-trisphosphate
IP3()
```

```
# Diacyl glycerol
DAG()

# Sink
Sink()

PI4P()
end molecule types


begin parameters
NA          6.022e23          # Avogadro's number; molecules/mole.
celldensity 1e9               # Cells/L
Fx                0.02              # Fraction of cell volume to simuate;
unitless.
ECFvol          1/(celldensity)           # Extracellular volume; L/cell.
simECFvol       ECFvol*Fx           # Simulated fraction of extracellular
volume; L.
Cellvol         1.4e-12             # Cytoplasmic volume; L. Value from
Faeder et al.
simCellvol  Cellvol*Fx        # Simulated fraction of cell volume; L.
ProteinTot  3e5               # Protein copy number per cell.
SimProteinTot     Fx*ProteinTot           # Simulated fraction of protein
copy number.
LigTot            5e4
SimLigTot   Fx*LigTot


# Exposure of hapten, forward rate constant. (Xu et al, Table I.)
lambda_p    1.7e-2      # /s

# Exposure of hapten, reverse rate constant.
# From lambda_p and ratio of forward and reverse rate constants. (Xu et al.)
lambda_m    5.4e-2 # /s

# Ligand binding from solution, forward rate constant (Xu et al.). Set in rnf
file.
kfl         0     # 8e6/(NA*simECFvol) /M/s

# Crosslinking by ligand, forward rate constant (Xu et al.).
kxl           1.3/SimProteinTot #/M/s

# Ligand binding, reverse rate constant. (Xu et al.)
krl         1.4e-1  #/s

# Lyn binding receptor through unique domain, forward rate constant (Faeder
et al.)
kfRecLyn1   4.2e7/(NA*simCellvol)  #/M/s

# Lyn binding receptor through unique domain, reverse rate constant (Faeder
et al.)
krRecLyn1   20 #/s

# Lyn binding through SH2 domain, forward rate constant (Faeder et al.)
kfRecLyn2   kfRecLyn1 #/M/s
```

```
# Lyn binding through SH2 domain, reverse rate constant (Faeder et al.)
krRecLyn2   0.12 #/s

# Receptor phosphorylation (beta chain) by Lyn (Faeder et al.)
kpLynB1           30 #/s

# Receptor phosphorylation (beta chain) by activated Lyn (Faeder et al.)
kpLynB2           100 #/s

# Receptor phosphorylation (gamma chain) by Lyn (Faeder et al.)
kpLynG1        1 #/s

# Receptor phosphorylation (gamma chain) by activated Lyn (Faeder et al.)
kpLynG2     3 #/s

# Autoinhibitory Lyn intramolecular bond, forward rate constant (Barua et
al.)
kfLynIn           10 #/s

# Autoinhibitory Lyn intramolecular bond, reverse rate constant (Barua et
al.)
krLynIn           3e-4 #/s

# Binding of Syk SH2 domains to dually-phosphorylated ITAM, forward rate
constant (Faeder et al.)
kfRecSyk    5.1e7/(NA*simCellvol)  #/M/s

# Binding of Syk SH2 domains to dually-phosphorylated ITAM, reverse rate
constant (Faeder et al.)
krRecSyk    0.13 #/s


kpSykSyk0   100
kpSykSykP   200
kpLynSyk1   30
kpLynSyk2   100
kfPagLynSH3 8.4e6/(NA*simCellvol)
krPagLynSH3 30
kfPagLynSH3_2     1e3
kfPagLynSH2 2.5e7/(NA*simCellvol)
kfPagLynSH2_2     1e3
krPagLyn2point    30
kpLynPag    1000
kfCskPag    2.5e4/(NA*simCellvol)
krCskPag    3e-3
kpCskLyn    1000
eff         5
kfRecFyn1   kfRecLyn1
krRecFyn1   krRecLyn1
kfRecFyn2   kfRecLyn2
krRecFyn2   krRecLyn2
kfFynIn           kfLynIn
krFynIn           krLynIn
kpFynB1           kpLynB1/eff
kpFynB2           kpLynB2/eff
```

```
kpFynG1           kpLynG1/eff
kpFynG2           kpLynG2/eff
kpFynSyk1   kpLynSyk1/eff
kpFynSyk2   kpLynSyk2/eff
kfPagFynSH3 kfPagLynSH3
krPagFynSH3 krPagLynSH3
kfPagFynSH3_2     kfPagLynSH3_2
kfPagFynSH2 kfPagLynSH2
kfPagFynSH2_2     kfPagLynSH2_2
krPagFyn2point    krPagLyn2point
kpFynPag    kpLynPag/eff
kpCskFyn    kpCskLyn

kfSykLat    kfRecLyn1
krSykLat    krRecLyn1
kpSykLat1   kpSykSyk0

kpSykLat2   kpSykSykP

KD_LatPlcg  62e-9
krLatPlcg   krRecLyn2
kfLatPlcg   (krLatPlcg/KD_LatPlcg)/(NA*simCellvol)


# Parameters for IP3 generation are taken from Lenoci et al (2011)
kfPlcgPip2  1e9/(NA*simCellvol)
krPlcgPip2  1
kcPlcgP     3.2e2
kcPlcg0     kcPlcgP/50


krLatGrb2   krRecLyn2
kfLatGrb2   2.5e8/(NA*simCellvol)

KD_Grb2Gab2 8e-6
krGrb2Gab2  krRecLyn2
kfGrb2Gab2  (krGrb2Gab2/KD_Grb2Gab2)/(NA*simCellvol)

kfFynGab2 kfSykLat
krFynGab2 krSykLat
kpFynGab2 kpFynB1

kfGab2Pi3k kfRecLyn2
krGab2Pi3k krRecLyn2

kfPi3kPip2  kfPlcgPip2
krPi3kPip2  krPlcgPip2
kpPi3k            kcPlcgP
kfBtkPip3   1.4e6/(NA*simCellvol)
krBtkPip3   1
kfBtkPlcg   kfSykLat
krBtkPlcg   krSykLat
kpBtkPlcg   kpSykSykP
kfShipRec   kfRecLyn2/5
krShipRec   krRecLyn2/5
kfShipPip3  kfPlcgPip2
```

```
krShipPip3  krPlcgPip2
kdpShipPip3       kcPlcgP
kfShipPip2        kfBtkPip3
krShipPip2  krBtkPip3
kfLatGrap2  6.6e6/(NA*simCellvol)
krLatGrap2  1

KD_Grb2Gab2 8e-6
krGrb2Gab2  1
kfGrb2Gab2  (1/KD_Grb2Gab2)/(NA*simCellvol)

kfGrap2Lcp2 9.5e6/(NA*simCellvol)
krGrap2Lcp2 0.06

KD_Lcp2Plcg1      1e-6
krLcp2Plcg1 1
kfLcp2Plcg1 (krLcp2Plcg1/KD_Lcp2Plcg1)/(NA*simCellvol)

kPten 1

kfP5  1
krP5 1
p           1e-3
dp          5
end parameters

begin seed species
Lig(hap~b,hap~b) SimLigTot
Rec(fab,fab,b_Y218~0,b_Y224~0,g_Y65_Y76~0) SimProteinTot
Lyn(U,SH2,SH3,Y397~0,Y508~0) SimProteinTot
Fyn(U,SH2,SH3,Y420~0,PTK,Y531~0) SimProteinTot
Syk(tSH2,Y346~0,PTK,Y519_Y520~0) SimProteinTot
Pag1(PRS1,PRS2,Y317~0,Y165_Y183~0,Y386_Y409~0) SimProteinTot
Csk(SH2) SimProteinTot
Lat(Y136~0,Y175~0) SimProteinTot
Plcg1(SH2,SH3,PLC,Y783~0) SimProteinTot
Grb2(SH2,cSH3) SimProteinTot
Gab2(PRS,Y441~0) SimProteinTot
PI45P2(bind) SimProteinTot
Pi3k(PI3Kc,p85_SH2) SimProteinTot
Btk(PH,PTK) SimProteinTot
Inpp5d(SH2,C2,IPP) SimProteinTot
Grap2(SH2,SH3) SimProteinTot
Lcp2(RxxK,PRS) SimProteinTot
PI4P()             SimProteinTot
Sink()             SimProteinTot
end seed species

begin reaction rules


# Haptens are transiently exposed.
Lig(hap~b) <-> Lig(hap~e) lambda_p,lambda_m

# Binding from solution when a hapten is exposed. Does not depend on state of
second hapten.
```

```
Lig(hap~e,hap~e) + Rec(fab) -> Lig(hap~e,hap~e!1).Rec(fab!1) kfl
Lig(hap~b,hap~e) + Rec(fab) -> Lig(hap~b,hap~e!1).Rec(fab!1) kfl

# Receptor crosslinking when both haptens are exposed.
Lig(hap~e,hap~e!1).Rec(fab!1) + Rec(fab) -> \
Lig(hap~e!2,hap~e!1).Rec(fab!1).Rec(fab!2) kxl

# Dissociation of hapten from receptor site
Lig(hap~e!1).Rec(fab!1) -> Lig(hap~e) + Rec(fab) krl

# Lyn unique domain binds receptor
Rec(b_Y218~0) + Lyn(U,SH3,SH2) -> \
Rec(b_Y218~0!1).Lyn(U!1,SH3,SH2) kfRecLyn1

# Lyn unique domain dissociates from receptor
Rec(b_Y218~0!1).Lyn(U!1) -> \
Rec(b_Y218~0) + Lyn(U) krRecLyn1

# Lyn SH2 domain binds receptor
Rec(b_Y218~P) + Lyn(U,SH3,SH2) -> \
Rec(b_Y218~P!1).Lyn(U,SH3,SH2!1) kfRecLyn2

# Lyn SH2 domain dissociates from receptor
Rec(b_Y218~P!1).Lyn(SH2!1) ->\
Rec(b_Y218~P) + Lyn(SH2) krRecLyn2

# Lyn SH2 domain binds Lyn pY508, forming an intramolecular bond
Lyn(U,SH3,SH2,Y508~P) -> \
Lyn(U,SH3,SH2!1,Y508~P!1) kfLynIn

# Lyn SH2 dissociates from Lyn pY508
Lyn(SH2!1,Y508~P!1) -> \
Lyn(SH2,Y508~P) krLynIn

## Lyn phosphorylates Y218 in the beta subunit of the receptor.
# Lyn bound by its unqiue domain
Lyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y218~0) -> \
Lyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y218~P)
kpLynB1

# Lyn bound by its SH2 domain
Lyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y218~0) ->
\
Lyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y218~P)
kpLynB2

## Lyn phosphorylates Y224 in the beta subunit of the receptor.
# Lyn bound by its unqiue domain
Lyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y224~0) ->\
Lyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y224~P)
kpLynB1

# Lyn bound by its SH2 domain
Lyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y224~0) ->\
Lyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y224~P)
kpLynB2
```

```
## Lyn phosphorylates Y65 and Y76 in the gamma subunit of the receptor.
# Lyn bound by its unqiue domain
Lyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~0) -
>\
Lyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P)
kpLynG1

# Lyn bound by its SH2 domain
Lyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~0)
->\
Lyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P)
kpLynG2

# Syk binds the dually phosphorylated gamma subunit of the receptor
Syk(tSH2) + Rec(g_Y65_Y76~P) <-> Syk(tSH2!1).Rec(g_Y65_Y76~P!1)
kfRecSyk,krRecSyk

## Lyn phosphorylates Syk
# Lyn bound by its unique domain
Lyn(U!1).Rec(fab!2,b_Y218~0!1).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P!4).
Syk(tSH2!4,Y346~0) ->\
Lyn(U!1).Rec(fab!2,b_Y218~0!1).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P!4).
Syk(tSH2!4,Y346~P)  kpLynSyk1

# Lyn bound by its SH2 domain
Lyn(SH2!1).Rec(fab!2,b_Y218~P!1).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P!4
).Syk(tSH2!4,Y346~0) ->\
Lyn(SH2!1).Rec(fab!2,b_Y218~P!1).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P!4
).Syk(tSH2!4,Y346~P)  kpLynSyk2

## Syk trans-phosphorylates the activation loop tyrosine of another Syk
molecule
# Syk unphosphorlyated on activation loop
Syk(tSH2!1,Y519_Y520~0).Rec(fab!2,g_Y65_Y76~P!1).Lig(hap~e!2,hap~e!3).Rec(fab
!3,g_Y65_Y76~P!4).Syk(tSH2!4,Y519_Y520~0) ->\
Syk(tSH2!1,Y519_Y520~0).Rec(fab!2,g_Y65_Y76~P!1).Lig(hap~e!2,hap~e!3).Rec(fab
!3,g_Y65_Y76~P!4).Syk(tSH2!4,Y519_Y520~P) kpSykSyk0

# Syk phosphorylated on activation loop
Syk(tSH2!1,Y519_Y520~P).Rec(fab!2,g_Y65_Y76~P!1).Lig(hap~e!2,hap~e!3).Rec(fab
!3,g_Y65_Y76~P!4).Syk(tSH2!4,Y519_Y520~0) ->\
Syk(tSH2!1,Y519_Y520~P).Rec(fab!2,g_Y65_Y76~P!1).Lig(hap~e!2,hap~e!3).Rec(fab
!3,g_Y65_Y76~P!4).Syk(tSH2!4,Y519_Y520~P) kpSykSykP

## SH3 domain of Lyn binds PRS2 in Pag1
# Association when Lyn is free
Lyn(U,SH3,SH2) + Pag1(PRS2,Y386_Y409) -> \
Lyn(U,SH3!1,SH2).Pag1(PRS2!1,Y386_Y409) kfPagLynSH3

# Association when Lyn is tethered to Pag1 by an SH2 domain-pY interaction
# Lyn, already tethered in Pag1 by SH2, binds Pag1 via SH3 domain
Lyn(U,SH3,SH2!2).Pag1(PRS2,Y386_Y409~P!2) -> \
Lyn(U,SH3!1,SH2!2).Pag1(PRS2!1,Y386_Y409~P!2) kfPagLynSH3_2

# Dissociation of Lyn SH3
```

```
Lyn(SH3!1,SH2).Pag1(PRS2!1,Y386_Y409) -> \
Lyn(SH3,SH2) + Pag1(PRS2,Y386_Y409)  krPagLynSH3


# SH2 domain of Lyn binds a pY docking site in Pag1
# Association when Lyn is free
Lyn(U,SH3,SH2) + Pag1(PRS2,Y386_Y409~P) -> \
Lyn(U,SH3,SH2!2).Pag1(PRS2,Y386_Y409~P!2) kfPagLynSH2

# Association when Lyn is tethered to Pag1 via an SH3 domain-PRS interaction
Lyn(U,SH3!1,SH2).Pag1(PRS2!1,Y386_Y409~P) -> \
Lyn(U,SH3!1,SH2!2).Pag1(PRS2!1,Y386_Y409~P!2) kfPagLynSH2_2

# Dissociation of Lyn from Pag1, breaking two-point attachment
# Because association of the Lyn SH3 domain with Pag1 PRS2 is rapid (relative
to the lifetime of a pY-SH2 bond) when Lyn is tethered by its SH2 domain,
# we omit dissociation of the SH2 domain alone, as a simplification.
Lyn(SH3!1,SH2!2).Pag1(PRS2!1,Y386_Y409~P!2) -> \
Lyn(SH3,SH2) + Pag1(PRS2,Y386_Y409~P)  krPagLyn2point

# Lyn phosphorylates Y386 and Y409 in Pag1
Lyn(SH3!1,Y508).Pag1(PRS2!1,Y386_Y409~0) -> \
Lyn(SH3!1,Y508).Pag1(PRS2!1,Y386_Y409~P) kpLynPag

# Lyn phosphorylates Y165 and Y183 in Pag1
Lyn(SH3,SH2!1,Y508).Pag1(Y165_Y183~0,Y386_Y409~P!1) -> \
Lyn(SH3,SH2!1,Y508).Pag1(Y165_Y183~P,Y386_Y409~P!1) kpLynPag

Lyn(SH3!2,SH2!1,Y508).Pag1(PRS2!2,Y165_Y183~0,Y386_Y409~P!1) -> \
Lyn(SH3!2,SH2!1,Y508).Pag1(PRS2!2,Y165_Y183~P,Y386_Y409~P!1) kpLynPag

Lyn(SH3!1,SH2,Y508).Pag1(PRS2!1,Y165_Y183~0) -> \
Lyn(SH3!1,SH2,Y508).Pag1(PRS2!1,Y165_Y183~P) kpLynPag

Lyn(SH2!1,Y508).Pag1(Y317~0,Y386_Y409~P!1) -> \
Lyn(SH2!1,Y508).Pag1(Y317~P,Y386_Y409~P!1) kpLynPag

# Lyn phosphorylates Y317 in Pag1
Lyn(SH3,SH2!1,Y508).Pag1(Y317~0,Y386_Y409~P!1) -> \
Lyn(SH3,SH2!1,Y508).Pag1(Y317~P,Y386_Y409~P!1) kpLynPag

Lyn(SH3!2,SH2!1,Y508).Pag1(PRS2!2,Y317~0,Y386_Y409~P!1) -> \
Lyn(SH3!2,SH2!1,Y508).Pag1(PRS2!2,Y317~P,Y386_Y409~P!1) kpLynPag

Lyn(SH3!1,SH2,Y508).Pag1(PRS2!1,Y317~0) -> \
Lyn(SH3!1,SH2,Y508).Pag1(PRS2!1,Y317~P) kpLynPag


# SH2 domain of Csk binds pY317 docking site in Pag1
Csk(SH2) + Pag1(Y317~P) <-> \
Csk(SH2!3).Pag1(Y317~P!3) kfCskPag,krCskPag

# Csk cis phosphorylates Y508 in Lyn
Lyn(SH3,SH2!1,Y508~0).Pag1(Y386_Y409~P!1,Y317~P!2).Csk(SH2!2) ->\
Lyn(SH3,SH2!1,Y508~P).Pag1(Y386_Y409~P!1,Y317~P!2).Csk(SH2!2) kpCskLyn
```

```
Lyn(SH3!1,SH2,Y508~0).Pag1(PRS2!1,Y317~P!2).Csk(SH2!2) ->\
Lyn(SH3!1,SH2,Y508~P).Pag1(PRS2!1,Y317~P!2).Csk(SH2!2) kpCskLyn

Lyn(SH3!1,SH2!3,Y508~0).Pag1(PRS2!1,Y386_Y409~P!3,Y317~P!2).Csk(SH2!2) ->\
Lyn(SH3!1,SH2!3,Y508~P).Pag1(PRS2!1,Y386_Y409~P!3,Y317~P!2).Csk(SH2!2)
kpCskLyn

# Fyn unique domain binds receptor
Rec(b_Y218~0) + Fyn(U,SH3,SH2) ->\
Rec(b_Y218~0!1).Fyn(U!1,SH3,SH2) kfRecFyn1

# Fyn unique domain dissociates from receptor
Rec(b_Y218~0!1).Fyn(U!1) ->\
Rec(b_Y218~0) + Fyn(U) krRecFyn1

# Fyn SH2 domain binds receptor
Rec(b_Y218~P) + Fyn(U,SH3,SH2) ->\
Rec(b_Y218~P!1).Fyn(U,SH3,SH2!1) kfRecFyn2

# Fyn SH2 domain dissociates from receptor
Rec(b_Y218~P!1).Fyn(SH2!1) ->\
Rec(b_Y218~P) + Fyn(SH2) krRecFyn2

# Fyn SH2 domain binds Fyn pY531, forming an intramolecular bond
Fyn(U,SH3,SH2,Y531~P) -> Fyn(U,SH3,SH2!1,Y531~P!1) kfFynIn

# Fyn SH2 domain dissociates from pY531
Fyn(SH2!1,Y531~P!1) ->\
Fyn(SH2,Y531~P) krFynIn

# Fyn phosphorylates Y218 in the beta subunit of the receptor
# Fyn bound by its unique domain
Fyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y218~0) ->\
Fyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y218~P)
kpFynB1

# Fyn bound by its SH2 domain
Fyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y218~0) ->
Fyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y218~P)
kpFynB2

# Fyn phosphorylates Y224 in the beta subunit of the receptor
# Fyn bound by its unique domain
Fyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y224~0) ->\
Fyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y224~P)
kpFynB1

# Fyn bound by its SH2 domain
Fyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y224~0) ->
Fyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,b_Y224~P)
kpFynB2

# Fyn phosphorylates Y65 and Y76 in the gamma subunit of the receptor
# Fyn bound by its unique domain
Fyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~0) -
>\
```

```
Fyn(U!1).Rec(b_Y218~0!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P)
kpFynG1

# Fyn bound by its SH2 domain
Fyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~0)
->\
Fyn(SH2!1).Rec(b_Y218~P!1,fab!2).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P)
kpFynG2

## Fyn phosphorylates Syk
# Fyn bound by its unique domain
Fyn(U!1).Rec(fab!2,b_Y218~0!1).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P!4).
Syk(tSH2!4,Y346~0) ->\
Fyn(U!1).Rec(fab!2,b_Y218~0!1).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P!4).
Syk(tSH2!4,Y346~P)  kpFynSyk1

# Fyn bound by its SH2 domain
Fyn(SH2!1).Rec(fab!2,b_Y218~P!1).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P!4
).Syk(tSH2!4,Y346~0) ->\
Fyn(SH2!1).Rec(fab!2,b_Y218~P!1).Lig(hap~e!2,hap~e!3).Rec(fab!3,g_Y65_Y76~P!4
).Syk(tSH2!4,Y346~P)  kpFynSyk2

## SH3 domain of Fyn binds PRS1 in Pag1
# Association when Fyn is free
Fyn(U,SH3,SH2) + Pag1(PRS1,Y165_Y183) -> \
Fyn(U,SH3!1,SH2).Pag1(PRS1!1,Y165_Y183) kfPagFynSH3

# Association when Fyn is tethered to Pag1 via an SH2 domain-pY interaction
Fyn(U,SH3,SH2!2).Pag1(PRS1,Y165_Y183~P!2) -> \
Fyn(U,SH3!1,SH2!2).Pag1(PRS1!1,Y165_Y183~P!2) kfPagFynSH3_2

# Dissociation of Fyn SH3
Fyn(SH3!1,SH2).Pag1(PRS1!1,Y165_Y183) -> \
Fyn(SH3,SH2) + Pag1(PRS1,Y165_Y183)  krPagFynSH3

## SH2 domain of Fyn binds a pY in Pag1
# Association when Fyn is free
Fyn(U,SH3,SH2) + Pag1(PRS1,Y165_Y183~P) -> \
Fyn(U,SH3,SH2!2).Pag1(PRS1,Y165_Y183~P!2) kfPagFynSH2

# Association when Fyn is tethered to Pag1 via SH3 domain-PRS interaction
Fyn(U,SH3!1,SH2).Pag1(PRS1!1,Y165_Y183~P) -> \
Fyn(U,SH3!1,SH2!2).Pag1(PRS1!1,Y165_Y183~P!2) kfPagFynSH2_2

# Dissociation of Fyn from Pag1, breaking two-point attachment
Fyn(SH3!1,SH2!2).Pag1(PRS1!1,Y165_Y183~P!2) -> \
Fyn(SH3,SH2) + Pag1(PRS1,Y165_Y183~P)  krPagFyn2point

# Fyn phosphorylates Y386 and Y409 in Pag1
Fyn(SH3,SH2!1,Y531).Pag1(Y386_Y409~0,Y165_Y183~P!1) -> \
Fyn(SH3,SH2!1,Y531).Pag1(Y386_Y409~P,Y165_Y183~P!1) kpFynPag

Fyn(SH3!2,SH2!1,Y531).Pag1(PRS1!2,Y386_Y409~0,Y165_Y183~P!1) -> \
Fyn(SH3!2,SH2!1,Y531).Pag1(PRS1!2,Y386_Y409~P,Y165_Y183~P!1) kpFynPag

Fyn(SH3!1,SH2,Y531).Pag1(PRS1!1,Y386_Y409~0) -> \
```

```
Fyn(SH3!1,SH2,Y531).Pag1(PRS1!1,Y386_Y409~P) kpFynPag

# Fyn phosphorylates Y317 in Pag1
Fyn(SH3,SH2!1,Y531).Pag1(Y317~0,Y165_Y183~P!1) -> \
Fyn(SH3,SH2!1,Y531).Pag1(Y317~P,Y165_Y183~P!1) kpFynPag

Fyn(SH3!2,SH2!1,Y531).Pag1(PRS1!2,Y317~0,Y165_Y183~P!1) -> \
Fyn(SH3!2,SH2!1,Y531).Pag1(PRS1!2,Y317~P,Y165_Y183~P!1) kpFynPag

Fyn(SH3!1,SH2,Y531).Pag1(PRS1!1,Y317~0) -> \
Fyn(SH3!1,SH2,Y531).Pag1(PRS1!1,Y317~P) kpFynPag

# Csk cis phosphorylates Y531 in Fyn
Fyn(SH3,SH2!1,Y531~0).Pag1(Y165_Y183~P!1,Y317~P!2).Csk(SH2!2) ->\
Fyn(SH3,SH2!1,Y531~P).Pag1(Y165_Y183~P!1,Y317~P!2).Csk(SH2!2) kpCskFyn

Fyn(SH3!1,SH2,Y531~0).Pag1(PRS1!1,Y317~P!2).Csk(SH2!2) ->\
Fyn(SH3!1,SH2,Y531~P).Pag1(PRS1!1,Y317~P!2).Csk(SH2!2) kpCskFyn

Fyn(SH3!1,SH2!3,Y531~0).Pag1(PRS1!1,Y165_Y183~P!3,Y317~P!2).Csk(SH2!2) ->\
Fyn(SH3!1,SH2!3,Y531~P).Pag1(PRS1!1,Y165_Y183~P!3,Y317~P!2).Csk(SH2!2)
kpCskFyn


# Syk phosphorylates Lat Y136
Syk(tSH2!+,PTK) + Lat(Y136~0) -> Syk(tSH2!+,PTK!1).Lat(Y136~0!1) kfSykLat
Syk(PTK!1).Lat(Y136~0!1) -> Syk(PTK) + Lat(Y136~0) krSykLat
Syk(PTK!1,Y519_Y520~P).Lat(Y136~0!1) -> Syk(PTK,Y519_Y520~P) + Lat(Y136~P)
kpSykLat2
Syk(PTK!1,Y519_Y520~0).Lat(Y136~0!1) -> Syk(PTK,Y519_Y520~0) + Lat(Y136~P)
kpSykLat1


# Lat pY136 binds Plcg1
Lat(Y136~P) + Plcg1(SH2) <-> Lat(Y136~P!1).Plcg1(SH2!1) kfLatPlcg,krLatPlcg

# Syk phosphorylates Lat Y175
Syk(tSH2!+,PTK) + Lat(Y175~0) -> Syk(tSH2!+,PTK!1).Lat(Y175~0!1) kfSykLat
Syk(PTK!1).Lat(Y175~0!1) -> Syk(PTK) + Lat(Y175~0) krSykLat
Syk(PTK!1,Y519_Y520~P).Lat(Y175~0!1) -> Syk(PTK,Y519_Y520~P) + Lat(Y175~P)
kpSykLat2
Syk(PTK!1,Y519_Y520~0).Lat(Y175~0!1) -> Syk(PTK,Y519_Y520~0) + Lat(Y175~P)
kpSykLat1

# Lat pY175 binds Grb2
Lat(Y175~P) + Grb2(SH2) <-> Lat(Y175~P!1).Grb2(SH2!1) kfLatGrb2,krLatGrb2

# Lat pY175 binds Grap2
Lat(Y175~P) + Grap2(SH2) <-> Lat(Y175~P!1).Grap2(SH2!1) kfLatGrap2,krLatGrap2

# Grap2 binds Lcp2
Grap2(SH3) + Lcp2(RxxK) <-> Grap2(SH3!1).Lcp2(RxxK!1) kfGrap2Lcp2,krGrap2Lcp2

# Grb2 binds Gab2
Grb2(cSH3) + Gab2(PRS) <-> Grb2(cSH3!1).Gab2(PRS!1) kfGrb2Gab2,krGrb2Gab2
```

```
# Lcp2 binds Plcg1
Lcp2(PRS) + Plcg1(SH3) <-> Lcp2(PRS!1).Plcg1(SH3!1) kfLcp2Plcg1,krLcp2Plcg1

# Fyn phosphorylates Gab2
Fyn(U!+,SH2,PTK) + Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~0) ->\
Fyn(U!+,SH2,PTK!3).Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~0!3)
kfFynGab2

Rec(b_Y218~P!4).Fyn(U,SH2!4,PTK) +
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~0) ->\
Rec(b_Y218~P!4).Fyn(U,SH2!4,PTK!3).Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!
2,Y441~0!3)  kfFynGab2

Fyn(PTK!1).Gab2(Y441~0!1) -> Fyn(PTK) + Gab2(Y441~0) krFynGab2

Fyn(PTK!1).Gab2(Y441~0!1) -> Fyn(PTK) + Gab2(Y441~P) kpFynGab2

# Gab2 binds PI 3-kinase
Gab2(Y441~P) + Pi3k(p85_SH2) <-> Gab2(Y441~P!1).Pi3k(p85_SH2!1)
kfGab2Pi3k,krGab2Pi3k

# PI 3-kinase generates PIP3
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~P!3).Pi3k(p85_SH2!3,PI3Kc) +
PI45P2(bind) ->\
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~P!3).Pi3k(p85_SH2!3,PI3Kc!4)
.PI45P2(bind!4) kfPi3kPip2

Pi3k(PI3Kc!1).PI45P2(bind!1) -> Pi3k(PI3Kc) + PI45P2(bind) krPi3kPip2

Pi3k(PI3Kc!1).PI45P2(bind!1) -> Pi3k(PI3Kc) + PI345P3(bind) + PI45P2(bind)
kpPi3k
Pi3k(PI3Kc!1).PI45P2(bind!1) -> Pi3k(PI3Kc) + PI345P3(bind) kpPi3k
DeleteMolecules


# Btk binds PIP3
Btk(PH) + PI345P3(bind) <-> Btk(PH!1).PI345P3(bind!1) kfBtkPip3,krBtkPip3

# Btk phosphorylates Plcg1
Btk(PH!+,PTK) + Plcg1(SH2!+,Y783~0) -> Btk(PH!+,PTK!1).Plcg1(SH2!+,Y783~0!1)
kfBtkPlcg
Btk(PTK!1).Plcg1(Y783~0!1) -> Btk(PTK) + Plcg1(Y783~0) krBtkPlcg
Btk(PTK!1).Plcg1(Y783~0!1) -> Btk(PTK) + Plcg1(Y783~P) kpBtkPlcg

# Inpp5d binds phosphorylated Y224 in the beta subunit of the receptor
Inpp5d(SH2,C2) + Rec(b_Y224~P) -> Inpp5d(SH2!1,C2).Rec(b_Y224~P!1) kfShipRec
Inpp5d(IPP,C2!+) + Rec(b_Y224~P) -> Inpp5d(IPP!1,C2!+).Rec(b_Y224~P!1)
100*kfShipRec
Inpp5d(SH2!1).Rec(b_Y224~P!1) -> Inpp5d(SH2) + Rec(b_Y224~P) krShipRec

Inpp5d(SH2!+,IPP) + PI345P3(bind) <-> Inpp5d(SH2!+,IPP!1).PI345P3(bind!1)
kfShipPip3,krShipPip3
Inpp5d(SH2!+,IPP!1).PI345P3(bind!1) -> Inpp5d(SH2!+,IPP) 0.01*kdpShipPip3
DeleteMolecules

PI345P3(bind) + Sink() -> Sink() 0.001*kdpShipPip3 DeleteMolecules
```

```
# Other lipid reactions
PI345P3(bind) -> PI45P2(bind) kPten DeleteMolecules
PI4P() -> PI45P2(bind) kfP5 DeleteMolecules
PI45P2(bind) -> PI4P() krP5 DeleteMolecules
# Nonspecific phosphorylation and dephosphorylation

Rec(b_Y218~0) -> Rec(b_Y218~P) p
Rec(b_Y218~P) -> Rec(b_Y218~0) dp

Rec(b_Y224~0) -> Rec(b_Y224~P) p
Rec(b_Y224~P) -> Rec(b_Y224~0) dp

Rec(g_Y65_Y76~0) -> Rec(g_Y65_Y76~P) p
Rec(g_Y65_Y76~P) -> Rec(g_Y65_Y76~0) dp

Lyn(Y508~0) -> Lyn(Y508~P) p
Lyn(Y508~P) -> Lyn(Y508~0) dp

Fyn(Y531~0) -> Fyn(Y531~P) p
Fyn(Y531~P) -> Fyn(Y531~0) dp

Syk(Y346~0) -> Syk(Y346~P) p
Syk(Y346~P) -> Syk(Y346~0) dp

Syk(Y519_Y520~0) -> Syk(Y519_Y520~P) p
Syk(Y519_Y520~P) -> Syk(Y519_Y520~0) dp

Pag1(Y317~0) -> Pag1(Y317~P) p
Pag1(Y317~P) -> Pag1(Y317~0) dp

Pag1(Y165_Y183~0) -> Pag1(Y165_Y183~P) p
Pag1(Y165_Y183~P) -> Pag1(Y165_Y183~0) dp

Pag1(Y386_Y409~0) -> Pag1(Y386_Y409~P) p
Pag1(Y386_Y409~P) -> Pag1(Y386_Y409~0) dp

Gab2(Y441~0) -> Gab2(Y441~P) p
Gab2(Y441~P) -> Gab2(Y441~0) dp

Lat(Y136~0) -> Lat(Y136~P) p
Lat(Y136~P) -> Lat(Y136~0) dp

Lat(Y175~0) -> Lat(Y175~P) p
Lat(Y175~P) -> Lat(Y175~0) dp

Plcg1(Y783~0) -> Plcg1(Y783~P) p
Plcg1(Y783~P) -> Plcg1(Y783~0) dp


end reaction rules

begin observables
Molecules Lat_pY136 Lat(Y136~P!?)
Molecules Lat_pY175 Lat(Y175~P!?)
Molecules Rec_pY218 Rec(b_Y218~P!?)
```

```
Molecules Rec_pY224 Rec(b_Y224~P!?)
Molecules Rec_pY65_Y76 Rec(g_Y65_Y76~P!?)
Molecules Syk_pY346 Syk(Y346~P!?)
Molecules Gab2_pY441 Gab2(Y441~P!?)
Molecules PI3K_recruited
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~P!3).Pi3k(p85_SH2!3)
Molecules PI345P3 PI345P3(bind!?)

Molecules Plcg1_pY783   Plcg1(Y783~P!?)
Molecules LynUbound Lyn(U!+)
Molecules LynSH2bound   Lyn(SH2!+)
Molecules Shipbound Inpp5d(SH2!+)
Molecules Plc_pip2 Plcg1(PLC!1).PI45P2(bind!1)
Molecules Lat_Plc Lat(Y136~P!1).Plcg1(SH2!1)


Molecules IP3     IP3()
Molecules PI34P2  PI34P2(bind!?)
Molecules PI45P2  PI45P2(bind!?)

Molecules Pag_Y165 Pag1(Y165_Y183~P!?)
Molecules Pag_Y409 Pag1(Y386_Y409~P!?)
Molecules LynIn     Lyn(SH2!1,Y508~P!1)
Molecules FynIn Fyn(SH2!1,Y531~P!1)
Molecules Lyn_pY508 Lyn(Y508~P!?)
Molecules Fyn_pY508 Fyn(Y531~P!?)

Molecules Btk_recruited Btk(PH!+)
# Free ligand
Molecules LigFree  Lig(hap,hap)

# Crosslinked receptor dimer
Molecules Crosslinks Lig(hap~e!1,hap~e!2).Rec(fab!1).Rec(fab!2)

# Ligand with both haptens buried
Molecules DNP_exposed_0_of_2     Lig(hap~b!?,hap~b!?)

# Ligand with one hapten exposed
Molecules  DNP_exposed_1_of_2    Lig(hap~b!?,hap~e!?)

# Ligand with both haptens exposed
Molecules  DNP_exposed_2_of_2    Lig(hap~e!?,hap~e!?)

# Total ligand
Molecules LigTotal Lig(hap!?,hap!?)

Molecules Syk_pY519_520 Syk(Y519_Y520~P)

end observables

writeXML();
```

References

Barker S, Caldwell K, Pfeiffer J, Wilson B. Wortmannin-sensitive phosphory- lation, translocation, and activation of PLCγ1, but not PLCγ2, in antigen- stimulated RBL-2H3 mast cells. Mol Biol Cell (1998) 9:483–96. doi:10.1091/ mbc.9.2.483

Barua D, Goldstein B. A mechanistic model of early FcεRI signaling: lipid rafts and the question of protection from dephosphorylation. PLoS One (2012) 7:e51669. doi:10.1371/journal.pone.0051669

Barua D, Hlavacek WS, Lipniacki T. A computational model for early events in B cell antigen receptor signaling: analysis of the roles of Lyn and Fyn. J Immunol (2012) 189:646–58. doi:10.4049/jimmunol.1102003

Calloway N, Owens T, Corwith K, Rodgers W, Holowka D, Baird B. Stimulated association of STIM1 and Orai1 is regulated by the balance of Ptdins(4,5)P2 between distinct membrane pools. J Cell Sci (2011) 124:2602–10. doi:10.1242/ jcs.084178

Cao L, Yu K, Banh C, Nguyen V, Ritz A, Raphael B, et al. Quantitative time- resolved phosphoproteomic analysis of mast cell signaling. J Immunol (2007) 179:5864–76.

Cho S, Velikovsky C, Swaminathan CP, Houtman JC, Samelson LE, Mariuzza RA. Structural basis for differential recognition of tyrosine-phosphorylated sites in the linker for activation of T cells (LAT) by the adaptor gads. EMBO J (2004) 23:1441–51. doi:10.1038/sj.emboj.7600168

Chylek LA, Stites EC, Posner RG, Hlavacek WS. Innovations of the rule-based modeling approach. In: Prokop A, Csukás B, editors. Systems Biology: Integrative Biology and Simulation Tools (Vol. 1). Dordrecht: Springer (2013). p. 273–300.

Condé C, Gloire G, Piette J. Enzymatic and non-enzymatic activities of SHIP- 1 in signal transduction and cancer. Biochem Pharmacol (2011) 82:1320–34. doi:10.1016/j.bcp.2011.05.031

Consortium TU. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res (2014) 42:D191–8. doi:10.1093/nar/gkt1140

Das R, Hammond S, Holowka D, Baird B. Real-time cross-correlation image analysis of early events in IgE receptor signaling. Biophys J (2008) 94:4996–5008. doi:10.1529/biophysj.107.105502

Dolmetsch R, Xu K, Lewis R. Calcium oscillations increase the efficiency and specificity of gene expression. Nature (1998) 392:933–6. doi:10.1038/31960 71. Ma W, Trusina A, El-Samad H, Lim W, Tang C. Defining network topologies that can achieve biochemical adaptation. Cell (2009) 138:760–73. doi:10.1016/j.cell.2009.06.013 72. Mangan S, Alon U. Structure and function of the feed-forward loop net-work motif. Proc Natl Acad Sci U S A (2003) 100:11980–5. doi:10.1073/pnas.2133841100

Edmead CE, Fox BC, Stace C, Ktistakis N, Welham MJ. The pleckstrin homol- ogy domain of Gab-2 is required for optimal interleukin-3 signalsome-mediated responses. Cell Signal (2006) 18:1147–55. doi:10.1016/j.cellsig.2005.09.002

Faeder JR, Blinov ML, Goldstein B, Hlavacek WS. Combinatorial complexity and dynamical restriction of network flows in signal transduction. Syst Biol (Stevenage) (2005) 2:5–15. doi:10.1049/sb:20045031

Faeder JR, Blinov ML, Hlavacek WS. Rule-based modeling of biochemical sys- tems with BioNetGen. Methods Mol Biol (2009) 500:113–67. doi:10.1007/978- 1- 59745- 525- 1_5

Faeder JR, Hlavacek WS, Reischl I, Blinov ML, Metzger H, Redondo A, et al. Investigation of early events in FcεRI-mediated signaling using a detailed math- ematical model. J Immunol (2003) 170:3769–81.

Gilfillan A, Tkaczyk C. Integrated signalling pathways for mast-cell activation. Nat Rev Immunol (2006) 16:218–30. doi:10.1038/nri1782

GoldsteinB,FaederJR,HlavacekWS,BlinovML,RedondoA,WofsyC.Modeling the early signaling events mediated by FcεRI. Mol Immunol (2002) 38:1213–9. doi:10.1016/S0161-5890(02)00066- 4

Gu H, Maeda H, Moon JJ, Lord JD, Yoakim M, Nelson BH, et al. New role for Shc in activation of the phosphatidylinositol 3-kinase/Akt pathway. Mol Cell Biol (2000) 20:7109–20. doi:10.1128/MCB.20.19.7109- 7120.2000

Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, Shumaker DE, et al. Sundials: suite of nonlinear and differential/algebraic equation solvers. ACM Trans Math Softw (2005) 31:363–96. doi:10.1145/1089014.1089020

Holowka D, Calloway N, Cohen R, Gadi D, Lee J, Smith NL, et al. Roles for Ca2+ mobilization and its regulation in mast cell functions. Front Immunol (2012) 3:104. doi:10.3389/fimmu.2012.00104

Holowka D, Sil D, Torigoe C, Baird B. Insights into immunoglobulin E receptor signaling from structurally defined ligands. Immunol Rev (2007) 217:269–79. doi:10.1111/j.1600-065X.2007.00517.x

Houtman JC, Higashimoto Y, Dimasi N, Cho S, Yamaguchi H, Bowden B, et al. Binding specificity of multiprotein signaling complexes is determined by both cooperative interactions and affinity preferences. Biochemistry (2004) 43:4170–8. doi:10.1021/bi0357311

Jang IK, Zhang J, Gu H. Grb2, a simple adapter with complex roles in lym- phocyte development, function, and signaling. Immunol Rev (2009) 232:150–9. doi:10.1111/j.1600-065X.2009.00842.x

Jouvin MH, Adamczewski M, Numerof R, Letourneur O, Vallé A, Kinet JP. Differential control of the tyrosine kinases Lyn and Syk by the two signaling chains of the high affinity immunoglobulin E receptor. J Biol Chem (1994) 269: 5918–25.

Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GS, Venugopal AK, et al. NetPath: a public resource of curated signal transduction pathways. Genome Biol (2010) 11:R3. doi:10.1186/gb- 2010- 11- 1- r3

Kawakami Y, Kitaura J, Satterthwaite AB, Kato RM, Asai K, Hartman SE, et al. Redundant and opposing functions of two tyrosine kinases, Btk and Lyn, in mast cell activation. J Immunol (2000) 165:1210–9.

Kimura T, Sakamoto H, Appella E, Siraganian RP. The negative signaling mol- ecule SH2 domain-containing inositol-polyphosphate 5-phosphatase (SHIP) binds to the tyrosine-phosphorylated beta subunit of the high affinity IgE recep- tor. J Biol Chem (1997) 272:13991–6. doi:10.1074/jbc.272.21.13991

Lewitzky M, Kardinal C, Gehring NH, Schmidt EK, Konkol B, Eulitz M, et al. The C-terminal SH3 domain of the adapter protein Grb2 binds with high affinity to sequences in Gab1 and SLP-76 which lack the SH3-typical P-x-x-P core motif. Oncogene (2001) 20:1052–62. doi:10.1038/sj.onc.1204202

Lim WA, Lee C, Tang C. Design principles of regulatory networks: search- ing for the molecular algorithms of the cell. Mol Cell (2013) 49:202–12. doi:10.1016/j.molcel.2012.12.020

Liu Y, Barua D, Liu P, Wilson BS, Oliver JM, Hlavacek WS, et al. Single-cell mea- surements of IgE-mediated FcεRI signaling using an integrated microfluidic platform. PLoS One (2013) 8:e60159. doi:10.1371/journal.pone.0060159

Mangan S, Zaslaver A, Alon U. The coherent feedforward loop serves asa sign-sensitive delay element in transcription networks. J Mol Biol (2003)334:197–204. doi:10.1016/j.jmb.2003.09.049

mast cell degranulation. Nat Immunol (2002) 3:741–8. doi:10.1038/ni817

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science (2002) 298:842–7. doi:10.1126/science.298.5594.824

Nag A, Faeder JR, Goldstein B. Shaping the response: the role of FcεRI and Syk expression levels in mast cell signaling. IET Syst Biol (2010) 33(4–47):334–47. doi:10.1049/iet-syb.2010.0006

Nag A, Monine M, Perelson A, Goldstein B. Modeling and simulation of aggre- gation of membrane protein LAT with molecular variability in the number of binding sites for cytosolic Grb2-SOS1-Grb2. PLoS One (2012) 7:e28758. doi:10.1371/journal.pone.0028758

Nag A, Monine MI, Blinov ML, Goldstein B. A detailed mathematical model predicts that serial engagement of IgE-Fc epsilon RI complexes can enhance Syk activation in mast cells. J Immunol (2010) 185:3268–76. doi:10.4049/jimmunol. 1000326

Nag A, Monine MI, Faeder JR, Goldstein B. Aggregation of membrane proteins by cytosolic cross-linkers: theory and simulation of the LAT-Grb2-SOS1 system. Biophys J (2009) 96:2604–23. doi:10.1016/j.bpj.2009.01.019

Oh-hora M, Rao A. Calcium signaling in lymphocytes. Curr Opin Immunol (2008) 20:250–8. doi:10.1016/j.coi.2008.04.004

On M, Billingsley JM, Jouvin MH, Kinet JP. Molecular dissection of the FcRβ signaling amplifier. J Biol Chem (2004) 279:45782–90. doi:10.1074/jbc. M404890200

PaarJ,HarrisN,HolowkaD,BairdB.Bivalentligandswithrigiddouble-stranded DNA spacers reveal structural constraints on signaling by FcεRI. J Immunol (2002) 169:856–64.

Parravicini V, Gadina M, Kovarova M, Odom S, Gonzalez-Espinosa C, Furumoto

Posner R, Geng D, Haymore S, Bogert J, Pecht I, Licht A, et al. Trivalent anti- gens for degranulation of mast cells. Org Lett (2007) 9:3551–4. doi:10.1021/ ol071175h

Qiu Y, Kung HJ. Signaling network of the Btk family kinases. Oncogene (2000) 19:5651–61. doi:10.1038/sj.onc.1203958

Saitoh S, Arudchandran R, Manetz T, Zhang W, Sommers C, Love P, et al. Lat is essential for FcεRI-mediated mast cell activation. Immunity (2000) 12:525–35. doi:10.1016/S1074-7613(00)80204- 6

Samelson LE. Signal transduction mediated by the T cell antigen receptor: the role of adapter proteins. Annu Rev Immunol (2002) 20:371–94. doi:10.1146/ annurev.immunol.20.092601.111357

Seet BT, Berry DM, Maltzman JS, Shabason J, Raina M, Koretzky GA, et al. Effi- cient T-cell receptor signaling requires a high-affinity interaction between the Gads C-SH3 domain and the SLP-76 RxxK motif. EMBO J (2007) 26:678–89. doi:10.1038/sj.emboj.7601535

Shelby SA, Holowka D, Baird B, Veatch SL. Distinct stages of stimulated FcεRI receptor clustering and immobilization are identified through superresolution imaging. Biophys J (2013) 105:2343–54. doi:10.1016/j.bpj.2013.09.049

Sil D, Lee J, Luo D, Holowka D, Baird B. Trivalent ligands with rigid DNA spacers reveal structural requirements for IgE receptor signaling in RBL mast cells. ACS Chem Biol (2007) 2:674–84. doi:10.1021/cb7001472

Stauffer TP, Meyer T. Compartmentalized IgE receptor-mediated signal trans- duction in living cells. J Cell Biol (1997) 139:1447–54. doi:10.1083/jcb.139.6. 1447

Tyson J, Novak B. Functional motifs in biochemical reaction networks. Annu Rev Phys Chem (2010) 61:219–40. doi:10.1146/annurev.physchem.012809.103457

Veatch SL, Chiang EN, Sengupta P, Holowka DA, Baird BA. Quantitative nanoscale analysis of IgE-FcεRI clustering and coupling to early signaling pro- teins. J Phys Chem B (2012) 116:6923–35. doi:10.1021/jp300197p

Xu K, Goldstein B, Holowka D, Baird B. Kinetics of multivalent antigen DNP- BSA binding to IgE-Fc epsilon RI in relationship to the stimulated tyrosine phosphorylation of Fc epsilon RI. J Immunol (1998) 160:3225–35.

Y, et al. Fyn kinase initiates complementary signals required for IgE-dependent

Yablonski D, Kadlecek T, Weiss A. Identification of a phospholipase C- gamma1 (PLC-gamma1) SH3 domain-binding site in SLP-76 required for T-cell receptor-mediated activation of PLC-gamma1 and NFAT. Mol Cell Biol (2001) 21:4208–18.

doi:10.1128/MCB.21.13.4208- 4218.2001

Young RM, Holowka D, Baird B. A lipid raft environment enhances Lyn kinase activity by protecting the active site tyrosine from dephosphorylation. J Biol Chem (2003) 278:20746–52. doi:10.1074/jbc.M211402200

Zhang J, Billingsley ML, Kincaid RL, Siraganian RP. Phosphorylation of Syk activation loop tyrosines is essential for Syk function. An in vivo study using a specific anti-Syk activation loop phosphotyrosine antibody. J Biol Chem (2000) 275:35442–7. doi:10.1074/jbc.M004549200

Zhang W, Sloan-Lancaster J, Kitchen J, Trible RP, Samelson LE. LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. Cell (1998)92:83–92. doi:10.1016/S0092- 8674(00)80901- 0

Zhang W, Trible RP, Samelson LE. LAT palmitoylation: its essential role in mem- brane microdomain targeting and tyrosine phosphorylation during T cell acti- vation. Immunity (1998) 9:239–46. doi:10.1016/S1074- 7613(00)80606- 8

Chapter 2: Models of intracellular signaling dynamics[1]

Introduction

Immune cells must process information about their changing environment to respond to signs of damage and infection. These cells possess surface receptors that bind extracellular ligands and initiate intracellular signaling, with information propagating through networks of molecular interactions. The regulation or dysregulation of these networks influences activities of the immune system, and can in some cases lead to pathological conditions such as allergies, asthma, and autoimmunity. Thus, an understanding of signaling in immune cells is an important part of understanding human health and disease. Cell signaling networks are complex, and are thus challenging to reason about using intuition alone, but computational models can enable us to expand our reasoning abilities.

Intracellular signals are propagated largely through enzyme- catalyzed reactions and noncovalent interactions. The outcomes of these biomolecular interactions are changes in the quantity of various chemical species (e.g., multiprotein complexes and various protein phosphoforms). These dynamics are governed by the same laws of chemistry and physics that govern other chemical reactions; thus, developing chemical kinetic models of intracellular signaling can be a path towards better understanding of nonlinear behaviors observed in cell signaling systems.

[1] Part of this work was previously published in: Chylek LA, Holowka DA, Baird BA, Hlavacek WS. (2014) An interaction library for the FcεRI signaling network. *Frontiers in Immunology* **5**, 172.

Several nonlinear behaviors have been observed in mast cells. Some of these behaviors include: phosphorylation responses with variable dependence on receptor phosphorylation, oscillations, and desensitization. In this chapter we will consider the first and second of these phenomena, with the third being covered in the next chapter.

The first topic that we consider is the apparent branching of pathways in FcεRI signaling. Previous studies have investigated the signaling properties of the structurally defined ligands, Y16 and Y46 (Sil et al., 2007; see also Appendix 1). These ligands differ in the spacing between DNP sites, with Y16 having approximately 5.2 nm and Y46 having 13.1 nm between sites. These studies revealed that Y46 induces less phosphorylation that Y16. However, it was also found that this reduction of receptor phosphorylation attenuates some downstream signaling events more than others. We used the rule library of Chapter 1 (Figure 1) to develop models of interactions relevant to these signaling events. We then investigated how the presence and absence of feedback loops may enable a subset of downstream signaling events to be robust to variance in upstream receptor phosphorylation.

The second topic that we consider is oscillation of intracellular Ca^{2+}. The concentration of Ca^{2+} in a mast cell's cytoplasm is a key variable influenced by FcεRI-initiated signaling processes. Following receptor crosslinking, Phospholipase Cγ1 (Plcg1) is activated and cleaves PIP_2, yielding the second messenger IP3, whose binding to IP3 receptors on the endoplasmic reticulum stimulates release of Ca^{2+} from these stores, leading to store operated Ca^{2+} entry from the extracelluar medium. Oscillations in Ca^{2+} are a common occurrence during mast cell activation, and these oscillations are linked to granule exocytosis. Ca^{2+} oscillations occur downstream of oscillations in abundance of PIP_2 and activation of the Rho-family GTPase Cdc42 (Wilkes et al.,

2014). However, the details of how the dynamics of these proteins and lipids give rise to Ca^{2+} oscillations in mast cells has remained somewhat enigmatic. We developed a model that produces Ca^{2+} oscillations and compared its predictions to preliminary experimental data.

Methods

Our rule-based models are specified in BioNetGen Language (Faeder et al., 2009). Our simulations are based on the law of mass action and an assumption of well-mixed reaction compartments. In the models, the following compartments are considered implicitly: the cytosol, the plasma membrane, and the extracellular fluid surrounding a single cell. Simulations were performed using CVODE, the built-in deterministic simulator of BioNetGen, which takes as input the ODEs derived from a rule-specified reaction network. Our illustrations of rules are based on published guidelines for model visualization and were drawn with the help of a template available online (http://bionetgen.org/index.php/ Extended_Contact_Maps).

In our bifurcation analyses, we found stable steady states through simulations that were started from arbitrary initial conditions or nearby steady states. The bifurcation parameter was an input signal taken in the model of interest to control the rate of activation of Syk and Fyn, which were each deactivated through a first-order process. Thus, as the input signal increases, so too do the steady-state levels of active Syk and Fyn. In simulations performed to find stable steady states, the bifurcation parameter was systematically varied from a low to high value, and vice versa.

To characterize signaling dynamics for specific observables (i.e., model outputs), we calculated rise time as the time required for the observable to reach 95% of its final steady-state value. For comparison between two models, a ratio of rise times was calculated.

Results and Discussion

Differential impact of receptor phosphorylation on downstream events

We used our rule library (of Chapter 1) to develop models for investigation of signaling dynamics. We focused on the adaptor protein Lat, which is known for its role as a signaling hub in both T cells and mast cells. This role arises in large part from its capacity to recruit multiple adaptors and enzymes that regulate lipid metabolism and production of second messengers. Most of Lat's interactions depend on prior Lat phosphorylation, which is catalyzed primarily by Syk. However, studies of FcεRI signaling induced by structurally defined antigens have revealed that not all "downstream" events are equally dependent on Lat phosphorylation. Specifically, a panel of trivalent antigens, Y16 and Y36, which are composed of 2,4-dinitrophenyl (DNP) hapten groups conjugated to rigid double-stranded DNA scaffolds and differing in the distance between DNP groups, was evaluated for the ability to induce phosphorylation of signaling proteins, Ca^{2+} mobilization, and degranulation by DNP-sensitized mast cells (Sil et al., 2007). It was found that phosphorylation of FcεRI and Lat, as well as store-operated Ca^{2+} entry and degranulation, were strongly dependent on spacing between DNP groups, with the shortest spacing examined associated with the strongest responses. In contrast, it was also found that release of Ca^{2+} from intracellular stores did not show as strong a dependence on the distance between hapten sites. Given that Ca^{2+} release is thought to occur as a result of activities of proteins that depend on Lat, how can this apparent uncoupling between Lat phosphorylation and Ca^{2+} mobilization be explained?

We hypothesize that compensatory mechanisms mediated by Fyn and Gab2 (Parravicini et al., 2003) are involved in this phenomenon. Gab2 can be phosphorylated by Fyn, and can then recruit PI3K. As discussed in Chapter 1, production of PIP_3 by PI3K contributes to activation of

Plcg1. A product of Plcg1 is IP3, which induces release of Ca^{2+} from intracellular stores. Thus, if Gab2 recruitment and activation is less sensitive to differences in Lat phosphorylation level, then Gab2 may open an avenue by which Ca^{2+} mobilization could escape control of Lat. We used our rule library to build models to determine if Gab2 could potentially enable Ca^{2+} mobilization when Lat phosphorylation is diminished.

We first considered a model in which Syk and Fyn are independent inputs. Our initial model (Listing 1) essentially consists of the first coherent feed-forward loop shown in Chapter 1, Figure 4: Lat recruits Plcg1, as well as PI3K through Gab2 and Grb2. Btk is recruited to PIP_3 and activates Plcg1 through phosphorylation. Fyn participates by phosphorylating Gab2. To model the differences between antigens observed to induce the most and least Lat phosphorylation, we considered different levels of active Syk consistent with the approximately fourfold difference in Lat phosphorylation observed experimentally for the shortest and longest ligands (Sil et al., 2007). The level of active Fyn was kept constant. Simulations of this model revealed that differences in Lat phosphorylation level were maintained or amplified in downstream events. According to the model, a decrease in Lat phosphorylation (arising from lowered Syk activity) causes at least proportionate decreases in the levels of activated Plcg1, activated Btk, Lat-associated PI3K, PIP_3, and IP3 (Figure 1). We also considered a scenario in which activity of Syk and Fyn are both controlled by the magnitude of an input signal, which is likely a more realistic scenario because both kinases are recruited to phosphorylated FcεRI. We varied the strength of this signal and evaluated the resulting steady-state levels of outputs (Figure 2). Consistent with results from the first scenario, decreased signal strength leads to decreased Lat phosphorylation, and is ac-

companied by even steeper decreases in activation of other signaling molecules. Thus, the interactions included in this model are insufficient to explain the experimental observation of Ca^{2+} mobilization in the absence of strong Lat phosphorylation.

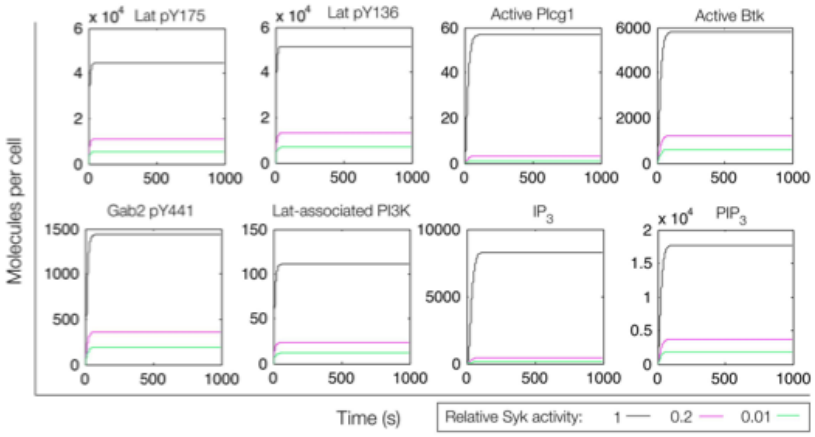


Figure 2.1. Simulation of a model of the feed-forward loop connecting Lat to IP3 production. Different lines indicate different relative levels of Syk activity, as indicated in the legend. In these simulations, Syk activity was set at the indicated level and held constant.

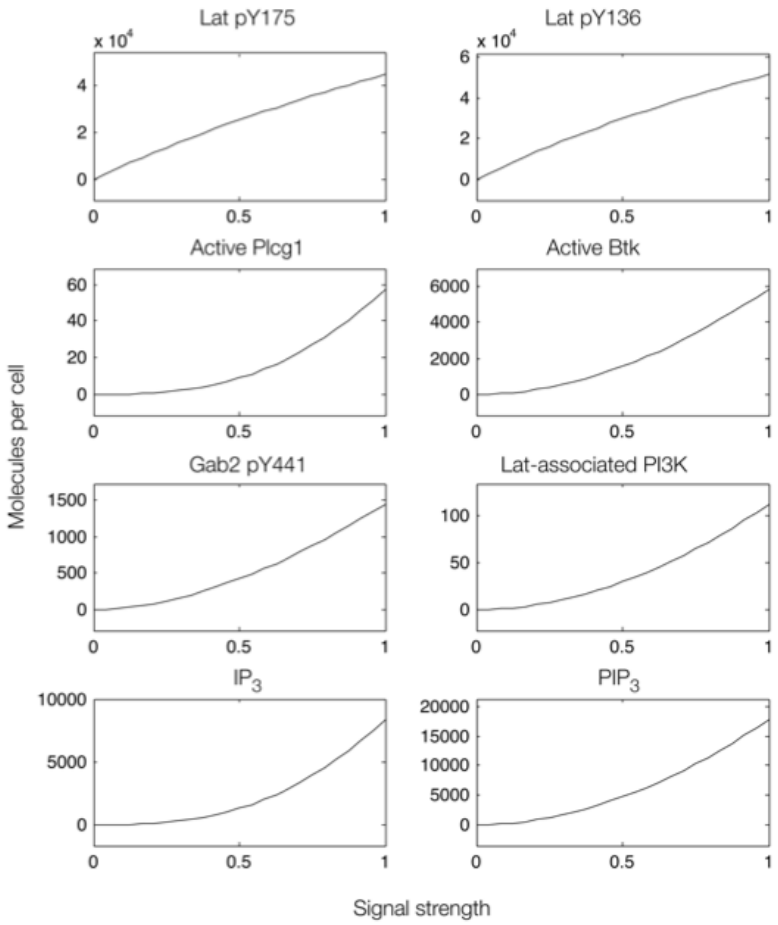


Figure 2.2. Steady-state dose–response curves, which were found by simulation of the feed-forward loop connecting Lat to IP_3 production. The differences in phosphorylation level between the two phosphorylation sites in Lat (top panels) results from different affinities of the binding partners that interact with each site. Active Plcg1 is taken to be Plcg1 that is both recruited to Lat and phosphorylated, and active Btk is taken to be Btk recruited to PIP_3.

In an extension of the initial model (Listing 2), we incorporated additional interactions from the rule library, those responsible for the positive feedback involving Gab2 interaction with PIP_3 (see Chapter 1, Figure 4). We reasoned that, with the addition of these interactions, once PIP_3 production is initiated, PIP_3 production may become self-sustaining, because PIP_3 is able to recruit Gab2 to the plasma membrane, which in turn is able to recruit PI3K. Simulated time courses with the same level of active Fyn and different levels of active Syk, as in Figure 1, are

shown in Figure 3. These results indicate that certain signaling readouts downstream of Lat are buffered against reduced Lat phosphorylation. For example, there is less than a fourfold difference in peak IP3 levels between the conditions of high (black line) and intermediate (magenta line) Lat phosphorylation. In contrast, the model without Gab2-mediated positive feedback predicts a greater than 100-fold difference.

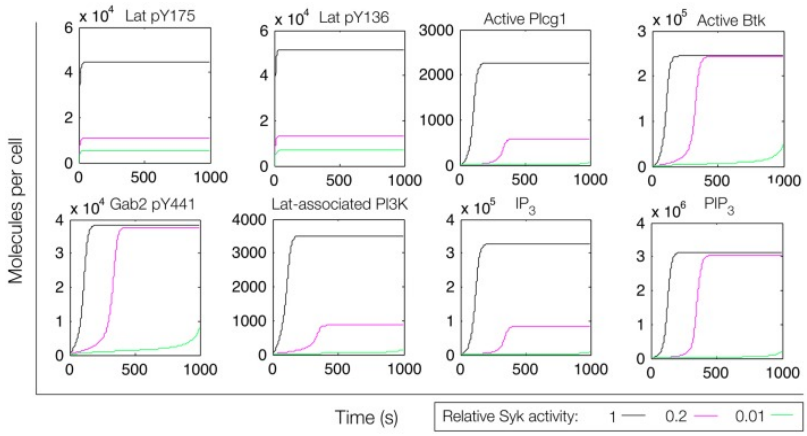


Figure 2.3. **Simulation of a model of the feed-forward loop connecting Lat to IP3 production with consideration of a Gab2-mediated positive feedback loop**. Different color lines indicate different relative levels of Syk activity.

To further investigate the role of positive feedback, we modulated an input signal controlling both Fyn and Syk activity, as in Figure 2. Steady-state simulation results from this model are shown in Figure 4, which differ from those obtained with the first model. First, the total numbers of signaling molecules in activated forms are greater than for the case without feedback, as long as the signal strength is above a certain level. Second, within certain input ranges, the model shows bi-stability, i.e., existence of two stable steady states, as indicated by signal strength values that correspond to more than one steady-state output value. In these cases, the system's final steady state will depend on the system's initial state; in other words, it is a history-dependent

process. Bi-stability has also been characterized in TCR signaling (Lipniacki et al., 2008, Das et al., 2009) and BCR signaling (Barua et al., 2012, Mukherjee et al., 2013). Third, we found that certain signaling readouts downstream of Lat are now buffered against reduced Lat phosphorylation (Figure 4), decreasing less sharply when signal is reduced. Together, these results suggest that Gab2-mediated positive feedback may enable committed, all-or-none decisions that lead to high levels of IP3 as long as Lat phosphorylation is above a threshold. When the level of Lat phosphorylation falls below this threshold, positive feedback is unable to enhance IP3 production (Figure 4). Thus, some amount of PIP_3 must be generated through Lat-dependent mechanisms before the Fyn/Gab2 pathway can contribute to production of IP_3. Giving some support to this hypothesis, it has previously been observed that Ca^{2+} release from stores induced by Y16 or Y46 is greatly reduced in the presence of the PI3K inhibitor Wortmannin, which would reduce the quantity of PIP_3 (Sil et al., 2007)

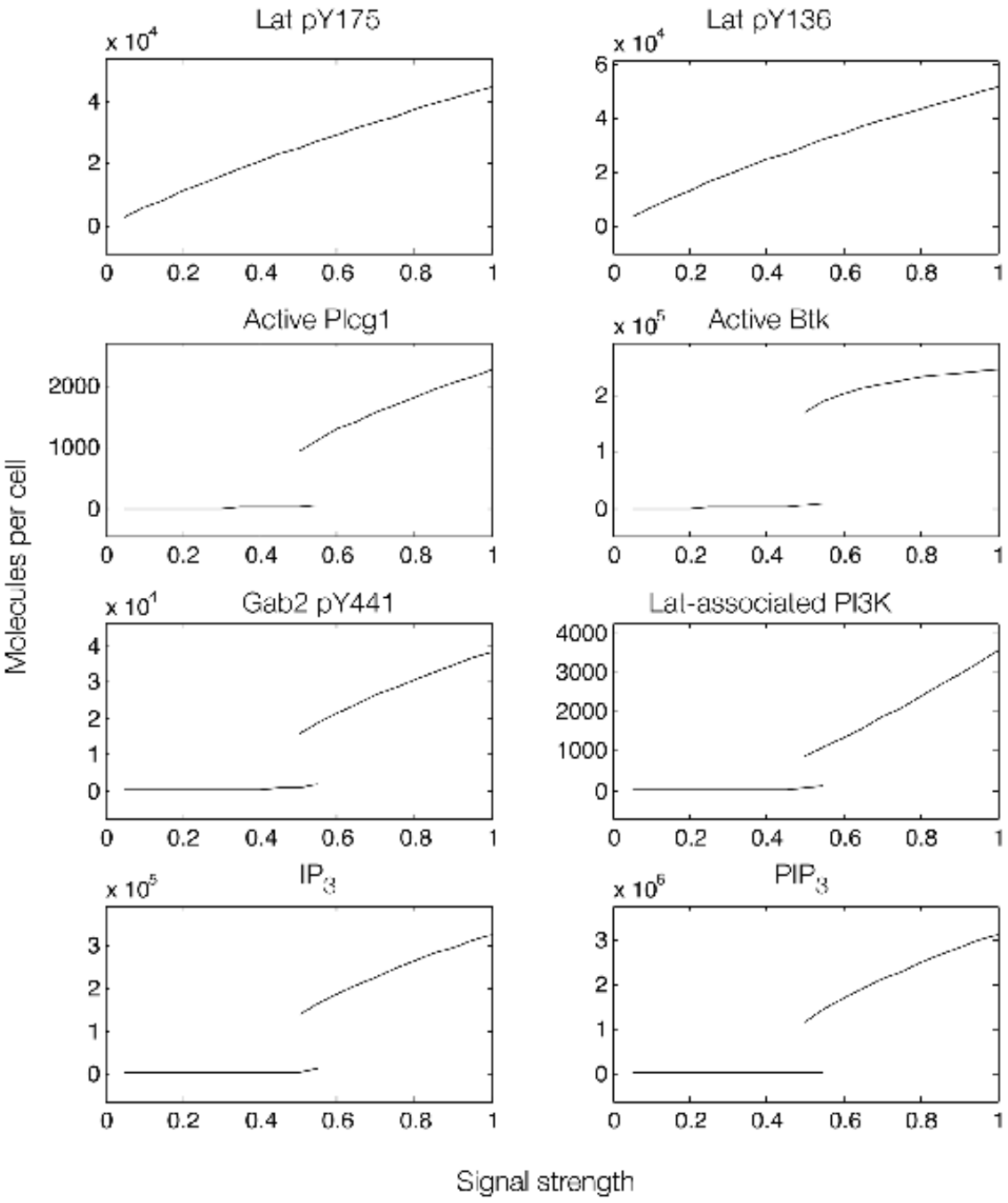


Figure 2.4. Steady-state dose-response curves, which were found by simulation of the feed-forward loop connecting Lat to IP_3 production when Gab2-mediated positive feedback is considered. Each plot is a bifurcation diagram; the bifurcation parameter is signal strength, which governs the rate of production of active Syk and Fyn. Only stable steady states are shown. As can be seen, the model predicts the possibility of bistability (existence of two steady states for a single signal-strength value).

We also considered how positive feedback affects the dynamics of signaling. We calculated the rise time for IP_3 at different input levels as predicted by the models with and without positive feedback. We found that positive feedback caused IP3 level to reach its steady state more slowly (Figure 5A). Rise time for the model with positive feedback peaked in the bistable region, where the system transitions from a low steady state to a higher steady state (Figure 5B). The slower rise in IP_3 level qualitatively mimics the experimentally observed dynamics of Ca^{2+} release from stores caused by antigens that induce low levels of Lat phosphorylation (Sil et al., 2007). These same antigens induce minimal store-operated calcium entry (SOCE) and minimal degranulation, which suggests that SOCE may be sensitive to the kinetics of IP3 production. In other words, this is a potential explanation for why SOCE is ligand length-dependent despite the (relatively) length independent release of Ca^{2+} from stores.

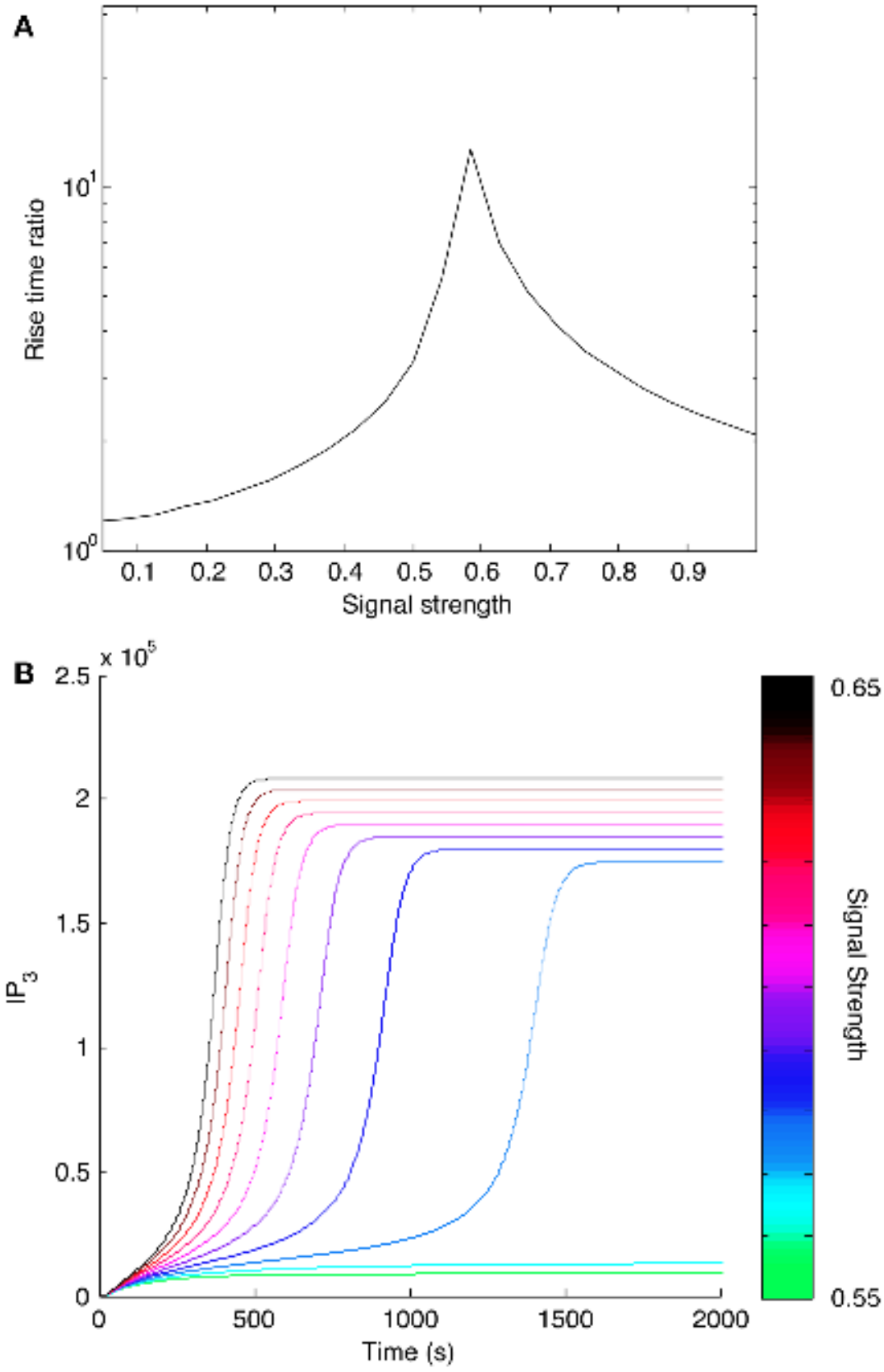

A
Rise time ratio
10^1
10^0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Signal strength
B
x 10^5
2.5
2
1.5
1
0.5
0
IP_3
0
500
1000
1500
2000
Time (s)
0.65
0.55
Signal Strength

Figure 2.5. Effect of positive feedback on signaling dynamics. (A) Rise time for IP_3 synthesis was calculated as the time needed to reach 95% of the final steady-state level. Rise times were calculated for different levels of input, or signal strength. Rise times for the model with positive feedback were divided by rise times for the model without positive feedback and plotted against corresponding input level. All indicated rise time ratios are greater than one, meaning that the model with positive feedback takes more time to reach its final steady state. (B) Time courses for IP_3 production in a narrow range of input levels surrounding the peak shown in (A). The input level corresponding to each curve is indicated with the color bar at the right.

There are several experimental tests that could be pursued to evaluate the role of Gab2-mediated positive feedback. One predicted effect of the feedback loop is bistability of several signaling readouts (Figure 4), including PIP_3. Testing for bistability usually benefits from single-cell measurements, because cell-to-cell variability may result in different cells having different bifurcation points. At the single-cell level, PIP_3 production can be monitored using PH domain constructs (Halet, 2005). When the strength of an input signal, such as ligand-induced receptor aggregation, crosses a threshold level, the quantity of PIP_3 is expected to increase dramatically in a switch-like manner. Another characteristic arising from bistability is hysteresis, meaning history dependence. As signal strength is reduced from a high level [e.g., by breaking up IgE-FcεRI aggregates with a monovalent hapten (Shelby et al., 2013)], PIP_3 level is expected to switch back to a low state. However, this switch is predicted to occur at a lower input level than that required to induce a transition from weak signaling to robust signaling. Controlling input level would benefit from an understanding of how ligand dose relates to receptor aggregation, which can be obtained with a model for ligand–receptor interactions (Monine et al., 2010, Appendix 1).

A second approach would involve disruption of the Gab2 feedback loop, which would be expected to increase sensitivity of mast cell degranulation to Lat phosphorylation. Mutation of the Gab2 PH domain, which binds PIP_3 and is therefore a key component of the feedback, would be expected to inhibit Ca^{2+} mobilization. However, such manipulation of endogenous Gab2

would be technically challenging, making this strategy potentially difficult to implement. An alternative approach would be to use RNAi to knock down expression of either *Gab2* or *Fyn*, which would be predicted to similarly inhibit Ca^{2+} mobilization.

Oscillations in early signaling events and intracellular Ca^{2+}

From the standpoint of network topologies, oscillations can have several possible sources (Kholodenko, 2006). One possibility is a negative feedback loop combined with a time delay and ultrasensitivity; once an output climbs above a threshold, the negative feedback pulls it to a lower state before the output is able to climb again. Another possible source is a combination of positive and negative feedbacks that give rise to a bistable system. Several different network structures incorporating opposing feedbacks have been found to be capable of producing oscillations (for example, see Tsai et al., 2008), and one of these structures has been investigated in the context of immunoreceptor signaling (Barua et al., 2012).

A previously formulated model for B-cell antigen receptor (BCR) signaling (Barua et al., 2012) predicted that a confluence of positive and negative feedbacks gives rise to oscillations in the quantity of activated Src-family kinases Lyn and Fyn. The negative feedback in this system arises from the scaffold protein that recruits the protein tyrosine kinase Csk, which is known as Cbp (Csk-binding protein) or Pag1(PAG). This protein is localized at lipid-ordered regions of the membrane. This scaffold has several tyrosine sites that can undergo phosphorylation, some of which interact with the SH2 domains of Lyn and Fyn. Other sites are capable of interacting with Csk. When Csk is colocalized with Lyn/Fyn, it is able to phosphorylate the Src-family kinase on its inhibitory C-terminal tyrosine. When this site is phosphorylated, the kinase is more likely to assume an auto-inhibitory conformation that reduces its catalytic activity.

Positive feedback arises because Src-family kinases phosphorylate FcεRI sites that they are then able to bind to; thus, recruitment of kinases to the receptor is perpetuated. Through the tug-of-war between positive and negative feedback, the system is pushed from one state to another, leading to oscillations in receptor phosphorylation and Src-family kinase activity.

We hypothesized that, if these FcεRI-proximal signaling readouts oscillate, they might contribute to oscillations of downstream readouts, particularly intracellular Ca^{2+} concentration. Using our rule library of Chapter 1, we adapted the model of Barua et al. (2012) for BCR signaling to the FcεRI system and extended it such that it includes Ca^{2+} release from endoplasmic reticulum stores (Fig. 6).

We altered the original model such that the receptor is FcεRI rather than BCR. We preserved the shared features of signal initiation: autoinhibitory intramolecular interactions of Lyn, phosphorylation of receptor immunoreceptor tyrosine-based activation motifs (ITAMs) by Lyn, phosphorylation of Lyn's activation loop tyrosine, phosphorylation of PAG, binding of Lyn and Csk to PAG, phosphorylation of Lyn's inhibitory C-terminal site by Csk, and recruitment of Syk to phosphorylated ITAMs. The model is extended to include a coarse-grained representation of activation of Dock GEF downstream of Syk, followed by activation of Cdc42 and PIP5 Kinase, which results in enhancement of PIP2 levels, enabling release of Ca^{2+} from the endoplasmic reticulum (Fig. 6). In short, this model proposes that oscillations in Lyn activity, resulting from positive and negative feedbacks, are perpetuated through multiple signaling steps to the level of Ca^{2+} release.

For the parameter values that lead to oscillations in Src-family kinase activity, the model predicts that oscillations in intracellular Ca^{2+} also occur (Fig. 7). We sought to evaluate this prediction experimentally.

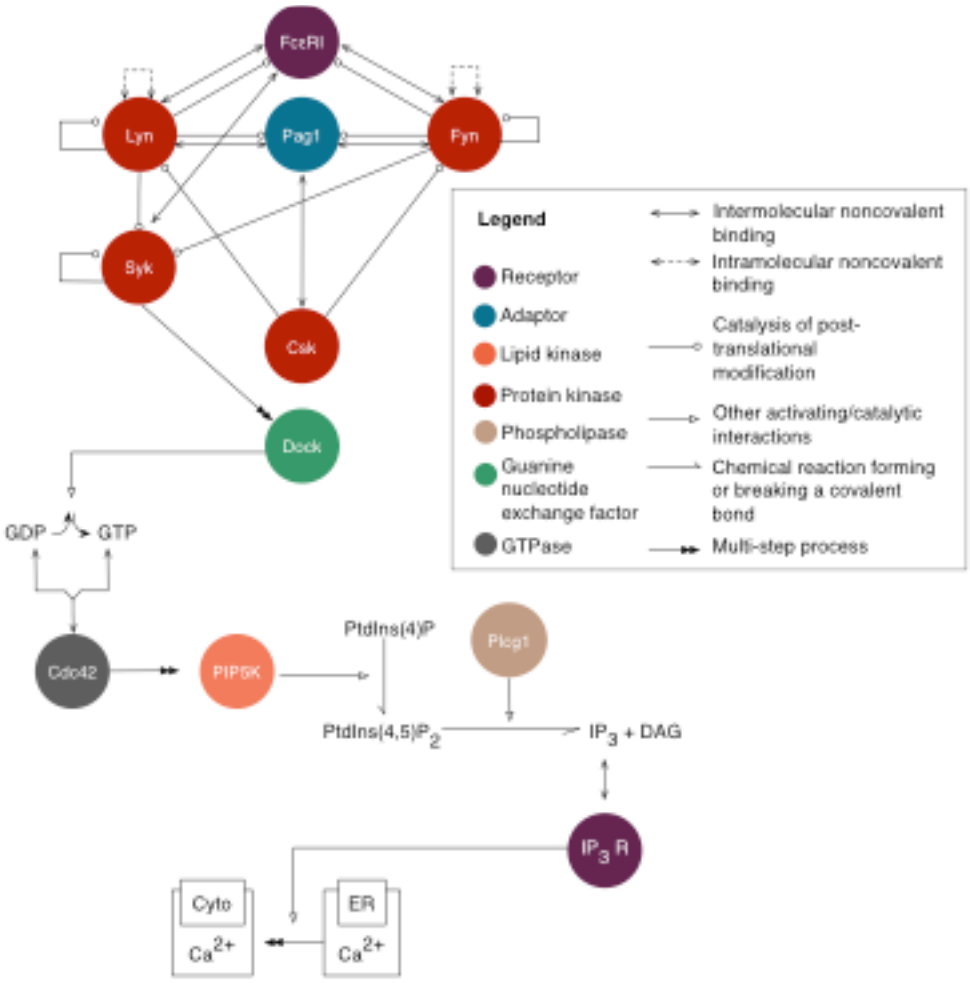


Figure 2.6. Diagram of the model of Ca^{2+} oscillations. In this model, a combination of positive and negative feedbacks among the Src-family kinases and Csk leads to oscillations in receptor phosphorylation. These oscillations are propagated into oscillations of Syk kinase activity, and downstream into oscillations in cytoplasmic Ca^{2+}.

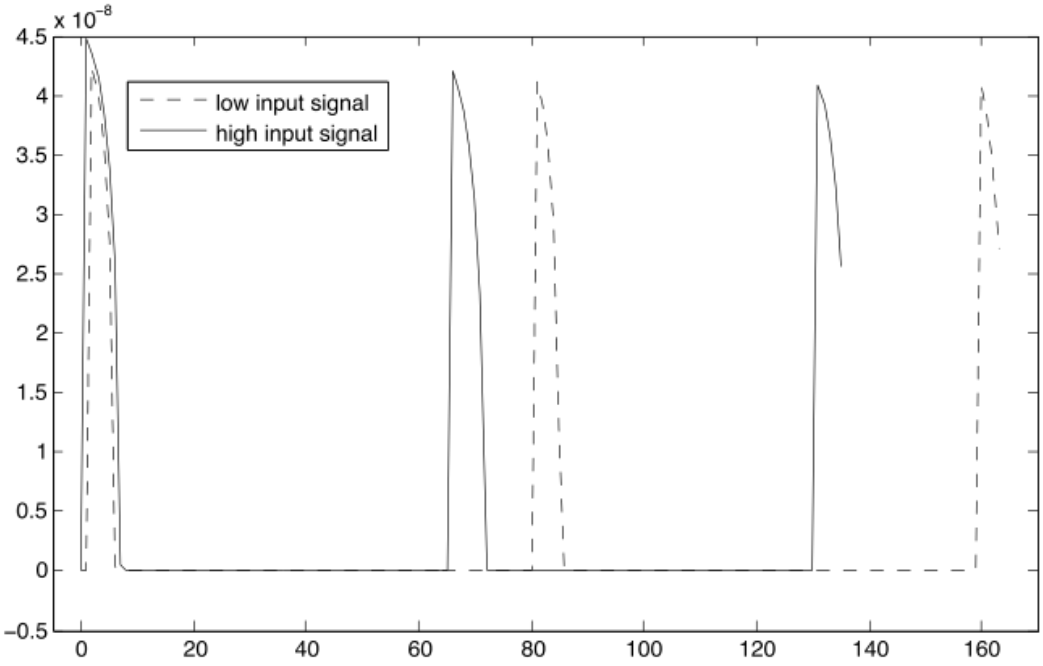
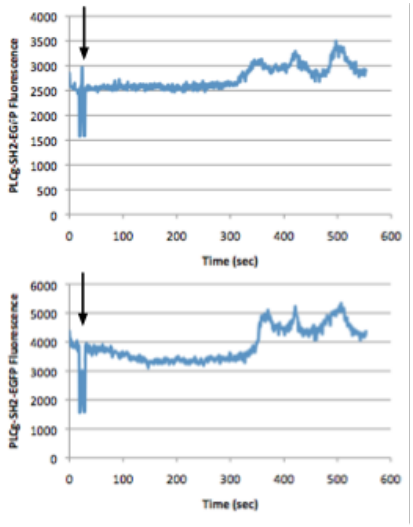


Figure 2.7. Predicted effect of increased antigen on the frequency of Ca^{2+} oscillations.

A first step was to find whether oscillations in initiatory signaling events are detectable. To do this, two different fluorescent proteins containing SH2 domains were used: a fluorescent construct of the SH2 domains of PLCγ (Fig. 8), and Syk-YFP (Fig. 9). When monitored using TIRF, both of these proteins were seen to be recruited to puncta (Fig. 10) at the plasma membrane in a manner that appears to be oscillatory. Recruitment of these proteins is representative of phosphorylation of molecules at the plasma membrane, which may include FcεRI and Lat, giving some credence to the hypothesis that receptor phosphorylation can oscillate. However, when the PLC SH2 construct is expressed in Syk- cells, fewer puncta form and oscillations are not detectable (M. Wilkes, data not shown), suggesting that the construct might be binding directly to Syk, or that Syk is otherwise necessary to facilitate recruitment (e.g., by phosphorylation of Lat).



dFigure 2.8. (From Marcus Wilkes.) Recruitment of PLCγ SH2-EGFP to puncta at the plasma membrane, as monitored through TIRF microscopy. The multivalent antigen DNP-BSA (DNP-conjugated bovine serum albumin) was added at the time indicated with an arrow. Responses from two different cells are shown.

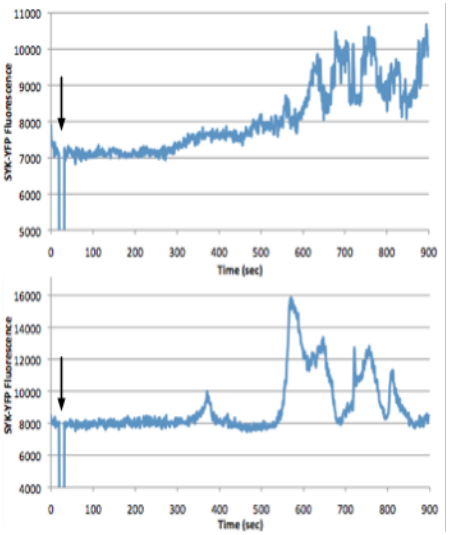


Figure 2.9. (From Marcus Wilkes.) Recruitment of Syk-YFP, monitored as puncta of YFP fluorescence at the plasma membrane, as determined through TIRF microscopy. The multivalent antigen DNP-BSA was added at the time indicated with an arrow.

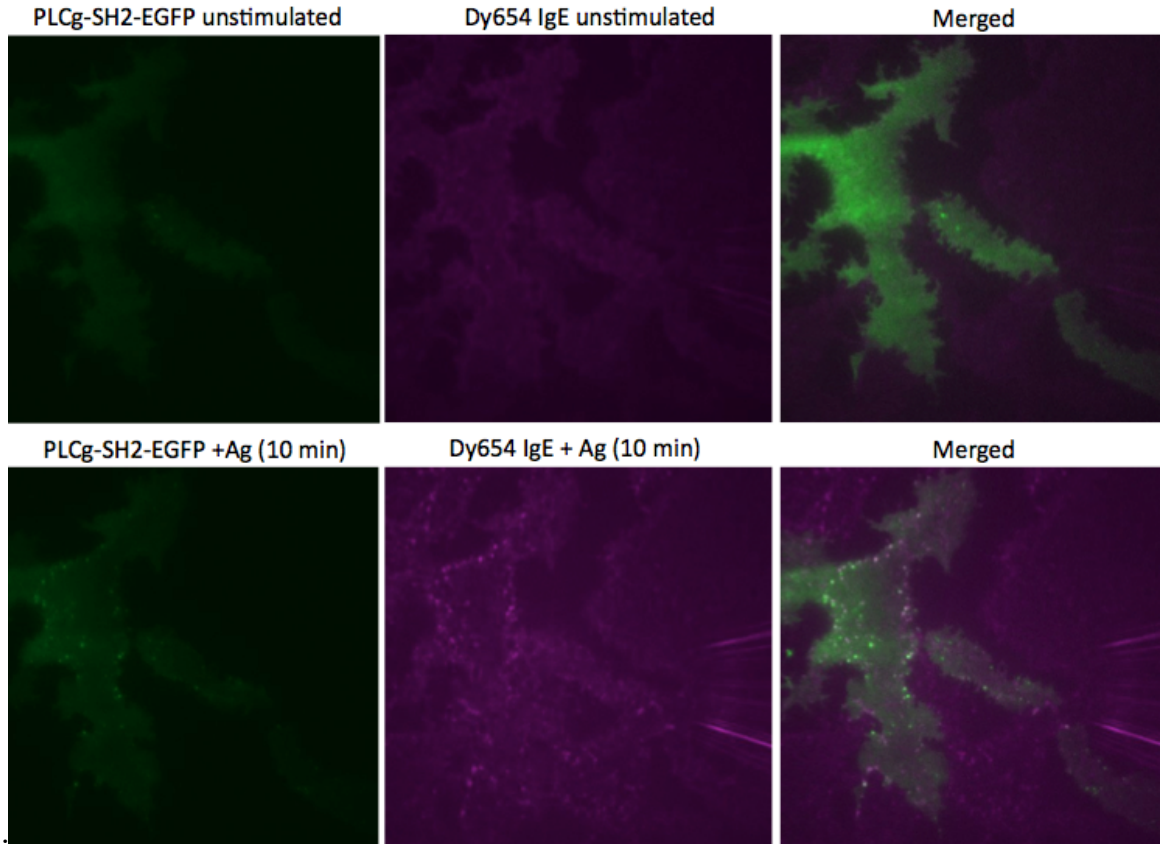


Figure 2.10. (From Marcus Wilkes) Upon antigen stimulation, the PLC SH2 construct is recruited to puncta co-localized with receptor clusters.

Oscillations were not observed in all experimental replicates; further investigations are necessary to ascertain whether or not oscillatory receptor phosphorylation is a genuine feature of FcεRI signaling. A prediction of the model is that an increase in antigen dose will increase the frequency of oscillations in receptor phosphorylation (Fig. 7), which was not observed; in the set of experiments where antigen (DNP-BSA) dose was varied, no oscillations were detected in recruitment of PLCγ-SH2-GFP, although clusters did form.

A way to pursue the issue of oscillations more systematically might be to use a well-characterized antigen, where the relationship of antigen dose to IgE- FcεRI aggregation and intracellular signaling can be characterized more rigorously than is possible for the chemically heterogeneous antigen DNP-BSA. Additionally, it may be possible to assess recruitment more precisely via FRET between FcεRI and Syk, as a read-out of Syk recruitment. Such techniques have been employed in investigation of BCR signal initiation (Tolar et al., 2005).

Once we have gathered enough data on signaling oscillations to be sufficiently confident, the model and experiments can be linked more tightly by optimizing the model. One aspect of this process can involve parameter fitting, for which we have developed a specialized tool (described in Appendix 2). Two areas that may be in need of quantitative refinement are the frequency of oscillations and the delay between antigen stimulation and the onset of oscillations. The frequency of oscillations depends on a number of parameters, prominently the rate constant for phosphorylation of the receptor by Lyn. Both the frequency of oscillations and the delay before onset are also influenced by the parameters describing antigen stimulation. Thus, it would be optimal to couple a model that accurately captures antigen-induced receptor aggregation to a model for intracellular oscillations. If further steps are necessary to align the model with experiments, the model could be expanded to include mechanisms that are not currently part of it, e.g., by including negative regulation by the tyrosine phosphatase Shp-1.

Finally, it may be necessary to investigate different combinations of feedbacks to see if any may give rise to Ca^{2+} oscillations. Interactions that are likely to have an influential role involve Protein Kinase C (PKC) and MARCKS. MARCKS is a protein that can bind to and thereby sequester membrane phosphoinositides, via its polybasic effector domain. When MARCKS is phosphorylated by PKC, its dissociation from the membrane is induced. The newly

exposed lipid species can then serve as substrates for phosphoinositide kinases, thus influencing intracellular levels of IP_3, while also facilitating exocytosis (Gadi et al., 2011). These interactions of MARCKS and PKC could be incorporated into our model in the usual manner: by specifying the relevant domains and residues of these proteins, and constructing rules to describe how these sites interact with each other as well as with existing model components (e.g., PIP_2). As discussed previously, many different circuitries are capable of producing oscillations, and recently developed software tools may accelerate the task of locating such feedbacks in the FcεRI signaling network (Sekar et al., 2015).

Conclusions

We have developed a set of models with which to investigate two nonlinear phenomena that have been observed in mast cell signaling: differential dependence of signaling events on receptor phosphorylation caused by different ligands, and oscillations in intracellular Ca^{2+}. In both cases, we have used our models to make experimentally testable predictions, and in the latter case preliminary experiments have been performed. Further pursuit of these tests may provide data to either contradict or support our current hypotheses. An interesting later step could involve combining aspects of these two models, assessing the nature of Ca^{2+} oscillations when cells are stimulated with antigens that induce different levels of receptor phosphorylation. Together, these studies will enhance our understanding of how Ca^{2+} dynamics, a central aspect of mast cell physiology, are influenced by the initiatory molecular steps in FcεRI signaling.

Listing 1: A model without Gab2 positive feedback.

```
# Example Model 1 (no Gab2 positive feedback) from "An interaction library
for the Fc(epsilon)RI signaling network."
# L.A. Chylek, W.S. Hlavacek, D.A. Holowka, and B.A. Baird

begin molecule types
Lat(Y136~0~P,Y175~0~P)
Grb2(SH2,cSH3)
Gab2(PRS,PH,Y441~0~P)
PI3K(SH2,cat)
Plcg1(SH2,cat,Y783~0~P)
PI(OH3~0~P,OH4~0~P,OH5~0~P,head~0~1)
Btk(PH,cat)
Syk(cat)
Fyn(cat)
pre
zero
end molecule types

begin observables
# Lat pY175
Molecules pY175 Lat(Y175~P!?)

# Lat pY136
Molecules pY136 Lat(Y136~P!?)

# Active Plcg1
Molecules actPlcg Plcg1(SH2!+,Y783~P)

# Active Btk
Molecules Btkrec Btk(PH!+)

# Gab2 pY441
Molecules pGab Gab2(Y441~P!?)

# Lat-associated PI3K
Molecules PI3K_Lat PI3K().Lat()

# IP3
Molecules IP3    PI(OH3~0!?,OH4~P,OH5~P,head~0!?)

# PIP3
Molecules PIP3  PI(OH3~P!?,OH4~P,OH5~P,head~1!?)

# PIP2
Molecules PIP2  PI(OH3~0!?,OH4~P,OH5~P,head~1!?)

# Grb2Gab2
Molecules Grb2Gab2 Grb2().Gab2()
end observables
```

```
begin parameters

NA          6.022e23          # Avogadro's number; molecules/mole.
celldensity 1e9               # Cells/L
Fx                1 #0.02                # Fraction of cell volume to si-
muate; unitless.
ECFvol          1/(celldensity)          # Extracellular volume; L/cell.
simECFvol       ECFvol*Fx           # Simulated fraction of extracellular
volume; L.
Cellvol         1.4e-12             # Cytoplasmic volume; L. Value from Fae-
der et al.
simCellvol  Cellvol*Fx        # Simulated fraction of cell volume; L.
ProteinTot  3e5               # Protein copy number per cell.
SimProteinTot     Fx*ProteinTot           # Simulated fraction of protein
copy number.
ActiveKinase1 0.5*SimProteinTot
ActiveKinase2 0.1*SimProteinTot
ActiveKinase3 0.05*SimProteinTot
kcSyk 100
kcFyn 100
KD1  79e-9*NA*simCellvol
KD2 62e-9*NA*simCellvol
krSH2 0.12
kfSH2_1 krSH2/KD1
kfSH2_2 krSH2/KD2
kfPH 1.4e6/(NA*simCellvol)
krPH 1
kfSH3 1e6/(NA*simCellvol)
krSH3 1
kcBtk 1e-3
kfP 1e9/(NA*simCellvol)
krP 1
kcP 10
kDeg 0.1
kPten 0.1
kX 5e-3
end parameters

begin functions
# A logistic growth function is used to model replenishment of PIP2 through,
e.g., activity of PIP5 kinase.
Replenish() = kX*PIP2*(1-PIP2/SimProteinTot)

end functions

begin species

# All molecules except active Syk and active Fyn are assumed to be present at
the same concentration
Lat(Y136~0,Y175~0) SimProteinTot
Grb2(SH2,cSH3) SimProteinTot
Gab2(PRS,PH,Y441~0) SimProteinTot
PI3K(SH2,cat) SimProteinTot
Plcg1(SH2,cat,Y783~0) SimProteinTot
PI(OH3~0,OH4~P,OH5~P,head~1) SimProteinTot
Btk(PH,cat) SimProteinTot
```

```
pre SimProteinTot

# Active Syk and active Fyn have an initial concentration of 0, which is
changed after a period of equilibration.
# See simulation commands at end of file for more information.
Syk(cat) 0
Fyn(cat) 0
end species

begin reaction rules

# Syk phosphorylates Lat Y136
Lat(Y136~0) + Syk(cat) <-> Lat(Y136~0!1).Syk(cat!1) kfSH3,krSH3
Lat(Y136~0!1).Syk(cat!1) ->Lat(Y136~P) + Syk(cat)  kcSyk

# Syk phosphorylates Lat Y175
Lat(Y175~0) + Syk(cat) <-> Lat(Y175~0!1).Syk(cat!1) kfSH3,krSH3
Lat(Y175~0!1).Syk(cat!1) ->Lat(Y175~P) + Syk(cat)  kcSyk

# Lat pY136 binds Plcg1
Lat(Y136~P) + Plcg1(SH2) <-> Lat(Y136~P!1).Plcg1(SH2!1) kfSH2_2,krSH2

# Lat pY175 binds Grb2
Lat(Y175~P) + Grb2(SH2) <-> Lat(Y175~P!1).Grb2(SH2!1) kfSH2_1,krSH2

# Fyn phosphorylates Gab2
#   Binding Gab2.Grb2.Lat
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~0) + Fyn(cat) -
>Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~0!3).Fyn(cat!3)  kfSH3
#   Unbinding
Gab2(Y441~0!1).Fyn(cat!1) -> Gab2(Y441~0) + Fyn(cat) krSH3
#   Phosphorylation
Gab2(Y441~0!1).Fyn(cat!1) -> Gab2(Y441~P) + Fyn(cat) kcFyn

# Gab2 binds Grb2
Grb2(cSH3) + Gab2(PRS) <-> Grb2(cSH3!1).Gab2(PRS!1) kfSH3,krSH3

# Gab2 binds PI3K
Gab2(Y441~P) + PI3K(SH2) <-> Gab2(Y441~P!1).PI3K(SH2!1) kfSH2_1,krSH2


# PI3K generates PIP3

#   PI3K.Gab2.Grb2.Lat binds PIP3
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~P!3).PI3K(SH2!3,cat) +
PI(OH3~0,OH4~P,OH5~P,head~1) ->
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,Y441~P!3).PI3K(SH2!3,cat!4).PI(OH
3~0!4,OH4~P,OH5~P,head~1) kfP

#   Unbinding
PI3K(cat!4).PI(OH3~0!4,OH4~P,OH5~P,head~1) -> PI3K(cat) +
PI(OH3~0,OH4~P,OH5~P,head~1) krP

#   Phosphorylation
PI3K(cat!4).PI(OH3~0!4,OH4~P,OH5~P,head~1) -> PI3K(cat) +
PI(OH3~P,OH4~P,OH5~P,head~1) kcP
```

```
# Btk binds PIP3
PI(OH3~P,OH4~P,OH5~P,head~1) + Btk(PH) <->
PI(OH3~P,OH4~P,OH5~P,head~1!1).Btk(PH!1) kfPH,krPH

# Btk phosphorylates Plcg1

#   Binding
Btk(PH!+,cat) + Plcg1(SH2!+,Y783~0) -> Btk(PH!+,cat!1).Plcg1(SH2!+,Y783~0!1)
kfSH3

#   Unbinding
Btk(cat!1).Plcg1(Y783~0!1) -> Btk(cat) + Plcg1(Y783~0) krSH3

#   Phosphorylation
Btk(cat!1).Plcg1(Y783~0!1) -> Btk(cat) + Plcg1(Y783~P) kcSyk


# Plcg1 generates IP3

#   Binding PIP2
Plcg1(cat,SH2!+,Y783~P) + PI(OH3~0,OH4~P,OH5~P,head~1) ->
Plcg1(cat!1,SH2!+,Y783~P).PI(OH3~0,OH4~P,OH5~P,head~1!1) 10*kfP

#   Unbinding
Plcg1(cat!1).PI(head~1!1) -> Plcg1(cat) + PI(head~1) krP

#   Catalysis
Plcg1(cat!1).PI(head~1!1) -> Plcg1(cat) + PI(head~0) kcP

# IP3 is cleared
PI(OH3~0,OH4~P,OH5~P,head~0) -> zero kDeg

# PIP3 is dephosphorylated to PIP2
PI(OH3~P,OH4~P,OH5~P,head~1) -> PI(OH3~0,OH4~P,OH5~P,head~1) kPten

# PIP2 is replenished. "pre" represents a precursor molecule. Replenish() is
a functional rate law defined in the "functions" block above.

pre -> pre + PI(OH3~0,OH4~P,OH5~P,head~1) Replenish()

# Dephosphorylation
Lat(Y175~P) -> Lat(Y175~0) 5
Lat(Y136~P) -> Lat(Y136~0) 5
Gab2(Y441~P) -> Gab2(Y441~0) 5
Plcg1(Y783~P) -> Plcg1(Y783~0) 5
end reaction rules


# Generate reaction network from rule set
generate_network({overwrite=>1});

# Execution of following command performs equilibration before Syk and Fyn
are added as inputs
simulate_ode({t_end=>1000,n_steps=>1000});
```

```
# The following command sets the Syk concentration. The different concentra-
tions considered were ActiveKinase1, ActiveKinase2, and ActiveKinase3, which
are specified in the parameters block above.
setConcentration("Syk(cat)","ActiveKinase1");

# The following command sets the Fyn concentration. Only one Fyn concentra-
tion (ActiveKinase1) was considered.
setConcentration("Fyn(cat)","ActiveKinase1");

# The following command simulates system behavior after addition of Syk and
Fyn inputs
simulate_ode({suffix=>stim1,t_end=>1000,n_steps=>1000});
```

Listing 2: A modeling including Gab2 positive feedback.

```
# Example Model 2 (with Gab2 positive feedback) from "An interaction library
for the Fc(epsilon)RI signaling network."
# L.A. Chylek, W.S. Hlavacek, D.A. Holowka, and B.A. Baird


begin molecule types

Lat(Y136~0~P,Y175~0~P)
Grb2(SH2,cSH3)
Gab2(PRS,PH,Y441~0~P)
PI3K(SH2,cat)
Plcg1(SH2,cat,Y783~0~P)
PI(OH3~0~P,OH4~0~P,OH5~0~P,head~0~1)
Btk(PH,cat)
Syk(cat)
Fyn(cat)
pre
zero
end molecule types

begin observables
# Lat pY175
Molecules pY175 Lat(Y175~P!?)

# Lat pY136
Molecules pY136 Lat(Y136~P!?)

# Active Plcg1
Molecules actPlcg Plcg1(SH2!+,Y783~P)

# Active Btk
Molecules Btkrec Btk(PH!+)

# Gab2 pY441
Molecules pGab Gab2(Y441~P!?)

# Lat-associated PI3K
Molecules PI3K_Lat PI3K().Lat()

# IP3
Molecules IP3    PI(OH3~0!?,OH4~P,OH5~P,head~0!?)

# PIP3
Molecules PIP3  PI(OH3~P!?,OH4~P,OH5~P,head~1!?)

# PIP2
Molecules PIP2  PI(OH3~0!?,OH4~P,OH5~P,head~1!?)


end observables
```

```
begin parameters

NA          6.022e23          # Avogadro's number; molecules/mole.
celldensity 1e9               # Cells/L
Fx                1                 # Fraction of cell volume to simuate;
unitless.
ECFvol          1/(celldensity)           # Extracellular volume; L/cell.
simECFvol       ECFvol*Fx           # Simulated fraction of extracellular
volume; L.
Cellvol         1.4e-12             # Cytoplasmic volume; L. Value from Fae-
der et al. (2003)
simCellvol  Cellvol*Fx        # Simulated fraction of cell volume; L.
ProteinTot  3e5               # Protein copy number per cell.
SimProteinTot     Fx*ProteinTot           # Simulated fraction of protein
copy number.
ActiveKinase1 0.5*SimProteinTot
ActiveKinase2 0.1*SimProteinTot
ActiveKinase3 0.05*SimProteinTot
kcSyk 100
kcFyn 100
KD1  79e-9*NA*simCellvol
KD2 62e-9*NA*simCellvol
krSH2 0.12
kfSH2_1 krSH2/KD1
kfSH2_2 krSH2/KD2
kfPH 1.4e6/(NA*simCellvol)
krPH 1
kfSH3 1e6/(NA*simCellvol)
kfSH3_3D 1e6/(NA*simCellvol)
kfSH3_2D 100*kfSH3_3D
krSH3 1
kcBtk 1e-3
kfP 1e9/(NA*simCellvol)
krP 1
kcP 10
kDeg 0.1
kPten 0.1
kX 5e-3
end parameters

begin functions
# A logistic growth function is used to model replenishment of PIP2 through,
e.g., activity of PIP5 kinase.
Replenish() = kX*PIP2*(1-PIP2/SimProteinTot)

end functions


begin species

# All molecules except active Syk and active Fyn are assumed to be present at
the same concentration
Lat(Y136~0,Y175~0) SimProteinTot
Grb2(SH2,cSH3) SimProteinTot
Gab2(PRS,PH,Y441~0) SimProteinTot
```

```
PI3K(SH2,cat) SimProteinTot
Plcg1(SH2,cat,Y783~0) SimProteinTot
PI(OH3~0,OH4~P,OH5~P,head~1) SimProteinTot
Btk(PH,cat) SimProteinTot
pre SimProteinTot

# Active Syk and active Fyn have an initial concentration of 0, which is
changed after a period of equilibration.
# See simulation commands at end of file for more information.
Syk(cat) 0
Fyn(cat) 0

end species

begin reaction rules

# Syk phosphorylates LAT Y136
Lat(Y136~0) + Syk(cat) <-> Lat(Y136~0!1).Syk(cat!1) kfSH3,krSH3
Lat(Y136~0!1).Syk(cat!1) ->Lat(Y136~P) + Syk(cat)  kcSyk

# Syk phosphorylates LAT Y175
Lat(Y175~0) + Syk(cat) <-> Lat(Y175~0!1).Syk(cat!1) kfSH3,krSH3
Lat(Y175~0!1).Syk(cat!1) ->Lat(Y175~P) + Syk(cat)  kcSyk

# Lat pY136 binds Plcg1
Lat(Y136~P) + Plcg1(SH2) <-> Lat(Y136~P!1).Plcg1(SH2!1) kfSH2_2,krSH2

# Lat pY175 binds Grb2
Lat(Y175~P) + Grb2(SH2) <-> Lat(Y175~P!1).Grb2(SH2!1) kfSH2_1,krSH2

# Fyn phosphorylates Gab2
#   Binding Gab2.Grb2.Lat, Gab2 PH domain free
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,PH,Y441~0) + Fyn(cat) -
>Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,PH,Y441~0!3).Fyn(cat!3)  kfSH3
#   Binding Gab2.Grb2.Lat, Gab2 PH domain bound
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,PH!+,Y441~0) + Fyn(cat) -
>Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,PH!+,Y441~0!3).Fyn(cat!3)  kfSH3
#   Binding Gab2.PIP3, Gab2 PRS free
Gab2(PRS,PH!+,Y441~0) + Fyn(cat) -> Gab2(PRS,PH!+,Y441~0!1).Fyn(cat!1) kfSH3
#   Binding Gab2.PIP3, Gab2 PRS bound to Grb2, but Grb2 not bound to Lat
Grb2(SH2,cSH3!2).Gab2(PRS!2,PH!+,Y441~0) + Fyn(cat) ->
Grb2(SH2,cSH3!2).Gab2(PRS!2,PH!+,Y441~0!1).Fyn(cat!1) kfSH3

#   Unbinding
Gab2(Y441~0!1).Fyn(cat!1) -> Gab2(Y441~0) + Fyn(cat) krSH3

#   Phosphorylation
Gab2(Y441~0!1).Fyn(cat!1) -> Gab2(Y441~P) + Fyn(cat) kcFyn

# Gab2 associates with Grb2
#   Grb2 is cytosolic and Gab2 is membrane associated
Grb2(SH2,cSH3)+Gab2(PRS,PH!+)->Grb2(SH2,cSH3!1).Gab2(PRS!1,PH!+) kfSH3_3D
#   Grb2 is membrane associated and Gab2 is cytosolic
Grb2(SH2!+,cSH3)+Gab2(PRS,PH)->Grb2(SH2!+,cSH3!1).Gab2(PRS!1,PH) kfSH3_3D
#   Grb2 is cytosolic and Gab2 is cytosolic
Grb2(SH2,cSH3)+Gab2(PRS,PH)->Grb2(SH2,cSH3!1).Gab2(PRS!1,PH) kfSH3_3D
```

```
#   Grb2 is membrane associated and Gab2 is membrane associated
Grb2(SH2!+,cSH3)+Gab2(PRS,PH!+)->Grb2(SH2!+,cSH3!1).Gab2(PRS!1,PH!+) kfSH3_2D

# Gab2 dissociates from Grb2
Grb2(cSH3!1).Gab2(PRS!1)->Grb2(cSH3)+Gab2(PRS) krSH3

# Gab2 binds PI3K
Gab2(Y441~P) + PI3K(SH2) <-> Gab2(Y441~P!1).PI3K(SH2!1) kfSH2_1,krSH2

# Gab2 binds PIP3
Gab2(PH) + PI(OH3~P,OH4~P,OH5~P,head~1) <->
Gab2(PH!1).PI(OH3~P,OH4~P,OH5~P,head~1!1) kfPH,krPH

# PI3K generates PIP3
#   PI3K.Gab2.Grb2.Lat binds PIP3. Gab2 PH is *not* bound
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,PH,Y441~P!3).PI3K(SH2!3,cat) +
PI(OH3~0,OH4~P,OH5~P,head~1) ->
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,PH,Y441~P!3).PI3K(SH2!3,cat!4).PI
(OH3~0!4,OH4~P,OH5~P,head~1) kfP
#   PI3K.Gab2.Grb2.Lat binds PIP3. Gab2 PH *is* bound.
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,PH!+,Y441~P!3).PI3K(SH2!3,cat) +
PI(OH3~0,OH4~P,OH5~P,head~1) ->
Lat(Y175~P!1).Grb2(SH2!1,cSH3!2).Gab2(PRS!2,PH!+,Y441~P!3).PI3K(SH2!3,cat!4).
PI(OH3~0!4,OH4~P,OH5~P,head~1) kfP
#   PI3K.Gab2.PIP3 binds PIP3. Gab2 PRS is not bound.
Gab2(PRS,PH!+,Y441~P!3).PI3K(SH2!3,cat) + PI(OH3~0,OH4~P,OH5~P,head~1) ->
Gab2(PRS,PH!+,Y441~P!3).PI3K(SH2!3,cat!2).PI(OH3~0!2,OH4~P,OH5~P,head~1) kfP
#   PI3K.Gab2.PIP3 binds PIP3. Gab2 PRS is bound but Grb2 is not bound to
Lat.
Grb2(SH2,cSH3!2).Gab2(PRS!2,PH!+,Y441~P!3).PI3K(SH2!3,cat) +
PI(OH3~0,OH4~P,OH5~P,head~1) ->
Grb2(SH2,cSH3!2).Gab2(PRS!2,PH!+,Y441~P!3).PI3K(SH2!3,cat!4).PI(OH3~0!4,OH4~P
,OH5~P,head~1) kfP

#   Unbinding
PI3K(cat!4).PI(OH3~0!4,OH4~P,OH5~P,head~1) -> PI3K(cat) +
PI(OH3~0,OH4~P,OH5~P,head~1) krP

#   Phosphorylation
PI3K(cat!4).PI(OH3~0!4,OH4~P,OH5~P,head~1) -> PI3K(cat) +
PI(OH3~P,OH4~P,OH5~P,head~1) kcP

# Btk binds PIP3
PI(OH3~P,OH4~P,OH5~P,head~1) + Btk(PH) <->
PI(OH3~P,OH4~P,OH5~P,head~1!1).Btk(PH!1) kfPH,krPH

# Btk phosphorylates Plcg1
#   Binding
Btk(PH!+,cat) + Plcg1(SH2!+,Y783~0) -> Btk(PH!+,cat!1).Plcg1(SH2!+,Y783~0!1)
kfSH3
#   Unbinding
Btk(cat!1).Plcg1(Y783~0!1) -> Btk(cat) + Plcg1(Y783~0) krSH3
#   Phosphorylation
Btk(cat!1).Plcg1(Y783~0!1) -> Btk(cat) + Plcg1(Y783~P) kcSyk
```

```
# Plcg1 generates IP3
#   Binding
Plcg1(cat,SH2!+,Y783~P) + PI(OH3~0,OH4~P,OH5~P,head~1) ->
Plcg1(cat!1,SH2!+,Y783~P).PI(OH3~0,OH4~P,OH5~P,head~1!1) 10*kfP
#   Unbinding
Plcg1(cat!1).PI(head~1!1) -> Plcg1(cat) + PI(head~1) krP
#   Catalysis
Plcg1(cat!1).PI(head~1!1) -> Plcg1(cat) + PI(head~0) kcP


# IP3 is cleared
PI(OH3~0,OH4~P,OH5~P,head~0) -> zero kDeg

# PIP3 is dephosphorylated to PIP2
PI(OH3~P,OH4~P,OH5~P,head~1) -> PI(OH3~0,OH4~P,OH5~P,head~1) kPten

# PIP2 is replenished. "pre" represents a precursor molecule. Replenish() is
a functional rate law defined in the "functions" block above.
pre -> pre + PI(OH3~0,OH4~P,OH5~P,head~1) Replenish()

# Dephosphorylation reactions
Lat(Y175~P) -> Lat(Y175~0) 5
Lat(Y136~P) -> Lat(Y136~0) 5
Gab2(Y441~P) -> Gab2(Y441~0) 5
Plcg1(Y783~P) -> Plcg1(Y783~0) 5
end reaction rules

# Generate reaction network from rule set
generate_network({overwrite=>1});

# Execution of following command performs equilibration before Syk and Fyn
are added as inputs
simulate_ode({t_end=>1000,n_steps=>1000});

# The following command sets the Syk concentration. The different concentra-
tions considered were ActiveKinase1, ActiveKinase2, and ActiveKinase3, which
are specified in the parameters block above.
setConcentration("Syk(cat)","ActiveKinase1");

# The following command sets the Fyn concentration. Only one Fyn concentra-
tion (ActiveKinase1) was considered.
setConcentration("Fyn(cat)","ActiveKinase1");

# The following command simulates system behavior after addition of Syk and
Fyn inputs
simulate_ode({suffix=>stim1,t_end=>1000,n_steps=>1000});
```

References

Barua D, Hlavacek WS, Lipniacki T. A computational model for early events in B cell antigen receptor signaling: analysis of the roles of Lyn and Fyn. J Immunol (2012) 189:646–58. doi:10.4049/jimmunol.1102003

Das J, Ho M, Zikherman J, Govern C, Yang M, Weiss A, et al. Digital signaling and hysteresis characterize ras activation in lymphoid cells. Cell (2009) 136:337–51. doi:10.1016/j.cell.2008.11.051

Faeder, J.R., Blinov, M.L., Hlavacek, W.S. Rule-based modeling of biochemical systems with BioNetGen. Methods Mol Biol. (2009) 500, 113-67.

Gadi, D., Wagenknecht-Wiesner, A., Holowka, D., Baird, B. Sequestration of phosphoinositides by mutated MARCKS effector domain inhibits stimulated Ca(2+) mobilization and degranulation in mast cells. Mol Biol Cell. (2011) 22, 4908-17.

Halet G. Imaging phosphoinositide dynamics using GFP-tagged protein domains. Biol Cell (2005) 97:501–18. doi:10.1042/BC20040080

Kholodenko, BN. Cell-signalling dynamics in time and space. Nat Rev Mol Cell Biol. (2006) 7:165-76.

Lipniacki T, Hat B, Faeder JR, Hlavacek WS. Stochastic effects and bistability in T cell receptor signaling. J Theor Biol (2008) 254:110–22. doi:10.1016/j.jtbi. 2008.05.001

Monine MI, Posner RG, Savage PB, Faeder JR, Hlavacek WS. Modeling multi-valent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. Biophys J (2010) 98:48–56. doi:10.1016/j.bpj. 2009.09.043

Mukherjee S, Zhu J, Zikherman J, Parameswaran R, Kadlecek TA, Wang Q, et al. Monovalent and multivalent ligation of the B cell receptor exhibit differential dependence upon Syk and Src family kinases. Sci Signal (2013) 6:ra1.

Parravicini V, Gadina M, Kovarova M, Odom S, Gonzalez-Espinosa C, Furumoto Y, et al. Fyn kinase initiates complementary signals required for IgE-dependent mast cell degranulation. Nat Immunol (2002) 3:741–8. doi:10.1038/ni817

Sekar JAP, Tapia JJ, Faeder JR. Visualizing regulation in rule-based models. (2015) arXiv:1509.00896.

Shelby SA, Holowka D, Baird B, Veatch SL. Distinct stages of stimulated FcεRI receptor clustering and immobilization are identified through superresolution imaging. Biophys J (2013) 105:2343–54. doi:10.1016/j.bpj.2013.09.049

Sil D, Lee J, Luo D, Holowka D, Baird B. Trivalent ligands with rigid DNA spacers reveal structural requirements for IgE receptor signaling in RBL mast cells. ACS Chem Biol (2007) 2:674–84. doi:10.1021/cb7001472

Tolar P, Sohn HW, Pierce SK. The initiation of antigen-induced B cell antigen receptor signaling viewed in living cells by fluorescence resonance energy transfer. Nat Immunol. (2005) 6:1168-76.

Tsai TY, Choi YS, Ma W, Pomerening JR, Tang C, Ferrell JE Jr. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. Science. (2008) 321:126-9.

Wilkes MM, Wilson JD, Baird B, Holowka D. Activation of Cdc42 is necessary for sustained oscillations of Ca2+ and PIP2 stimulated by antigen in RBL mast cells. Biol Open. (2014) 3:700-10.

Chapter 3: Frequency response properties of FcεRI signaling [1]

Abstract

We used microfluidics to start and stop signaling from the high-affinity receptor for IgE on mast cells via alternating streams of medium containing either a stimulatory multivalent or non-stimulatory monovalent antigen. Initial exposure to multivalent antigen (10 nM) elicits a robust secretory response, which ends abruptly upon subsequent exposure to excess monovalent antigen, which halts receptor signaling by breaking up receptor clusters. Surprisingly, the secretory response to a second period of multivalent antigen stimulation depends on the interval of monovalent antigen-induced signaling quiescence. For intervals of monovalent antigen exposure less than 60 min, the second response to stimulation is blunted relative to the first response. The opposite was observed for longer intervals. Through coordinated modeling and experimentation, we were able to attribute these history-dependent responses to antigen stimulation to timescale separation of fast Syk-driven positive signaling, slow Ship1-driven negative signaling, and slower proteasome-mediated degradation of Ship1 co-factors, including Shc1.

Introduction

The high-affinity receptor for IgE is found on the surface of mast cells and basophils, where it can interact with antigens, via bound IgE antibodies, and trigger secretion of

[1] This work is currently in review at *Science Signaling* as the following: Chylek LA,* Harmon B,* Liu Y,* Mahajan A, Schudel BR, Holowka DA, Baird BA, Wilson BS, Hlavacek WA, Singh AK. IgE receptor signaling encodes memories of antigen exposure.
*These authors contributed equally.

inflammatory mediators that contribute to allergic reactions. The receptor controls cellular responses through a signaling network whose proteins and lipids undergo time-dependent changes in abundance and phosphorylation. To characterize the timescales on which these changes occur, mast cells were exposed to pulsatile stimulation by a multivalent antigen in a microfluidic device. Two stimulatory pulses were separated by an interval of exposure to a monovalent, non-stimulatory antigen that halts signaling. We varied the non-stimulatory interval and quantified the effect on cumulative secretion during stimulation, thus characterizing the frequency response of the IgE receptor signaling network. When the non-stimulatory interval was short (<60 min), secretory responses to a second stimulation were diminished. For a 60 min interval, the response to a second stimulation was equivalent to the initial response. For intervals greater than 60 min, responses to the second stimulation were enhanced relative to the initial response. Through an iterative process of computational modeling and experimental tests, we found that these memory-like phenomena arise from a confluence of rapid, short-lived positive signals driven by the protein tyrosine kinase Syk; slow, long-lived negative signals driven by the lipid phosphatase Ship1; and slower degradation of Ship1 co-factors, including Shc1. Our results show that signaling processes occurring on different timescales give rise to history-dependent cellular responses.

Introduction

Central players in inflammation and allergic reactions include mast cells and basophils, which upon stimulation with a multivalent antigen, release histamine and other

inflammatory mediators in a process called degranulation. Stimulation occurs when a multivalent antigen induces aggregation of the high affinity receptor for IgE, also known as FceRI. Receptor aggregation leads to activation of several kinases, including the protein tyrosine kinase Syk, which phosphorylates an array of downstream targets to promote degranulation. Positive signals for degranulation generated by FceRI and Syk are held in check by negative regulatory processes (*1*). The dynamic interplay between positive and negative signals influences how a cell responds to inputs. Complex input waveforms, such as an antigen whose concentration varies over time, offer a means to elucidate signaling dynamics that can give rise to seemingly enigmatic phenomena, such as desensitization.

Desensitization can arise with repeated exposure to an antigen (*2,3,4,5,6,7*). A mast cell that has undergone nonspecific desensitization will show attenuated responses to an antigen it has previously encountered, as well as other antigens. Mechanisms inducing nonspecific desensitization are likely to operate at the level of receptor-proximal signaling because antigen stimulation of primary human mast cells dampens responses to an unrelated antigen, without affecting secretagogues that bypass the receptor (*8*). Several proteins, including the lipid phosphatases Ship1 (Inpp5d) and Pten and the protein tyrosine phosphatase Shp1 (Ptpn6), have been implicated in negative regulation of mast cell signaling (9), but the molecular processes governing desensitization have yet to be fully characterized. This is due in part to the technical challenge of exposing cells to stimuli that change over time. However, the question of how complex inputs affect cellular outputs can now be addressed with microfluidic devices.

Microfluidic technology allows for precise manipulation of fluids at timescales of seconds. This capability can be leveraged to expose single cells to complex waveform inputs, such as pulsatile, ramp, square-wave and sinusoidal stimuli. Indeed, microfluidic devices have been used to produce periodic stimuli to measure the frequency dependence of signal processing in the osmo-adaptation pathway of yeast (*10*), to quantify the bandwidth of the HOG MAP pathway in yeast (*11*), and to characterize responses of amoebae to pulses of chemoattractant (*12*).

Here, we used a microfluidic chip to characterize the frequency response of an antigen receptor signaling system that plays an important role in immunity. We find that the frequency response properties of the system allows antigen exposure (for a finite time) to either desensitize cells or to prime cells for a hyperactive response to a second exposure to antigen.

Results

Exposing mast cells to complex waveform inputs

To expose cells to alternating environments of stimulation and input quiescence, cells were incubated in a microfluidic device. The design of the device is illustrated in Fig. 1 (top panel). The chip has three inlets for loading cells, exchanging reagents, and buffer washing, as well as two outlets for collecting secreted material, immuno-stained cells, and waste. The channel's serpentine design minimizes dead volume and maximizes the effective surface area for seeding of cells. The microfluidic chip is integrated with miniaturized electronic valves, optical elements, actuated pressure controllers, and data

acquisition software, forming a self-contained platform that allows for precise control of the microenvironment of single cells and measurement of cellular responses to environmental perturbations. Significantly, a complete exchange of media/reagents can be accomplished within less than 20 seconds.

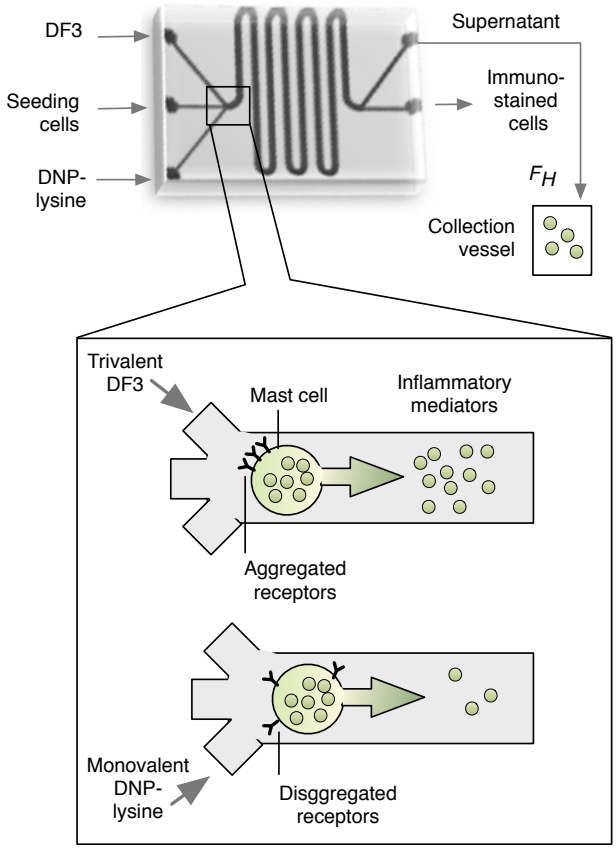


Figure 3.1. A microfluidic device for activation and deactivation of IgE receptor (FcεRI) signaling in mast cells. Top: An illustration of the microfluidic device with inlets, outlets (F_H is the exit flow rate), and serpentine channels. Bottom: DF3, a trivalent DNP ligand, induces aggregation of FcεRI via interaction with FcεRI-bound anti-DNP IgE, leading to release of inflammatory mediators including β-hexosaminidase. In contrast, monovalent DNP-lysine induces breakup of aggregates, thereby halting FcεRI signaling and attenuating release of inflammatory mediators.

We used the device to expose cells to alternating flows of medium containing multivalent or monovalent ligands, which are stimulatory and non-stimulatory, respectively. A schematic of how a mast cell responds to these ligands is shown in Fig. 1 (bottom panel). As our multivalent ligand, we chose the trivalent DNP ligand, DF3, because it is symmetric, structurally well-defined, and a potent secretagogue (*13*). DF3 was introduced through the microfluidic stream to crosslink DNP-specific IgE bound to cell-surface FcεRI and to initiate the intracellular signaling process. This process leads to release of inflammatory mediators within 1–2 min. A DF3 dose of 10 nM was chosen based on prior work determining that this dose is optimal for degranulation (*13*). Following exposure to DF3, exposure to monovalent DNP-lysine causes receptor aggregates to quickly break up, resulting in reversal of intracellular signaling (*14*). After ablation of receptor aggregation, DNP-lysine has no further effect because DNP-lysine does not induce FcεRI signaling (Fig. 2).

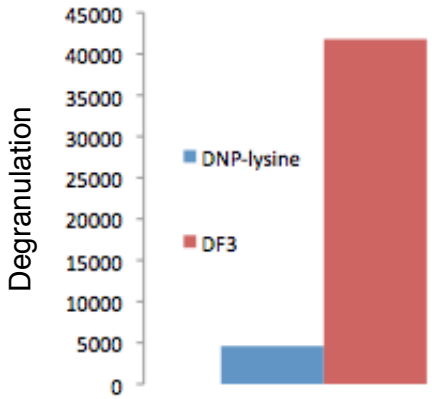


Figure 3.2. DNP-lysine does not induce degranulation. Cells were first exposed to a 5-min pulse of DNP-lysine and then a 5-min pulse of DF3, and degranulation was measured during each of these periods. DNP-lysine did not induce appreciable degranulation.

Thus, for long intervals between pulses of multivalent ligand, medium only was used after 5 min of DNP-lysine exposure, which was sufficient to disaggregate receptors and halt antigen-induced degranulation (Fig. 3).

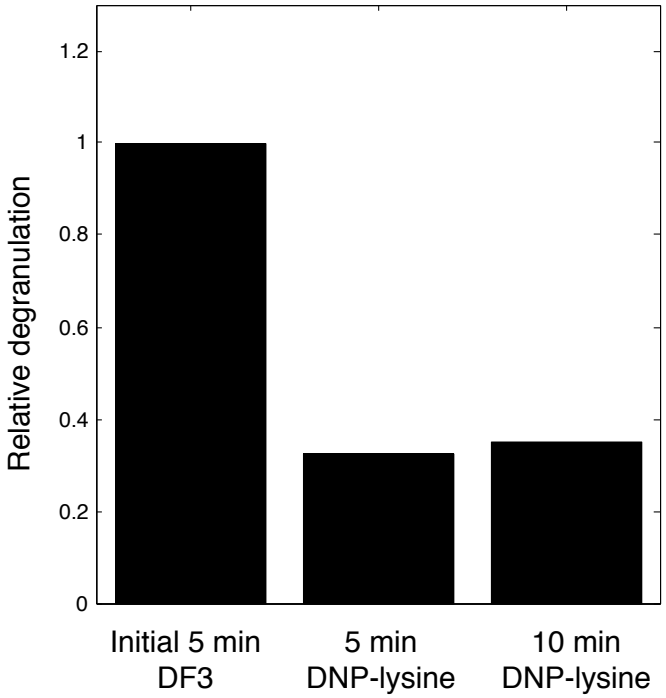


Figure 3.3. 5 min of DNP-lysine is sufficient to break up receptor aggregates and reduce degranulation. Degranulation was monitored during an initial 5 min pulse of DF3, and then during a pulse of DNP-lysine that lasted either 5 or 10 min. Note that 5 min and 10 min of DNP-lysine exposure result in similar degranulation levels.

The microfluidic device enabled us to rapidly change the environment of cells and to expose cells to abrupt pulses of stimulatory (DF3) and non-stimulatory (DNP-lysine) inputs. For example, in one protocol, we exposed cells to a 5-min pulse of DF3, followed by a 5-min pulse of DNP-lysine, and finally a second 5-min pulse of DF3. We characterized the magnitude of responses during the second DF3 pulse compared to the first. The responses that we measured were 1) cumulative amount of exocytosed β-hexosaminidase over a period of interest as measured by a β-hexosaminidase activity assay with a

fluorogenic substrate and 2) Syk phosphorylation at an activating site (Y346) (*16*) as measured by flow cytometry.

Our nomenclature is introduced in Fig. 4A. The duration of the initial pulse of DF3 is designated *S1*. The duration of the subsequent pulse of DNP-lysine is designated *I*, and the second pulse of DF3 is designated *S2*. The amount of degranulation/secretion occurring within a given time interval is designated *H*. In the experiments described below, we measured *H(S1)*, implemented a range of durations for *I*, and determined how *I* impacts *H(S2)*.

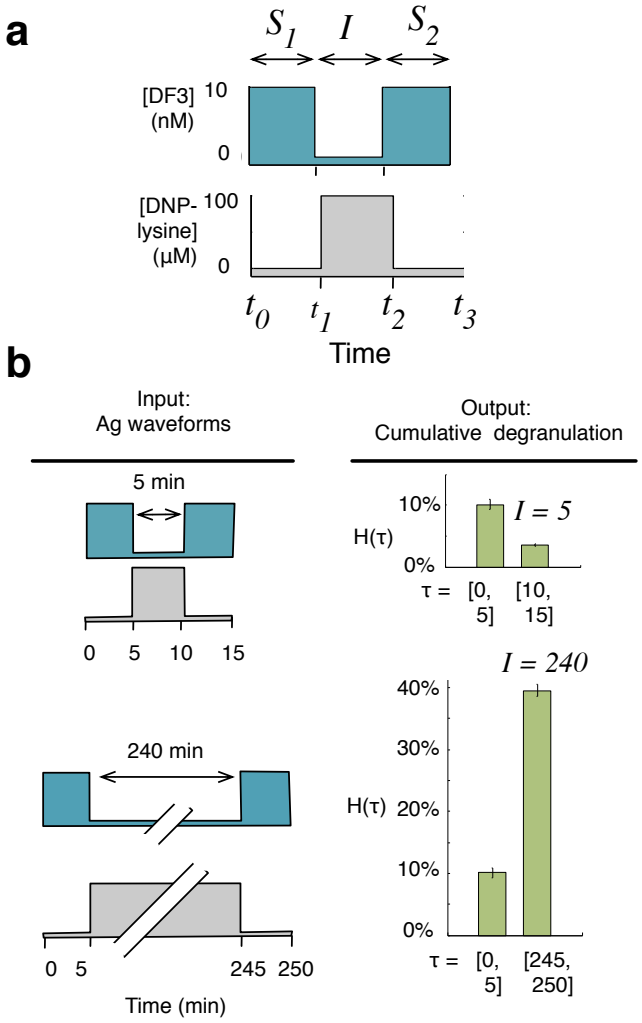


Figure 3.4. Responses to complex waveform inputs, implemented with the microfluidic device of Fig. 1. **a)** Between times t_0 and t_1, cells were exposed to a pulse of 10 nM DF3. This period is designated *S1*. Between times t_1 and t_2, cells were exposed to 1 μM DNP-lysine. This period is designated *I*. Finally, between times t_2 and t_3, cells were exposed to a second pulse of 10 nM DF3, designated *S2*. **b)** Responses during *S2* depend on the length of *I* (compare top and bottom panels). We measured degranulation during *S1* and *S2*, represented by *H(S1)* and *H(S2)*. As can be seen at top right in panel b, with $I = 5$ min, $H(S2) < H(S1)$. In contrast, as can be seen at bottom right in panel b, with $I = 240$ min, $H(S2) > H(S1)$.

Mast cells exhibit short- and long-term memory

We initially considered an *S1* duration of 5 min. Interestingly, we found that when *I* was short, e.g., 5 min, *H(S2)* was greatly diminished relative to *H(S1)* (Fig. 4B, top panel); i.e., the cells exhibited desensitization or "short-term memory." To determine how the duration of the first pulse influences desensitization, we compared results after varying the duration of the first exposure to DF3 (30 s – 5 min).

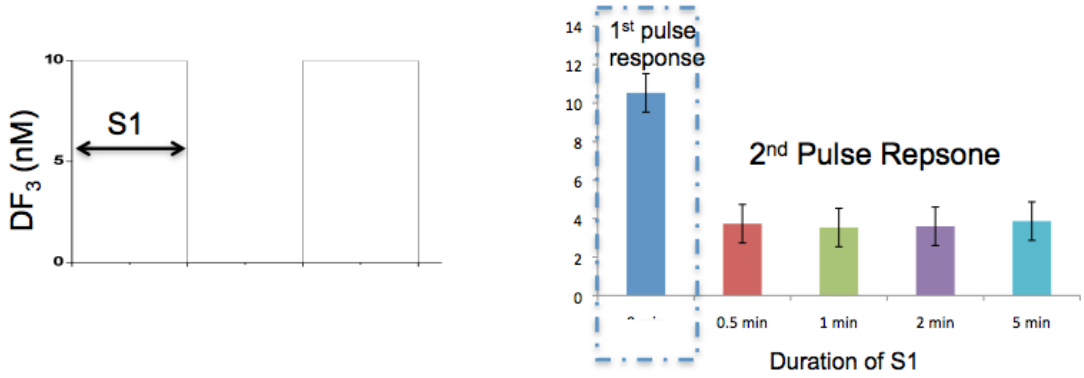


Figure 3.5. The impact of *S1* duration on responses during *S2*. The duration of *S1* was varied from 30 s to 5 min. The durations of *I* and *S2* were both 5 min in all cases. Desensitization occurs to a similar extent for all *S1* durations: the response during *S2* is alike for all of the *S1* durations considered.

As shown in Figure 5, initial exposures as short as 30 s, followed by a 5 min exposure to monovalent ligand, induced desensitization to a second pulse of the trivalent ligand. This phenomenon also occurs if cells are sensitized to both dansyl and DNP and then sequentially exposed to dansyl and DNP ligands (Figure 6). Thus, the desensitization process found here is nonspecific and insensitive to the duration of initial antigen exposure.

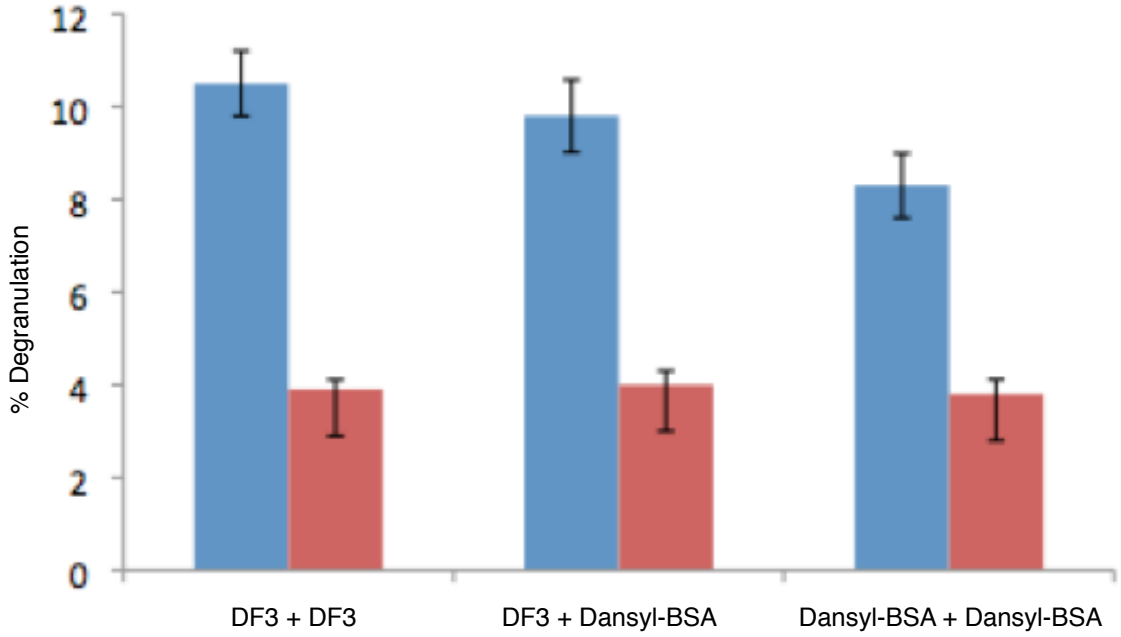


Figure 3.6. Desensitization is nonspecific. Cells were sensitized with half-and-half mixture of anti-DNP IgE, and anti-dansyl IgE. Cells were exposed to DF3 or a dansyl antigen in the indicated order. For all of these experiments, *S1* = 5 min, *I* = 5 min, and *S2* = 5 min. The blue bar represents the response to the first pulse and the red bar represents the response to the second pulse. Exposure to one antigen promotes desensitization to the other.

We also subjected cells to continuous stimulation with DF3 and found that degranulation continues to rise for as long as 10 min (Figure 7). Receptor internalization was not dramatic over the timescales considered (Figure 8). Thus, desensitization does not result from depletion of the cell's supply of β-hexosaminidase or reduction of receptors at the membrane, and instead arises from internal changes that occur after stimulation by DF3.

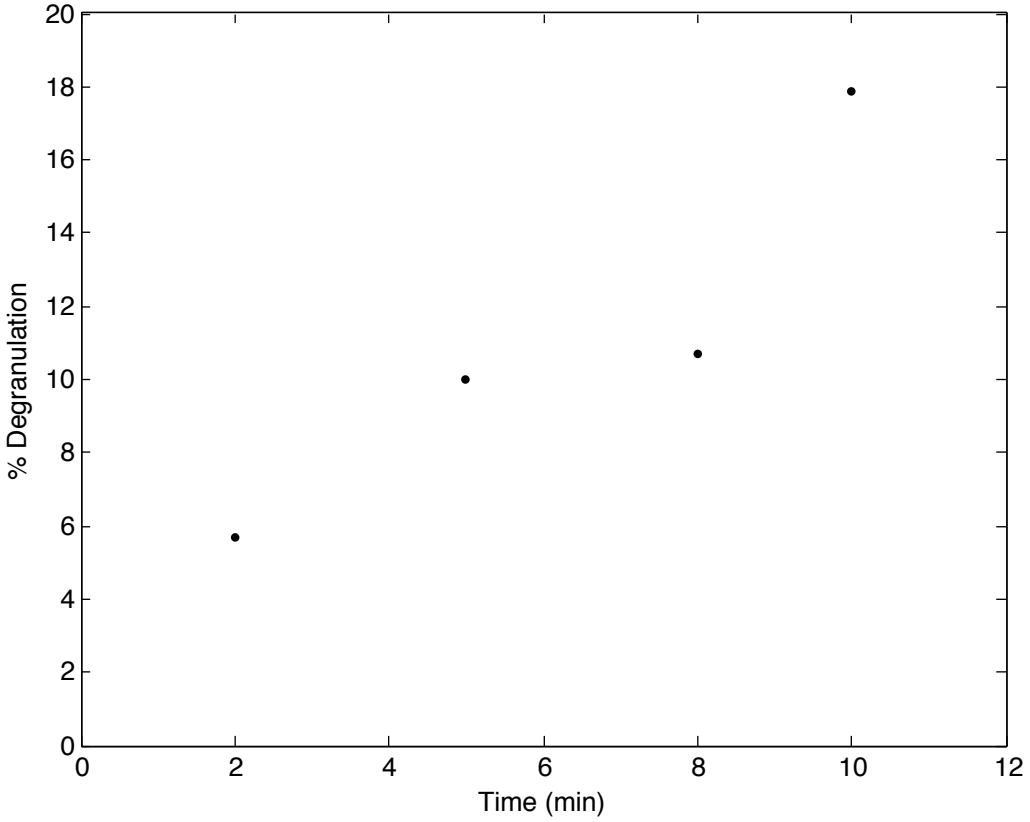


Figure 3.7. Degranulation during continuous DF3 stimulation. Cells were incubated with 10 nM DF3 for 2, 5, 8, and 10 min, and cumulative degranulation was measured during this time. The increase in degranulation over this time period indicates that desensitization is not the result of a shortage of granules.

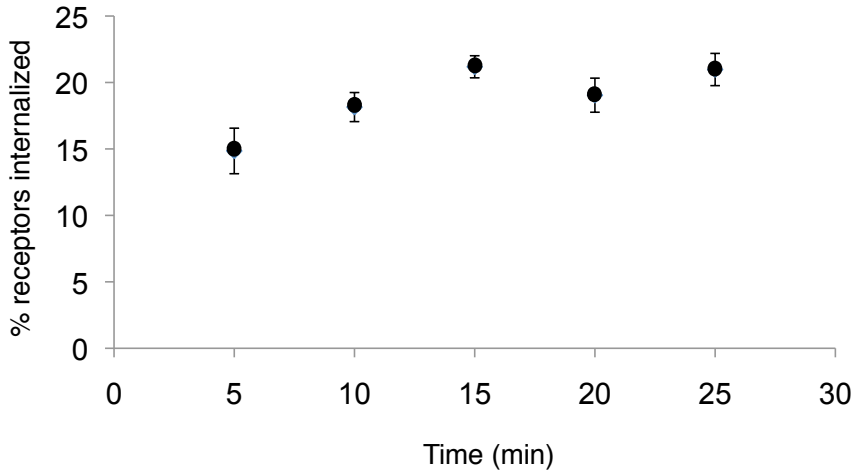


Figure 3.8. Receptor internalization over time following DF3 stimulation.

In contrast to the observations reported above, when *I* was long, e.g., 240 minutes, *H(S2)* was much greater than *H(S1)* (Fig. 4B, bottom panel). We term this phenomenon "long-term memory" or priming. These results show that initial desensitization is not a static condition and suggest that the signaling state within the cell changes over the course of *I* to determine the magnitude of subsequent degranulation responses. Thus, the memory phenomena are dynamical.

Ship1 activates short-term memory and the proteasome activates long-term memory

To understand the signaling dynamics that give rise to memory, we sought to perturb short- and long-term memory at the molecular level (Fig. 9A–C), and to incorporate this information into a model of the system (Fig. 9D–G) that could reproduce experimental measurements of degranulation for a range of *I* values.

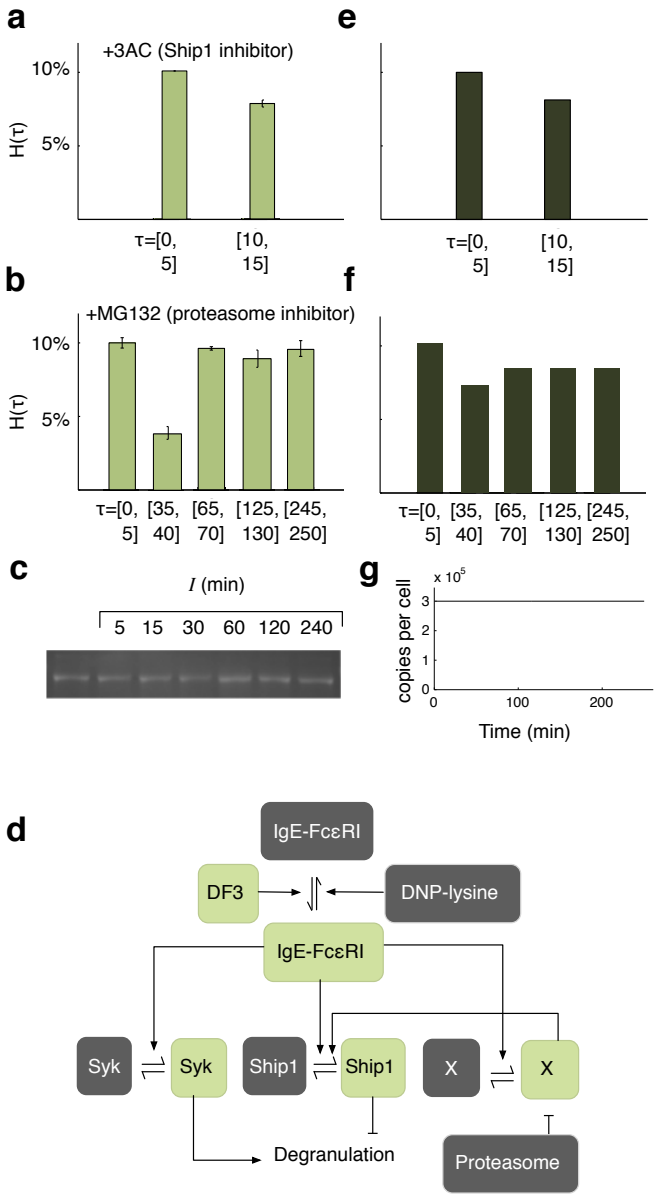


Figure 3.9. A mathematical model to explain the phenomena of desensitization and priming. A) Degranulation in the presence of Ship1 inhibitor. B) Degranulation in the presence of a proteasome inhibitor. C) Measured abundance of Ship1 over time. Cell density was normalized. D) Diagram of the mathematical model. Activation of IgE-FcεRI by DF3 triggers activation of Syk, Ship1, and the Ship1 co-factor *X*. Syk and Ship1 exert opposing influences on degranulation. The proteasome degrades *X*. E) Simulated degranulation in the presence of a Ship1 inhibitor. F) Simulated degranulation in the presence of a proteasome inhibitor. G) Simulated abundance of Ship1 over time.

We first considered which negative regulators might bring about desensitization.

The lipid phosphatase Ship1 is known to be important in downregulation of signaling via

FcεRI (*9*). To assess the role of Ship1 in desensitization, we treated cells with the Ship1-specific inhibitor 3AC (*17*). This inhibitor does not affect cell viability (Fig. 10). Cells then underwent a stimulation pattern with $S1$ = 5 min, I = 5 min and $S2$ = 5 min. We found that $H(S2)$ for 3AC-treated cells was greater than $H(S2)$ for untreated cells. Indeed, the second secretory response of treated cells was nearly identical to their initial response to DF3 (Fig. 9A). These results implicate Ship1 as having a role in desensitization. We also evaluated the effect of NSC, an inhibitor of the tyrosine phosphatase Shp1, a protein that has been shown to positively and negatively regulate mast cell degranulation (*9*). In contrast with inhibition of Ship1, inhibition of Shp1 did not attenuate the priming phenomenon (Fig. 11).

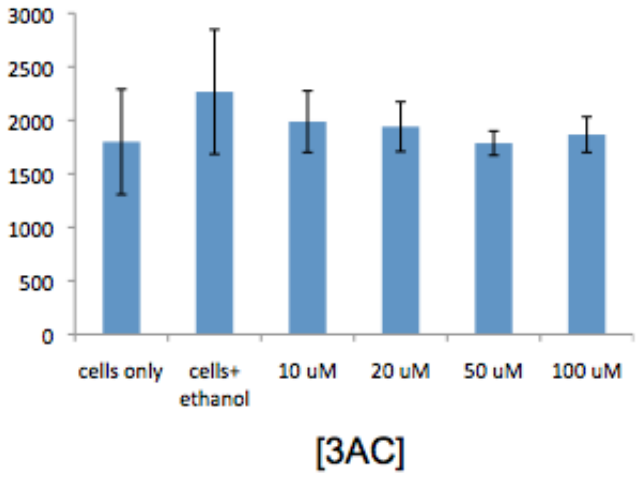


Figure 3.10. The effect of the Ship1 inhibitor, 3AC, concentration on cell viability. The inhibitor does not adversely affect viability compared to untreated cells, for the concentrations considered.

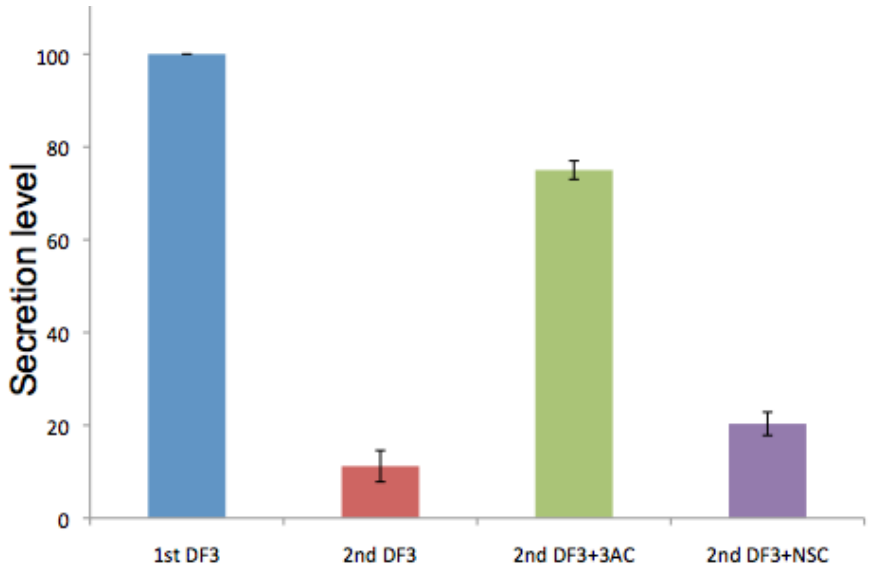


Figure 3.11. Inhibition of the tyrosine phosphoatase Shp1 (Ptpn6) by NSC has a minimal effect on desensitization. Cells were treated with NSC and then exposed to a stimulation pattern of *S1* = 5 min, *I* = 5 min, and *S2* = min. Degranulation was monitored during the two stimulatory periods for the untreated cells (1st and 2nd panels) and NSC-treated cells (3rd and 4th panels). We then considered what processes could counterbalance the effect of desensitization to bring about long-term memory. We hypothesized that protein degradation could be involved. To test this hypothesis, we treated cells with the proteasome inhibitor MG132. We found that long-term memory was impaired: unlike the case of the untreated cells, *H(S2)* after long *I* did not exceed *H(S1)* (Fig. 9B). Thus, the proteasome inhibits the desensitization induced by Ship1.

We assessed whether these two players, Ship1 and the proteasome, are directly linked; that is, whether Ship1 is degraded by the proteasome. Surprisingly, we found that Ship1 was not degraded following stimulation (Fig. 9C), indicating that the proteasome counteracts Ship1 activity in an indirect way. A possible mechanism of indirect negative regulation is proteasome-mediated degradation of one or more activators or co-factors of Ship1.

A computational model predicts signaling dynamics

To investigate the mechanisms controlling desensitization and priming, we constructed a mathematical model of the FcεRI signaling network based on known

mechanisms of FcεRI signaling, principles of chemical kinetics, and the experimental clues that we had generated so far: 1) Ship1 negatively regulates signaling and 2) the proteasome negatively but indirectly regulates Ship1. In formulating the model, we hypothesized that a co-factor of Ship1, which we termed protein *X*, is degraded by the proteasome. The model is illustrated in Fig. 9D. Model parameters were tuned for consistency with experimental data characterizing the magnitude of *H(S2)* measurements for a range of *I* values (Fig. 12). We used the calibrated model to simulate desensitization in the presence of Ship1 inhibition (Fig. 9E), priming in the presence of proteasome inhibition (Fig. 9F), and Ship1 abundance over time (Fig. 9G), and found consistency with experimental results (cf. panels E and F in Fig. 9 with panels A and B, respectively).

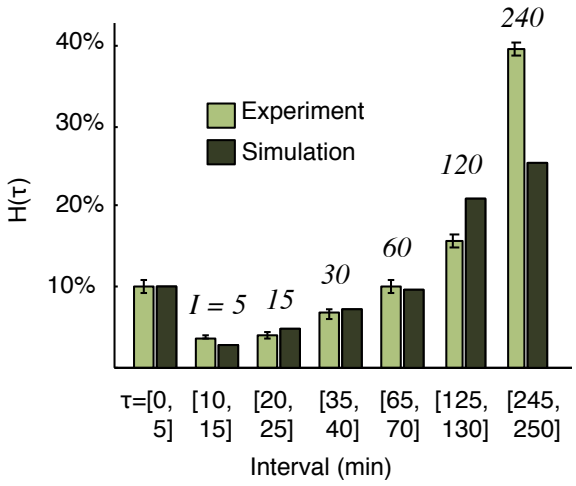


Figure 3.12. Simulated vs. measured degranulation during a second pulse of DF3, *H(S2)*, as a function of *I* values.

Through extensive simulation-based analysis of the model, we found that the ability of the model to reproduce the experimental data depends on the following assumptions, which are incorporated in the model: 1) Syk is activated and deactivated

quickly, 2) Ship1 is activated and deactivated slowly, and 3) the hypothesized co-factor of Ship1, *X*, is degraded very slowly. Our model predicts that the tyrosine kinase Syk, a known driver of positive signaling, undergoes rapid activation upon exposure to DF3, and rapid deactivation upon exposure to DNP-lysine. Rapid Syk activation and deactivation kinetics had been hypothesized previously but without direct experimental confirmation (*20*). Our model also predicts that activation of Syk in response to a second DF3 pulse is largely history-independent, meaning that Syk is expected to be reactivated in a similar manner regardless of the length of *I* (Fig. 13A). In contrast, our model predicts a slower and history-sensitive regulation pattern for Ship1 (Fig. 13B). In the model, Ship1 activation and deactivation is more gradual, such that a second DF3 pulse after $I = 5$ min results in greater Ship1 activation during *S2*, because Ship1 retains a "head start." However, over time Ship1 gradually loses its capacity to be activated due to degradation of the Ship1 co-factor *X*. With $I = 60$ min, Ship1 activation is similar to the initial pulse. With $I = 240$ min, Ship1 is refractory to activation due to lack of the hypothesized Ship1 co-factor (Fig. 13B). In our model, this loss of negative signaling is responsible for the hyperdegranulation occurring for longer *I* values.

Syk and Ship1 are activated and deactivated on different timescales

To test these predictions, we assessed Syk regulation through flow cytometric measurement of Syk phosphorylation at Y346 (rat numbering). Phosphorylation of this site relieves Syk auto-inhibition (*15, 16*). We found that phosphorylation of Y346 rapidly rises with DF3 stimulation and rapidly falls with DNP-lysine exposure as measured by

both flow cytometry and western blotting (Fig. 13C, 14, and 15), consistent with model predictions. We tested the prediction of slow Ship1 activation and deactivation kinetics using a Ship activity assay. As expected, Ship1's ability to dephosphorylate lipids becomes increasingly refractory to repeated stimulation as *I* grows longer (Fig. 13D).

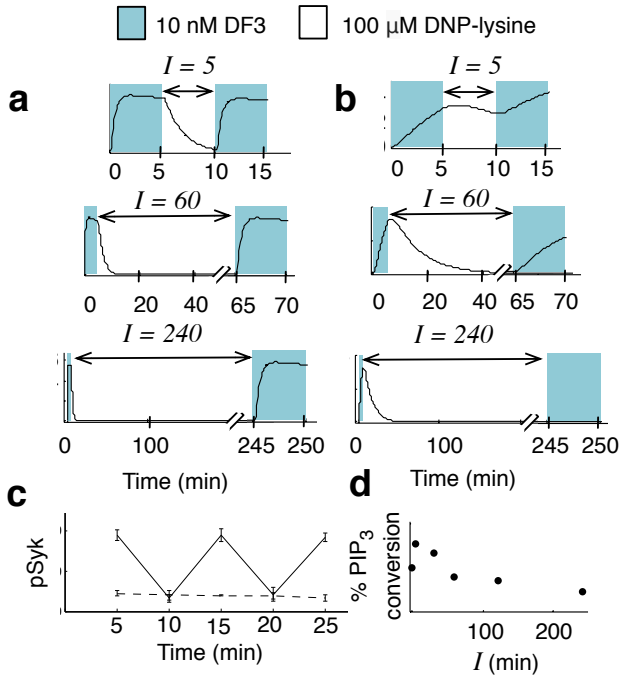


Figure 3.13. Simulated and measured dynamics of Syk and Ship1 activation/deactivation. A) Simulation of quantity of activated Syk molecules for varying *I* values. B) Simulation of quantity of activated Ship1 molecules for varying *I* values. C) Phosphorylation of Syk pY346, a readout of activation, as measured by flow cytometry during the same stimulation pattern as panel A (top). D) Activation of Ship1, as measured by a malachite green assay, upon DF3 stimulation following the indicated *I* values.

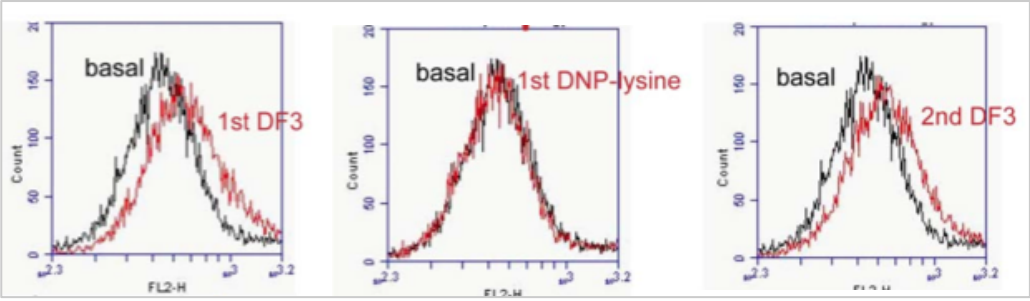


Figure pi. Histograms of flow cytometry measurements of Syk activation. Cells were exposed to a 5-min pulse of DF3, followed by a 5-min pulse of DNP-lysine, and then a second pulse of DF3. Measurements were taken at the end of each of these periods. These data are also summarized in Fig. 5D.

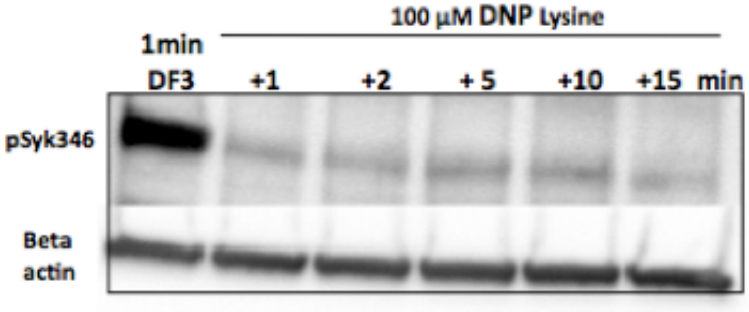


Figure 3.15. Kinetics of Syk deactivation. Cells were stimulated with DF3 (1st panel) and then exposed to DNP-lysine for the indicated time periods (2^{nd}-6^{th} panels). Phosphorylation of Syk Y346, a readout of Syk activation, was monitored using western blotting. Phosphorylation of this site decreases within 1 minute, indicating rapid deactivation.

A Shc1 knockdown accelerates long-term memory

Next, we sought to identify the hypothesized Ship1 co-factor that is degraded by the proteasome. We considered a set of proteins that are known to interact with and contribute to Ship1 activity.

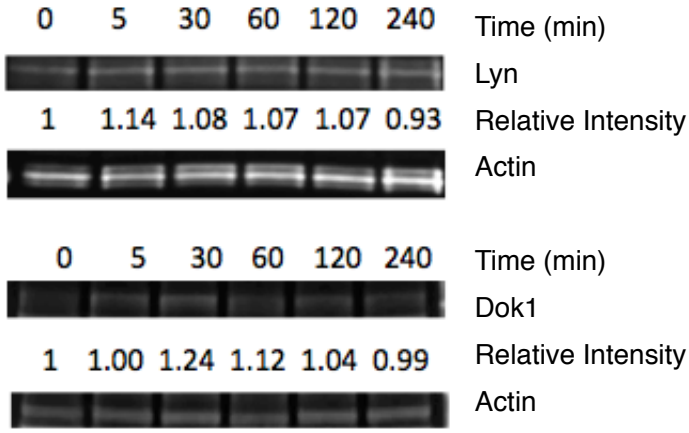


Figure 3.16. Abundances of Lyn (top) and Dok1 (bottom) over time following stimulation with DF3. Both of these proteins are implicated in activation of Ship1, but their steady abundances do not agree with the model's predicted time course of degradation for protein *X*.

The proteins that we were able to monitor by western blotting, such as Dok1 and Lyn, did not show changes in abundance following stimulation (Fig. 16), with one exception. Shc1 underwent noticeable degradation over time in agreement with the model-predicted dynamics of protein *X* (Fig. 17A). Interestingly, Ship1 is recruited to a ternary complex with Shc1 and the adaptor Lat (*18*), and Shc1 and Ship1 both interact with the β chain ITAM of FcεRI (*19*). Shc1 is also part of a ternary complex with Ship1 and PKC-δ, another negative regulator of FcεRI-induced mast cell signaling (*25*). We hypothesized that Shc1 may contribute to the priming and long-term memory phenomena.

To test this hypothesis, we considered the effect of a *Shc1* knockdown (KD). We first simulated the effect of the knockdown and found that hyperdegranulation is predicted to occur earlier (Fig. 17B). We then used siRNA to experimentally reduce the abundance of Shc1 in cells (Fig. 17C). Degranulation assays are consistent with our model's predictions, with hyperdegranulation arising earlier for the *Shc1* KD cells than for the WT cells (Fig. 17D). The model also predicted that Ship1 activity would be elevated during proteasome inhibition (Fig. 17E), which was confirmed experimentally as well (Fig. 17F). We conclude that Shc1 slows the onset of mast cell hyperactivity (i.e., priming), likely by supporting activation of Ship1.

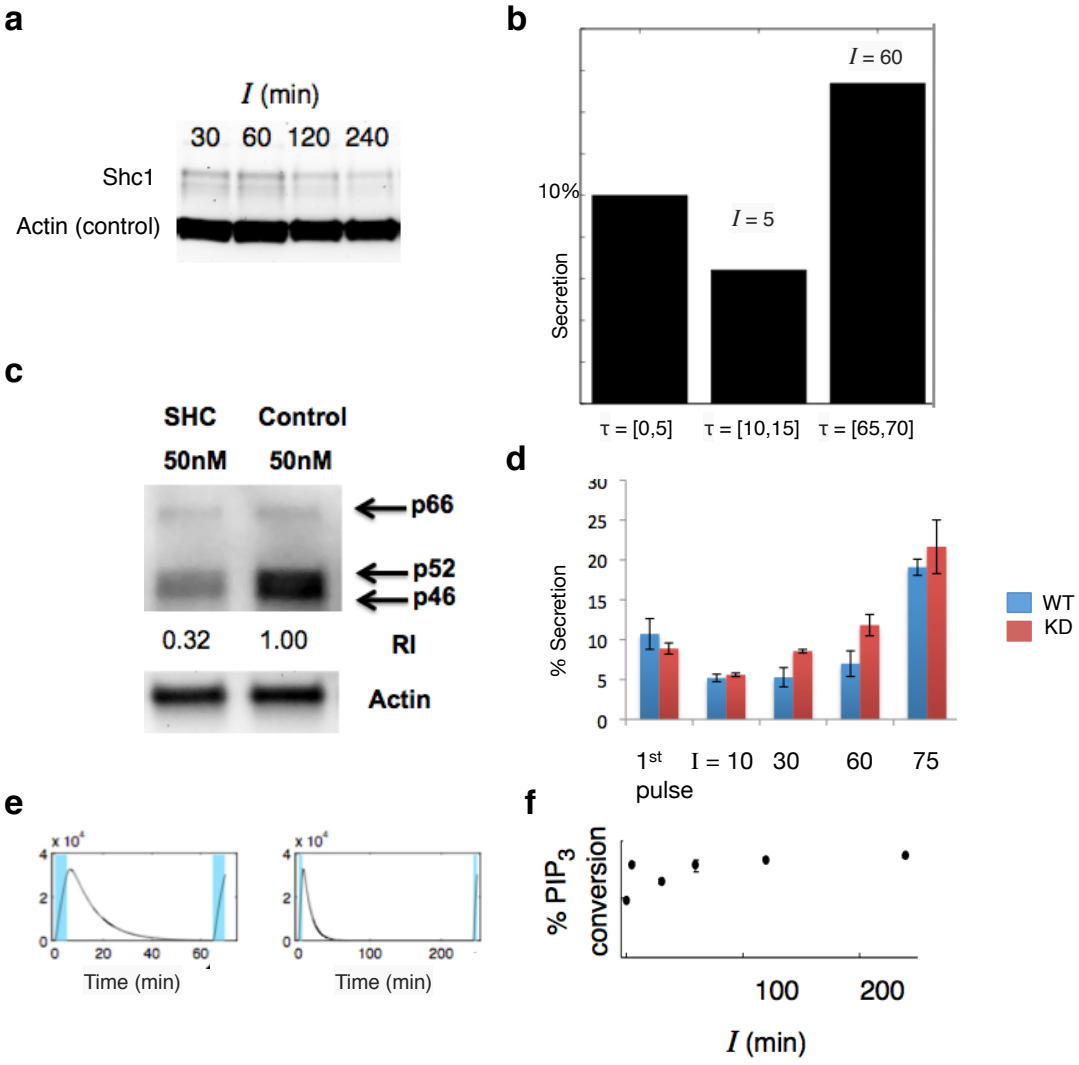


Figure 3.17. Role of Shc1 in memory responses. A) Measured Shc1 abundance over time. B) Predicted effect of *Shc1* knockdown on memory. Compare to simulations of WT cells in Fig. 12. C) Efficiency of *Shc1* knockdown in RBL-2H3 mast cells. Data are representative of at least three experiments. D) Degranulation of WT and *Shc1* KD cells for various *I* intervals. The plot represents results from a pair of duplicate experiments; two additional pairs of duplicate experiments were performed with similar results. E) Predicted effect of proteasome inhibition on Ship1 activation. The left panel shows dynamics of Ship1 activation for I = 60, and the right panel shows dynamics of activation for I = 240. Ship1 is robustly re-activated in both cases. Stimulatory periods are indicated with blue backgrounds. F) Measured Ship1 activation during proteosome inhibition. Compare to Ship1 activation in WT cells, Fig. 13d.

Discussion

A cell's response to stimulation is determined by the internal dynamics of its information processing systems, which in turn can be influenced by what the cell has experienced in the past. Investigating the details of this relationship requires fine control of the cellular microenvironment, quantification of resultant changes in the biochemical state of a cell, and an understanding of the molecular network that connects inputs to outputs. In achieving this goal, microfluidic technology and computational modeling are two valuable tools. We coupled the two to examine the internal dynamics of signaling induced by complex time-varying patterns of stimulation of the high-affinity receptor for IgE (FcεRI). We found that mast cells exhibit varied responses depending on the interval of time between periods of stimulation: short intervals resulted in attenuated responses to a second antigen exposure, whereas long intervals of input quiescence resulted in enhanced responses.

In investigating the mechanisms underlying these contrasting responses, pulsed stimulation enabled us to consider both the kinetics of activation (upon exposure to multivalent DF3) and deactivation (upon exposure to monovalent DNP-lysine) on timescales of minutes to hours. We uncovered different timescales on which activation and deactivation of signaling proteins occur. The protein tyrosine kinase Syk, which transmits signals that promote degranulation, undergoes rapid activation and deactivation in a history-independent manner. In contrast, the lipid phosphatase Ship1, which has a negative effect on degranulation (*21*), undergoes slower activation and deactivation. The slower kinetics of Ship1 activation and deactivation gives rise to history-dependent responses. History dependence also arises from the much slower process of antigen-

induced proteasome-mediated degradation of proteins, including Shc1, a co-factor of Ship1 (*18,22*).

Priming is a feature of several types of immune cells. In a general sense, priming prepares a cell to respond robustly to antigen stimulation at a later time. Other immune cells, such as T and B cells, undergo priming when they encounter their specific antigen and begin differentiation. Mast cells can be primed by IL-4 (*26*) or by pharmacological agents (*27*). Here, we have shown that mast cells can also be primed by exposure to an antigen. The possible biological relevance of this phenomenon is that prolonged exposure to antigens from, say, a parasite, which mast cells can defend an organism against (*28*), may induce inflammatory responses that are amplified over time.

Our results are consistent with known mechanisms of Syk and Ship1 activation, which involve recruitment to the plasma membrane. Syk has a tandem pair of Src homology 2 (SH2) domains that each bind weakly to phosphorylated immunoreceptor tyrosine-based activation motifs (ITAMs) in the γ chains of FcεRI. Phosphorylation of Y130 (rat numbering), which is located between the tandem SH2 domains, accelerates dissociation of Syk from ITAMs (*30*).

In contrast, Ship1 activation may be more sustained and long lived because of the tendency of Ship1 to join signaling complexes held together by multiple contacts. For example, Ship1 forms a ternary complex with Shc1 and Lat (*18*). Such complexes are likely to be long-lived because the components of these complexes can dissociate only after the sequential breakage of multiple noncovalent bonds (*23,24*). Given the number of proteins that interact with Ship1 and influence its recruitment to the plasma membrane

and its enzymatic activity, Shc1 may not be the sole target of the proteasome. Additional signaling proteins targeted for proteasome-mediated degradation may also contribute to the priming phenomenon.

The results presented here add another important role to Ship1's known repertoire. This protein has already been characterized as a "gatekeeper" of mast cell signaling due to its decisive role in determining the magnitude of mast cell responses by regulating the abundances of important phospholipids (*29*). Its presence also has a strong influence on the shape of a mast cell's dose-response curve; in the absence of Ship1, degranulation shows dramatic increase in magnitude across multiple antigen doses (*21*). Our findings, characterizing its activation/deactivation dynamics and its role in determining frequency responses, reveal that it has a strong impact on long-term cellular decision making.

The approach used here to uncover signaling dynamics in mast cells could be generalized to study the frequency response properties of other systems, e.g., how the frequency of stimuli impacts the magnitude of responses. Precisely timed addition and removal of a stimulus via microfluidic technology can potentially enable characterization of activation and deactivation dynamics in other cell signaling systems.

Methods

Experiments were performed by Dr. Yanli Liu and Dr. Brooke Harmon in the lab of Dr. Anup Singh at Sandia National Laboratory.

Cells and reagents

RBL-2H3 cells were maintained in MEM (Invitrogen, Carlsbad, CA), with 10% FBS, 100 U/ml penicillin, and 100 U/ml streptomycin (Invitrogen) and L-glutamine at 37°C under 5% CO_2. Mouse monoclonal anti-DNP-IgE was prepared as described by Liu et al. (*31*). Trivalent ligand DF3 and FAM-conjugated DF3 were custom synthesized by AnaSpec Inc. (Fremont, CA). DNP-lysine, 4-Methylumbelliferyl N-acetyl-β-D-glucosaminide (MUG) and phosphatase inhibitor were purchased from Sigma. RIPA lysis buffer and 10X Halt protease inhibitor were purchased from Thermo Scientific. SHC flexiTube siRNAs were purchased from Qiagen.

Integrated microfluidic platform for pulsed stimulation

Polydimethylsiloxane (PDMS)-based microfluidic devices were fabricated using standard soft lithography. The PDMS channel was permanently sealed to No. 1 glass cover slips. One end of PEEK tubing (125 μm i.d.) was directly inserted into the PDMS to insure tight sealing, while the other end of the tubing was connected to the fast response electronic valve controller and active pressure controller (APC) to achieve precise control of fluidics within the channel. The temperature of the system was maintained by a thermostat attached to the chip manifold. A graphical user interface (GUI) was programmed to interface the rabbit control box with pressure control units and electronic valves.

Prior to on-chip assays, the assembled devices were placed in a desiccator under vacuum for 2 h to evacuate the PDMS bulk and prevent bubble formation within the channels in the course of cell-based assays. The devices were cleaned using 10% freshly prepared bleach solution, 70% ethanol and Hank's buffered saline solution (HBSS). 10 μg/ml of

human plasma fibronectin was used to enhance the attachment of cells to the glass surface. RBL-2H3 cells in suspension were injected into the device to achieve 5,000-10,000 cells in serpentine channels. After a 30 min static incubation of cells at 37°C, cells were primed with 100 ng/ml of monoclonal anti-DNP IgE. Flow rates between 0.5 and 1.0 μL/min were used in the washing steps unless indicated otherwise.

Pulsed stimulation pattern over single cells within the channel was achieved by alternating trivalent ligand DF3 (for example, at 10 nM concentration for 5 min) and excess monovalent ligand DNP-lysine (for example, at 100 μM concentration for another 5 min). Downstream cell-based assays including microscopic imaging, flow cytometry analysis and degranulation assays were performed immediately before and after each DF3 pulse.

Degranulation Assay

RBL-2H3 cells were plated in 24-well plates and treated with 1 μg/ml of mouse monoclonal anti-DNP-IgE in complete medium overnight. Following overnight priming with IgE, cells were washed two times with Hanks Buffer to remove excess IgE, then stimulated with 10 nM DF3 in Hanks Buffer (0.5 ml/well) for 5 min at 37°C and 150 μl of sample was collected for the first 5 min time point. For all other time points the cells were washed 1X with Hanks buffer prior to addition of 100 mM of DNP-lysine in Hanks buffer for 5 min at 37°C. The cells were washed 1X with Hanks buffer to remove DNP-lysine then complete medium was added for the indicated time intervals: 15 min, 30 min, 60 min, 75 min, 90 min, 120 min, 240 min, 300 min (each interval includes an initial 5-min period of incubation with DNP-lysine). At these times medium was removed and cells were re-

stimulated with 10 nM DF3 for 5 min at 37°C and then 150 μL of supernatant was collected in microcentrifuge tubes. After supernatants were collected 0.5 ml/well of 1% Triton in Hanks buffer (TX-100) was added to cells to release total β-hexosaminidase from the cells. Then 150 μl of cell lysate was collected in microcentrifuge tubes. To monitor the release of the chemical mediator β-hexosaminidase upon pulsed stimulation, a live cell degranulation assay was developed with modifications of our previously reported protocol (*31*). Briefly, 10 μL of the collected samples was added to 10 μL of MUG (1 mM) in sodium acetate (pH 5) in 1 well of a black bottom 96 well plate. After 30 min incubation at 37°C, 200 mM glycine (pH 10) buffer was used to stop the reaction. Fluorescence was measured at EX360/EM450. Degranulation is quantified as the percentage of total β-hexosaminidase released in TX-100 cell lysates. The percent degranulation was defined as the amount of secreted β-hexosaminidase divided by the total β-hexosaminidase from whole cells times 100%. The background level of spontaneous secretion was subtracted from each sample. Experiments were performed in duplicate or triplicate three or more times and the average (± standard deviation) for each experiment was calculated.

Transfection with siRNA

Qiagen Flexi Tube rat Shc1 siRNA (Rn_Shc1_1 FlexiTube siRNA) was transfected into RBL-2H3 cells at a final concentration of 20 nM using Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's instructions. Cells at 6-hours post-transfection were plated in 24-well plates and treated with 1 μg/ml of mouse monoclonal anti-DNP-IgE

in complete medium overnight. Following overnight priming with IgE, cells were analyzed for degranulation upon DF3 stimulation.

Western blot analysis

Qiagen Flexi Tube rat Shc1 siRNA was transfected into RBL-2H3 cells at a final concentration of 20 nM using Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's instructions. At 24 hours post-transfection, cells were washed with PBS, trypsinized, resuspended in PBS, spun down, and then resuspended in Thermo Scientific RIPA buffer with 1X Halt protease inhibitors. Protein concentrations were measured using Bradford assay (Bio-Rad) and 25 μg of lysate was separated on Mini-Protean TGX precast gels (Bio-Rad), transferred to polyvinylidene difluoride (PVDF) membranes, blocked with 0.01% Tween 20 and 5% dry milk in TBS for 1 h and probed with anti-SHC (BD Biosciences) overnight in a block. After five washes with 0.01% Tween 20 in TBS (TBS-T), a Pierce horseradish peroxidase (HRP)-conjugated polyclonal goat anti-rabbit secondary antibody (Thermo Scientific) was added at 1:2500 for 30 min at room temperature in a block. Proteins were visualized with an Alpha Innotech (San Leandro, CA) imager using a SuperSignal West Femto sensitivity substrate (Thermo Scientific). The blots were then probed for 1 h with an HRP-conjugated rabbit anti-actin polyclonal antibody (Novus) at 1:2000 and were washed and visualized with an Alpha Innotech (San Leandro, CA) imager using a Super- Signal West Femto sensitivity substrate (Thermo Scientific).

Flow cytometry analysis of pSyk

Individual cells upon stimulation were fixed at desired time points with paraformaldehyde (PFA, 4%) for 15 min at room temperature and permeabilized with 0.2% Tween 20 for additional 10 min. The cells were stained with PE conjugated phospho-ZAP-70/Syk antibody (1:100) in Pharmingen staining buffer. Stained cells were detached from the channel surface after immersion in 1% trypsin-EDTA at 37°C for 5 min using a high-pressure flow (4 to 5 psi). Detached cells were collected, re-suspended in PBS buffer and analyzed using Accuri flow cytometer (BD Biosciences).

Internalization

Internalization of IgE-FcεRI complexes was measured based on established protocols using acid stripping and flow cytometry. Briefly, RBL cells were sensitized with IgE-Alexa^{647} and stimulated with pulsed DF3 ligand. Then, the cells were briefly washed with a stripping buffer (0.5 M of NaCl and 0.2 M acetic acid) at 4°C for 5 min to remove the surface IgE from cells. For quantification, identical samples were processed for flow cytometry, with and without acid stripping.

Inhibitor Treatment

RBL-2H3 cells were plated 24 hr before the inhibitor treatment. The following inhibitors were initially resuspended in dimethyl sulfoxide (DMSO) and were then diluted in MEM medium to obtain the final concentration: SHIP1 inhibitor 3AC (20 μM), SHP1/2 PTPase inhibitor NSC-87877 (1 μM) and proteasome inhibitor MG132 (10 μM). As a

control, the cells were also incubated with 50 μM DMSO, representing the highest concentration of DMSO to which the cells were exposed with the inhibitors listed above. Cells were pre-incubated with MEM, DMSO, and 3AC (Ship1 inhibitor), NSC-87877 (Shp1 inhibitor), or MG132 (proteasome inhibitor) for 1 to 3 hrs at 37°C. Following inhibitor treatment, the cells were examined by immunoblotting and degranulation analysis upon pulsed stimulation as described elsewhere in this report.

To test the viability of cells treated with chemical inhibitors, PrestoBlue cell viability reagents (Life technologies) were added to treated cells in MEM medium to make a 1X final concentration. The cells were incubated at 37°C for 10 min. Fluorescence was measured at an excitation wavelength of 560 nm and an emission wavelength of 590 nm.

Western blot analysis of protein/inositol phosphatases

A panel of protein/inositol phosphatases including Lyn, Dok1, Ship1, and Shc was examined for total expression level upon pulsed stimulation. Stimulated cells were treated with lysis buffer supplemented with phosphatase inhibitor and protease inhibitor (1:100). 8 μL of the lysates were separated on Novex NuPage 4-12% bis-tris precast gels (Life Technology), transferred to nitrocellulose membrane, blocked with 3% BSA in TBST (TBS + 0.05% Tween 20) for 1 hr, and then probed with desired antibody at 4°C for 12 hrs. After three washes with TBST, the blots were incubated with 1:5000 horseradish peroxidase (HRP) conjugated goat anti-rabbit IgG (Santa Cruz Biotechnology) for 2 hrs at room temperature, followed by an additional 3 washes with TBST. The target protein was

visualized with an Alpha Innotech (San Leandro, CA) imager using a Clarity western ECL substrate (Bio RAD).

Biochemical analysis of SHIP1 activity after membrane extraction

Membrane phosphatase SHIP1 was isolated using Mem-Per Plus Membrane Protein Extraction kit (Pierce) according to manufacturer's instruction. Briefly, after RBL cells were stimulated with pulsed DF3, cells were permeabilized with a mild detergent to release the cytosolic proteins followed by a second detergent to solubilize membrane proteins. The isolated membrane proteins were analyzed using a Malachite Green phosphatase assay kit (Echelon Biosciences). The isolated membrane proteins were incubated with water-soluble substrate PtdIns(3,4,5)P3 (PIP_3) for 60 min at 37°C. Free phosphate generated from the SHIP1 activity was determined by measuring absorbance at 640 nm in a microplate reader. Percent of PIP_3 conversion was determined for each time point as [(free phosphate in reaction, pmol) − (background phosphate, pmol)] × 100%/3000 pmol. Free phosphate in background is the value of phosphate in "substrate-only PIP_3" controls.

Modeling and Simulation

We formulated a chemical kinetic model for IgE receptor signaling using the BioNetGen Language (BNGL) (*32*). The model is rule based, which enables the representation of sites (e.g., domains and amino acid residues) within biomolecules (*33,34*). The model was simulated deterministically using BioNetGen, meaning that the

ordinary differential equations (ODEs) derived from the rules of the model were integrated numerically using CVODE (*35*) with BioNetGen's default settings for algorithmic parameters (*32*). The model input file is provided as Supplementary File 1.

References

1. S. Kraft, J.-P. Kinet. New developments in FcεRI regulation, function, and inhibition. *Nat Rev Immunol* **7**, 365-378 (2007).
2. Barbu A.E., Pecht I. Desensitization of mast cells' secretory response to an immuno-receptor stimulus. *Immunol Lett*. Aug **100**, 78-87 (2005).
3. G.R. Mendoza, K. Minagawa. Subthreshold and suboptimal desensitization human basophils. I. Kinetics of decay of releasability. *Int Arch Allergy Appl Immunol*, **68**, 101-107 (1982).
4. G.R. Mendoza, K. Minagawa. Subthreshold and suboptimal desensitization human basophils. II. Nonspecificity and irreversibility of desensitization. *Int Arch Allergy Appl Immunol*, **69**, 282–284 (1982).
5. J.J. Pruzansky, R. Paterson. Desensitization of human basophils with suboptimal concentrations of agonist. Evidence for reversible and irreversible desensitization. *Immunology*, **65**, 443–447 (1988).

6. A. Komiya, K. Hirai, M. Iikura, H. Nagase, H. Yamada, M. Miyamasu, K. Ohta, Y. Morita, C. Ra, K. Yamamoto, M. Yamaguchi. Induction of basophil desensitization in physiological medium: enhancement after IgE-dependent upregulation of surface IgE binding on basophils. *Int Arch Allergy Immonol*, **130**, 40–50 (2003).
7. M. Weetall, D. Holowka, B. Baird. Heterologous desensitization of the high affinity receptor for IgE (FcεRI) on RBL cells. *J Immunol*, **150**, 4072–4083 (1993).
8. Zhao W., Gomez G., Macey M., Kepley C.L., Schwartz L.B. *In vitro* desensitization of human skin mast cells. *J Clin Immunol*. **32**, 150-160 (2012).
9. Bounab, Y., Getahun, A., Cambier, J.C., Daëron, M. Phosphatase regulation of immunoreceptor signaling in T cells, B cells and mast cells. *Curr Opin Immunol*. 2013 **25**, 313-320 (2013).
10. Mettetal, J. T., Muzzey, D., Gomez-Uribe, C., van Oudenaarden, A. The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*. *Science* **319**, 482-484 (2008)
11. Hersen, P., McClean, M. N., Mahadevan, L., Ramanathan, S. Signal processing by the HOG MAP kinase pathway. Proc Natl Acad Sci USA. **105**, 7165-7170 (2008).
12. Wang, C. J., Bergmann, A., Lin, B., Kim, K., Levchenko, A. Diverse sensitivity thresholds in dynamic signaling responses by social amoebae. *Sci Sig* **5**, ra17 (2012).

13. Mahajan A., Barua D., Cutler P., Lidke D.S., Espinoza F.A., Pehlke C., Grattan R., Kawakami Y., Tung C.S., Bradbury A.R., Hlavacek W.S., Wilson B.S. Optimal aggregation of FcεRI with a structurally defined trivalent ligand overrides negative regulation driven by phosphatases. *ACS Chem Biol*. **9**, 1508-1519 (2014).
14. Mao, S.Y. and Metzger, H. Characterization of protein-tyrosine phosphatases that dephosphorylate the affinity IgE receptor. J Biol Chem 272, 14067-73 (1997)
15. Gradler U Structural and biophysical characterization of the Syk activation switch. J Mol Biol 425: 308-333 (2013)
16. Geahlen R.L. Syk and pTyr'd: Signaling through the B cell antigen receptor. *Biochem Biophys Acta* 1793: 1115-27 (2009)
17. Brooks, R., Fuhler, G.M., Iyer, S., Smith, M.J., Park, M.Y., Paraiso, K.H., Engelman, R.W., Kerr, W.G. SHIP1 inhibition increases immunoregulatory capacity and triggers apoptosis of hematopoietic cancer cells. *J Immunol*. **184**, 3582-3589. (2010)
18. Ulivieri, C., Fanigliulo, D., Masi, G., Savino, M.T., Gamberucci, A., Pelicci, P.G., Baldari, C.T. p66Shc is a negative regulator of FcεRI-dependent signaling in mast cells. *J Immunol*. **186**, 5095-5106 (2011).
19. Furumoto, Y., Nunomura, S., Terada, T., Rivera, J., Ra, C. The FcεRIβ immunoreceptor tyrosine-based activation motif exerts inhibitory control on MAPK and IκB kinase phosphorylation and mast cell cytokine production. J Biol Chem 279: 49177-49187 (2004)

20. Faeder J.R., Hlavacek W.S., Reischl I., Blinov M.L., Metzger H., Redondo A., Wofsy C., Goldstein B. Investigation of early events in FcεRI-mediated signaling using a detailed mathematical model. *J. Immunol*. **170**, 3769-81 (2003).

21. Huber, M., Helgason, C.D., Damen, J.E., Liu, L., Humphries, R.K., Krystal, G. The src homology 2-containing inositol phosphatase (SHIP) is the gatekeeper of mast cell degranulation. *Proc Natl Acad Sci U S A*. **95**, 11330-5. (1998)

22. Damen, J.E., Liu, L., Rosten, P., Humphries, R.K., Jefferson, A.B., Majerus, P.W., Krystal, G. The 145-kDa protein induced to associate with Shc by multiple cytokines is an inositol tetraphosphate and phosphatidylinositol 3,4,5-triphosphate 5-phosphatase. *Proc Natl Acad Sci U S A*. 93, 1689-1693. (1996)

23. Deeds E.J., Bachman J.A., Fontana W. Optimizing ring assembly reveals the strength of weak interactions. *Proc Natl Acad Sci USA* **109**, 2348-2353 (2012).

24. Barua D, Hlavacek W.S. Modeling the effect of APC truncation on destruction complex function in colorectal cancer cells. *PLOS Comput Biol* **9**, e1003217 (2013).

25. Huber, M., Gibbs, B.F. SHIP1 and the negative control of mast cell/basophil activation by supra-optimal antigen concentrations. *Mol Immunol*. 63, 32-37 (2015).

26. Bischoff, S.C. Role of mast cells in allergic and non-allergic immune responses: comparison of human and murine data. *Nat Rev Immunol* **7**, 93-104 (2007).

27. Nassiri, M., Babina, M., Dölle, S., Edenharter, G., Ruëff, F., Worm, M. Ramipril and metoprolol intake aggravate human and murine anaphylaxis: evidence for direct mast cell priming. *J Allergy Clin Immunol*.**135**, 491-499 (2015).

28. Kawakami, T., Galli, S.J. Regulation of mast-cell and basophil function and survival by IgE. *Nat Rev Immunol*. **2**, 773-786. (2002)

29. Huber, M., Helgason, C.D., Damen, J.E., Liu, L., Humphries, R.K., Krystal, G. The src homology 2-containing inositol phosphatase (SHIP) is the gatekeeper of mast cell degranulation. *Proc Natl Acad Sci U S A*. **95**, 11330-11335. (1998)

30. Zhang, Y., Oh, H., Burton, R.A., Burgner, J.W., Geahlen, R.L., Post, C.B. Tyr130 phosphorylation triggers Syk release from antigen receptor by long-distance conformational uncoupling. *Proc Natl Acad Sci U S A*. **105**, 11760-11765. (2008)

31. Liu, Y., Barua, D., Liu, P., Wilson, B.S., Oliver, J.M., Hlavacek, W.S., Singh, A.K. Single-cell measurements of IgE-mediated FcεRI signaling using an integrated microfluidic platform. *PLOS ONE* **8**, e60159 (2013).

32. Faeder, J.R., Blinov, M.L., Hlavacek, W.S. Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol Biol* **500**, 113-67 (2009).

33. Chylek, L.A., Harris, L.A., Tung, C.S., Faeder, J.R., Lopez, C.F., Hlavacek, W.S. Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. *Wiley Interdiscip Rev Syst Biol Med*. **6**, 13-36 (2014).

34. Stefan, M.I., Bartol, T.M., Sejnowski, T.J., Kennedy, M.B. Multi-state modeling of biomolecules. *PLoS Comput Biol*. 10, e1003844. (2014)

35. Hindmarsh, A.C. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans Math. Softw.* 31, 363-396. (2005)

Chapter 4: Fitting cell signaling models to data [1]

Abstract

Rule-based models are analyzed with specialized simulators, such as those provided by the BioNetGen and NFsim open-source software packages. Here, we present BioNetFit, a general-purpose fitting tool that is compatible with BioNetGen and NFsim. BioNetFit is designed to take advantage of distributed computing resources. This feature facilitates fitting (i.e. optimization of parameter values for consistency with data) when simulations are computationally expensive. BioNetFit can be used on stand-alone Mac, Windows/Cygwin, and Linux platforms and on Linux-based clusters running SLURM, Torque/PBS, or SGE. The BioNetFit source code (Perl) is freely available (http://bionetfit.nau.edu).

Introduction

Biomolecular interactions can be represented by formalized rules (Chylek et al., 2014a; Stefan et al., 2014). Collections of rules form rule-based models, which provide concise representations of biomolecular interaction networks and can be analyzed to obtain insights into how system-level behavior emerges from biomolecular interactions (Chylek et al., 2014a; Stefan et al., 2014). Rule-based models must be analyzed with specialized algorithms and software tools (Chylek et al., 2014a; Stefan et al., 2014), such as BioNetGen (Harris et al., 2015), which interprets models encoded in the BioNetGen language (BNGL) and provides deterministic, stochastic and hybrid forward simulation capabilities (Harris et al., 2015). To date, other critical methods of analysis, such as fitting (parameter estimation), have typically been applied *ad hoc* (e.g. by writing

[1] This work has been published as: BR Thomas, LA Chylek, J Colvin, S Sirimulla, AHA Clayton, WS Hlavacek, RG Posner. (2015) GenFit: a fitting tool compatible with BioNetGen, NFsim, and distributed computing environments. *Bioinformatics* pii, btv655.

problem-specific programs), as in the study of Kozer et al. (2013) or Chylek et al. (2014b), which leads to duplication of effort and hinders the ability to reproduce results.

Although many algorithms and software implementations are available for analysis of models specified in traditional forms (Press et al., 2007), such as that of ordinary differential equations (ODEs), standard techniques are not easily applied to rule-based models except for those rule-based models that can be recast into a traditional model form. Translating a set of rules into a reaction network (i.e. a list of reactions) or the corresponding ODEs for the mass-action kinetics of the network, a capability provided by BioNetGen, is expensive and sometimes impracticable. In such cases, direct simulation methods must be used (Chylek et al., 2014a). These methods have been called network-free methods, because they do not involve processing a set of rules to obtain a list of reactions. Rather, rules are used as reaction- event generators within a particle-based kinetic Monte Carlo (KMC) simulation. Network-free simulators, such as NFsim (Sneddon et al., 2011), which is compatible with BNGL-specified models, expand the range of biomolecular interaction networks that can be studied, because the cost of network-free simulation scales with the number of rules in a model, not the number of reactions implied by the rules (Danos et al., 2007). Software that provides a general-purpose fitting capability and that interfaces with a network-free simulator has not hitherto been available.

Here, we present BioNetFit, a general-purpose fitting program for rule-based models that is compatible with BioNetGen and NFsim. Because of the potentially high cost of network-free simulation (Chylek et al., 2014a), BioNetFit has been designed to be used on clusters. It can also be used on stand-alone computers. BioNetFit implements a genetic algorithm (Whitley, 1994;

Smith and Eiben, 2008). Below, we provide an overview of BioNetFit and a demonstration of its capabilities.

Methods

This software was developed in collaboration with Brandon Thomas and Dr. Richard Posner at Northern Arizona University.

Source code and a user manual, which includes installation instructions, are available at the BioNetFit Web site (http://bionetfit.nau.edu). Dependencies are described in the user manual and in Supplementary Methods. Collections of files needed to run example fitting jobs are included in the BioNetFit distribution in the 'examples' directory and six subdirectories named 'example1' through 'example6.' Step-by-step instructions for running these fitting jobs are included in the user manual. The fitting procedure implemented in BioNetFit is described in the Methods.

Running BioNetFit requires several problem-specific plain-text files: a BNGL model-specification (.bngl) file (i.e. a BioNetGen/NFsim input file); one or more data (.exp) files, each with two or more white space-separated columns; and a BioNetFit configuration (.conf) file (see the 'examples' directory in the BioNetFit distribution). If NFsim is being used, an .rnf file may also be required. To run BioNetFit, a user must identify the parameters that are free to vary in fitting in the .conf file and include a simulation command in the .bngl file for each .exp file. Each simulation command should generate outputs corresponding to the data in the associated .exp file. See the BioNetFit user guide for additional information.

Demonstration results were obtained by running BioNetFit on the Monsoon cluster at Northern Arizona University (http://nau.edu/hpc/). The .bngl, .exp and .conf files required to repeat the demonstration results presented in the Results section are provided as Supplementary Files S1–S6 onlinw. These files are also provided in the 'example1' and 'example2' subdirectories of the 'examples' directory within the BioNetFit distribution.

Results

To provide a demonstration of capabilities, we used BioNetFit together with BioNetGen, NFsim and experimental data of Kozer et al. (2013) to estimate parameter values for a published (Kozer et al., 2013) and an extended version of a model for activation of the epidermal growth factor (EGF) receptor (EGFR). Each version of the model accounts for EGF-induced oligomerization of EGFR. The published model is defined in Supplementary File S1 of the BioNetFit distribution, which is a .bngl file that directs BioNetGen's ODE solver to produce simulation results. The file is derived from the .bngl file used by Kozer et al. (2013). This version of the model limits EGFR oligomers to a maximum size of four receptors per oligomer. This limitation was introduced by Kozer et al. (2013) to avoid the need to use computationally expensive network-free simulation in fitting, which was performed using problem-specific code. The extended model is defined in Supplementary File S2 of the BioNetFit distribution, which is a .bngl file that directs NFsim to produce simulation results. The file is derived from the .bngl file used by Kozer et al. (2014). This version of the model imposes no constraints on the size of linear oligomers. It is otherwise identical to the published model of Kozer et al. (2013). The

steady-state dose-response and time-series data used in each of the two fitting runs are provided in Supplementary Files S3 and S4, which are .exp files. Supplementary Files S5 and S6 are the BioNetFit configuration (.conf) files for the two fitting runs. These files complement Supplementary Files S1 and S2, respectively.

In the fitting procedure, BioNetFit evaluated the consistency of simulation results with all available data (i.e. BioNetFit found a global fit). The chi-square metric was used to determine goodness of fit. Parameter estimates are summarized in Table 1. The number of free parameters was nine, which included five physical parameters (e.g. rate constants) and four scaling factors (parameters of a measurement model), which define linear relationships between measured quantities and model-predicted quantities. The quality of the fit for the original and extended models is illustrated in Figure 1. For best-fit parameter values, the extended model predicts that EGFR oligomers larger than tetramers do not form appreciably (Figure 2). This finding justifies the omission of these oligomers in the analysis of Kozer et al. (2013).

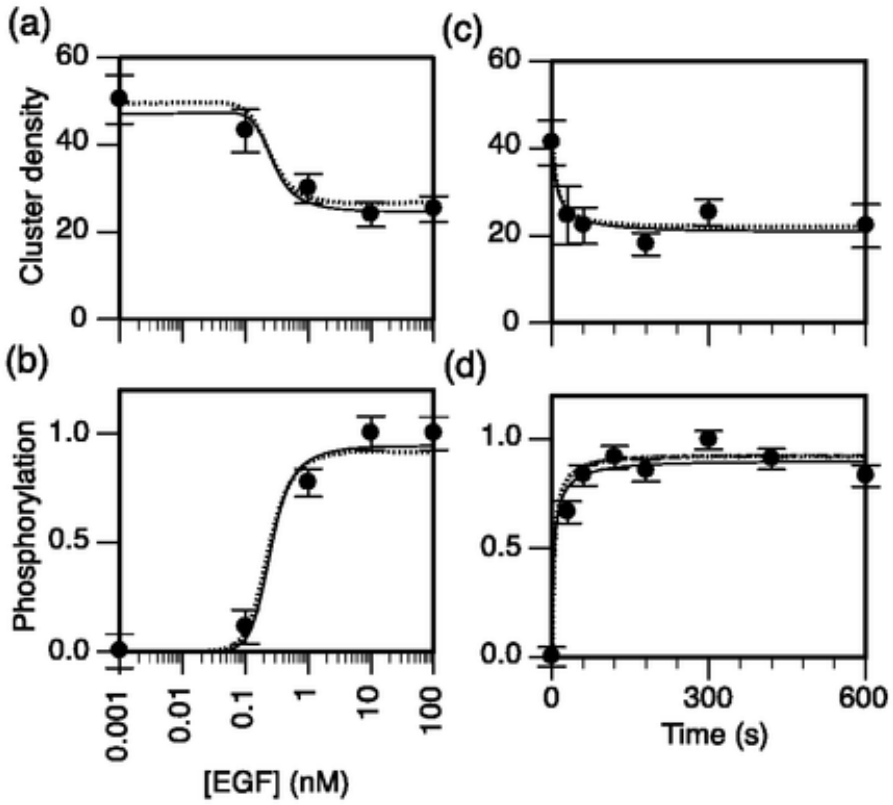


Figure 4.1. Plots illustrating the quality of fits found by BioNetFit. BioNetFit was used to estimate values for the free parameters of the model of Kozer et al. (2013) for EGFR activation. BioNetFit was also used to estimate values for the corresponding parameters of an extended model, which can only be analyzed using network-free simulation. In each case, parameter estimates are based on the data of Kozer et al. (2013), which are represented by points with error bars. The left panels, (a) and (b), show fits to the steady-state dose-response data. The right panels, (c) and (d), show fits to the time-series data. Solid curves show deterministic simulation results produced by BioNetGen's ODE solver. Dotted curves show stochastic simulation results produced by NFsim. Curves are based on the best-fit parameter values, which are summarized in Table 1. The stochastic simulation results shown here represent averages from multiple simulation runs. In fitting, each stochastic simulation run was repeated and the results averaged to obtain smoothed time courses. The CPU time typically required for a single stochastic simulation was ~3 min. The fitting procedure for the extended model involved a total of ~10 000 (stochastic) simulations. Because of parallel processing, the procedure took only ~40 h. The number of EGFR clusters (the sum of EGFR monomers, dimers, etc.) per μm2 is indicated on the vertical axis of panel (a) and (c). The relative level of EGFR phosphorylation (arbitrary units) is indicated on the vertical axis of panel (b) and (d).

BioNetFit's ability to take advantage of simulation runs performed in parallel on a cluster

enabled consideration of the extended model, which had to be simulated by calling NFsim. The

KMC (stochastic simulation) method implemented in NFsim tends to be computationally expensive, because in such methods, system state is advanced one reaction event at a time (Chylek et al., 2014a). Use of NFsim was required because the extended model places no constraint on the size of (linear) EGFR oligomers, which form via polymerization-like reactions. Network-free simulation is typically required to cope with polymerization-reactions (Chylek et al., 2014a). As expected, we found that simulations performed by NFsim (for the extended model) were significantly more expensive than simulations performed by BioNetGen's ODE solver (for the original model). The cost difference depends on various factors, including hardware, model parameter values and algorithmic parameter settings. For a representative CPU in the cluster we used and the parameter settings of Supplementary Files S1 and S2, an NFsim simulation is 5-fold more expensive than an ODE-based simulation. It should be noted that the cost difference depends sensitively on the value of one particular algorithmic parameter called *f*, which scales system size. In general, the cost of a discrete-event stochastic simulation, the type of simulation performed by NFsim, depends on the number of reaction events per unit time, which in turn depends on system size (measured by the number of reactive molecules in the system). We set *f* at 0.01, which corresponds to a system size of 1% of a cell. This setting reduces the cost of individual simulation runs (and memory requirements) at the expense of introducing noise (i.e. fluctuations arising from probabilistic reaction events). The noisiness of simulation results complicates the comparison of predicted and measured (population averaged) quantities. Because of this issue, in fitting, multiple NFsim simulations were performed for each goodness-of-fit evaluation. BioNetFit allows a user to cope with the noisiness of stochastic simulation results in a brute-force manner by repeating simulation runs a user-specified number

of times and then averaging the results. This procedure serves to smooth stochastic simulation results.

Results from two additional non-trivial fitting demonstrations are summarized in Supplementary Tables 2 and 3, which provide best-fit parameter estimates, and Figure 3 and 5, which illustrate the quality of fits obtained. The files needed to reproduce these results are provided in the 'example3' and 'example4' subdirectories of the 'examples' directory in the BioNetFit distribution. The demonstrations each require NFsim. The 'example3' demonstration concerns a rule-based model for interaction of a trivalent ligand with a bivalent cell-surface receptor and parameter estimation on the basis of equilibrium data collected via flow cytometry (Monine et al., 2010).

The 'example4' demonstration concerns a rule-based model for early events in T-cell receptor signaling and parameter estimation on the basis of temporal phosphoproteomic data collected via mass spectrometry-based proteomics (Chylek et al., 2014b). This model is illustrated in Figure 4. In the original study, parameters were hand-tuned for consistency with experimental data. We re-fitted the model using the same literature-based ranges as the original study. For this problem, because changes in phosphorylation levels were measured relative to basal phosphorylation levels, it was necessary to normalize model outputs to the basal steady state.

Conclusions

Parameter estimation is a critical step in model building and model-guided analysis of data. BioNetFit provides a fitting capability that is suitable for a broad range of rule-based modeling

applications because of its compatibility with BioNetGen and NFsim. Moreover, BioNetFit can potentially be used with traditionally formulated models (e.g. ODE models), because BioNetGen has the capability to translate models encoded in the SBML format (Hucka et al., 2003) into BNGL format (Harris et al., 2015). Importantly, BioNetFit is designed to take advantage of distributed computing resources, which can make fitting feasible even when simulation runs are computationally expensive.

Methods

BioNetFit is compatible with simulators specialized for rule-based models, namely those provided by the BioNetGen and NFsim open-source software packages. Several recent reviews about rule-based modeling (and enabling software) are available (Chylek et al., 2014a; 2015; Stefan et al., 2014). BioNetGen translates rule-based models encoded in the BioNetGen language (BNGL) into equivalent models having traditional forms (Blinov et al., 2004; Faeder et al., 2005; Blinov et al., 2006), such as ordinary differential equations (ODEs), and implements deterministic, stochastic, and hybrid simulation methods (Blinov et al., 2004; Faeder et al., 2009; Hogg et al., 2014; Harris et al., submitted). NFsim (Sneddon et al., 2011) implements a network-free simulation algorithm (Danos et al., 2007; Yang et al., 2008; Colvin et al., 2009; 2010; Yang and Hlavacek, 2011; Chylek et al., 2013; Chylek et al., 2014a), which is a particle-based kinetic Monte Carlo (KMC) method in which BNGL-encoded rules are used as reaction-event generators. Network-free simulation allows one to cope with rule-based models that cannot be recast into traditional forms (Chylek et al., 2013; Chylek et al., 2014a). In the past, as in the study of Chylek et al. (2014b) or Monine et al. (2010), when researchers have analyzed a model for which network-free simulation is necessary, they have had to write purpose-built scripts to perform fitting (i.e., to optimize parameter values for consistency with data). The lack of general-purpose software for fitting has limited the use of rule-based models to analyze data, as well as the reproducibility of fitting.

BioNetFit implements a genetic algorithm (Whitley, 1994; Smith and Eiben, 2008). The fitting procedure is outlined below. BioNetFit can be used on stand-alone Mac, Windows (with Cygwin) and Linux platforms. It can also be used on Linux-based clusters running SLURM,

Torque/PBS, or SGE. The BioNetFit source code (Perl) is freely available (http://BioNetFit.nau.edu) – it is distributed under the terms of the GNU General Public License (GPL), version 3.

Along with the BioNetFit source code, a user manual, which includes installation instructions, is available at the BioNetFit Web site (http://bionetfit.nau.edu). Collections of files needed to run examples of fitting jobs are included in the BioNetFit distribution in the 'examples' directory, which contains subdirectories named 'example1,' 'example2,' 'example3,' 'example4,' 'example5,' and 'example6.' Step-by-step instructions for running these fitting jobs are included in the user manual. The first two examples, which are related to the experimental/modeling studies of Kozer et al. (2013; 2014), are discussed in the Results section. These fitting jobs are computationally demanding; use of a cluster is recommended. Similarly, the third and fourth examples, which involve estimation of parameters of models considered by Monine et al. (2010) and Chylek et al. (2014b), are computationally demanding. The fifth and sixth examples, which involve synthetic data and a toy model for ligand-induced autophosphorylation of a receptor, can be executed on a laptop. For convenience, BioNetFit is distributed with current versions of BioNetGen (Harris et al., submitted) and NFsim (Sneddon et al., 2011). The current version of BioNetGen is 2.2.6. The current version of NFsim is 1.11.

Running BioNetFit requires a BNGL model-specification (.bngl) file (i.e., a BioNetGen/NFsim input file); one or more data (.exp) files, each with two or more white space-separated columns; and a BioNetFit configuration (.conf) file (see the 'examples' directory in the BioNetFit

distribution). If NFsim is being used, a .rnf file may also be required (see the 'example6' subdirectory in the 'examples' directory). All of these files are plain-text files.

Before running BioNetFit, a user needs to identify the parameters that are free to vary in fitting. Although free parameters may be identified by simply listing names of parameters in the .conf file, we recommend instead that the free parameters in the .bngl file be associated explicitly with identifiers listed in the .conf file. For example, in the *parameters* block of a .bngl file, we recommend that a user replace the line p = 1 with p = p__FREE__, where 'p' is the name of a parameter originally assigned a value of 1 in the .bngl file and 'p__FREE__' is listed in the .conf file as an identifier of a free parameter. As illustrated by this example, identifiers are likely to be more understandable if they are derived from parameter names, but this is not a requirement. These recommendations serve as a safeguard against potential parsing errors, which at present can arise if parameter names in the .bngl file appear in mathematical expressions. See the BioNetFit user guide for additional examples and further discussion.

A user must ensure that the .bngl file includes a simulation command for each .exp file. The simulation command should generate outputs corresponding to the data in the .exp file. Each simulation output to be used in fitting must be defined in the .bngl file (in either the *observables* or *functions* block). Furthermore, the outputs generated by a simulation command must be directed to either a .gdat file (in the case of time-course data) or a .scan file (in the case of steady-state dose-response data). The basename of the output file (which by default is the basename of the .bngl file that produces the output file) must be extended to include a suffix that matches the filename of the corresponding .exp file. For example, the files model_name.bngl,

data1.exp, and model_name_data1.gdat have names consistent with this convention. See the BioNetFit user guide for additional discussion of file naming conventions.

The BioNetFit configuration (.conf) file specifies the parameters of the genetic algorithm (for an introduction to genetic algorithms, see Whitley, 1994; Smith and Eiben, 2008), the objective function to be minimized in fitting (e.g., the sum-of-squares or chi-square function), the initial values of free parameters, the number of replicate simulation runs to be performed to smooth the results of stochastic simulations (optional), relationships between data input files and simulation output files, cluster and job-scheduling variables (e.g., the number of CPU cores to be used in parallel), paths to files/directories, and various BioNetFit options (e.g., the level of verbosity of console messages). Extensive documentation of the syntax and options of the .conf file is included in the BioNetFit user manual. Most .conf file settings have default values that are expected to be suitable for a wide array of applications.

BioNetFit has several dependencies. It requires Perl (https://www.perl.org), which is available on most Linux-based clusters and stand-alone computers. Cygwin (https://www.cygwin.com) must be installed to use BioNetFit on a Windows platform. For cluster computing, BioNetFit must be used with one of the following resource management tools, which are commonly installed on Linux-based clusters: SGE (http://gridscheduler.sourceforge.net), SLURM (http://slurm.schedmd.com), or Torque/PBS (http://www.adaptivecomputing.com/products/open-source/torque/). BioNetFit attempts to automatically detect the cluster environment in which it is running (e.g., by executing the command ‘which qsub’ and/or the command ‘which sbatch’) and to then configure itself appropriately. BioNetGen (http://bionetgen.org/) and NFsim

(http://emonet.biology.yale.edu/nfsim/) must be installed (in a user's home directory on a cluster) and available to accept commands from BioNetFit. Compilation of the NFsim C++ source code and/or the C/C++ source code for BioNetGen's run_network component may be required. BioNetFit and dependencies are compatible with 32- and 64-bit architectures.

The fitting procedure implemented in BioNetFit consists of the following steps:

1. Initialize. Generate a collection of N new BioNetGen/NFsim input (.bngl) files, each derived from the .bngl file provided by the user and differing only with respect to the values initially assigned by the user (either explicitly or via a procedure selected in the .conf file) to the K parameters identified by the user as being free to vary. Thus, the new .bngl files are in one-to-one correspondence with a "population" of parameter value sets $\{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$, where $\mathbf{P}_i = \{p_{i1}, \ldots, p_{iK}\}$ for $i=1, \ldots, N$. The value of N (the "population size") is an integer set by the user. The user's .bngl file should include definitions of a set of observables and functions (i.e., model outputs) that correspond to the measurements to be used in fitting. Data are provided by the user in the form of plain-text files, each with a .exp filename extension and containing white space-separated columns of numerical entries. For example, in the case of time-series data, the first column would consist of time points at which measurements were made. The series of measurements would then be given in the second column. BioNetFit can compare time-series data (which are provided in .exp files in which the first column of numerical entries has the header 'time') against simulation outputs reported in .gdat files, which are produced via *simulate* commands in the *actions* block of the user's .bngl file (Faeder et al., 2009). BioNetFit can also compare steady-state dose-response data [which are provided in .exp files in which

the first column of numerical entries has a header corresponding to the name of a (copy-number or concentration) parameter introduced in the *parameters* block of the user's .bngl file] against simulation outputs reported in .scan files, which are produced via *parameter_scan* commands in the *actions* block of the .bngl file (Faeder et al., 2009).

2. Calculate goodness-of-fit for each set of parameter values by passing the corresponding .bngl file to a simulator (BioNetGen or NFsim), retrieving the simulation results, and calculating the value of a user-selected objective function. Based on objective function evaluations, parameter value sets are ranked 1 to N according to goodness-of-fit. Evaluation of the objective function involves use of the .gdat and/or .scan file(s) generated by the simulator and the .exp file(s) provided by the user. The user should ensure that simulation calls are such that the .gdat/.scan file(s) contain $y(x)$-values at the x-values specified in the .exp file(s). Simulation jobs can be run in series or in parallel. In the case of stochastic network-free simulation with NFsim, a simulation run can be repeated a user-specified number of times to obtain smoothed (averaged) results. If BioNetGen is serving as the simulator, a .net file, which may be supplied by the user, is required to obtain simulation results. The .net file defines the reaction network implied by the rules in the user's .bngl file. If a .net file is required but not supplied by the user, the .net file is generated by calling BioNetGen's *generate_network* function (Faeder et al., 2009). BioNetFit operates such that the .net file is generated no more than once.
3. Stop if a stopping condition is satisfied and report the parameter value set with the best goodness-of-fit. A user can specify that BioNetFit stop a) after performing a maximum number of iterations or "generations" (G_{max}), b) after finding an acceptable goodness-of-

fit according to the user's specifications, or c) after converging to parameter values that would not be altered with further iterations.

4. Create a new collection of N BioNetGen/NFsim input files by repeating the following steps $N/2$ times:
    a. Select two parameter value sets $\mathbf{P}_i$ and $\mathbf{P}_j$ from the previous collection of N parameter value sets. Each parameter value set is selected with a probability inversely proportional to its rank.
    b. Create two new BioNetGen/NFsim input files by randomly exchanging the corresponding elements of $\mathbf{P}_i$ and $\mathbf{P}_j$. In other words, BioNetFit replaces p_{ik} with p_{jk} for $k=1,\ldots,K$ with probability Q_1, which is set by the user. Thus, the parameter Q_1 sets the rate of "recombination."
    c. With a user-defined probability Q_2, each element of $\mathbf{P}_i$ and $\mathbf{P}_j$ is multiplied by a randomly chosen factor that falls within a user-specified range (e.g., 0.8 to 1.2). Thus, the parameter Q_2 sets the rate of "mutation."
5. Return to Step 2.

The algorithm described above is an example of a "genetic algorithm" (Whitley, 1994; Smith and Eiben, 2008), a type of algorithm found to perform well for a range of problems. It is a biologically inspired heuristic global optimization method – its good performance is partly explained by its ability to avoid becoming trapped near local extrema. In other words, mutation and recombination allow the algorithm to move away from local extrema, which is important for finding a global extremum in a noisy fitness landscape. Many variations of the genetic algorithm concept, a heuristic global optimization approach, have been discussed in the literature. Many

other heuristic global optimization approaches have also been discussed in the literature. The relative performance of heuristic global optimization methods is problem specific.

The .conf files in the 'examples' directory of the BioNetFit distribution can be used as templates for new applications of BioNetFit and to guide the selection of values for algorithmic parameters (viz., N, Q_1, and Q_2).

In the example fitting jobs discussed in the Results section, deterministic simulation runs were performed using BioNetGen's ODE solver (with default settings), which is CVODE in the SUNDIALS package (Hindmarsh et al., 2005). Stochastic simulation runs were performed using NFsim.

Information about the performance of BioNetFit, which implements the algorithm described above, is reported in Fig.1 for Examples 1 and 2. For Example 3, simulation/fitting runs are faster, whereas for Example 4, simulation/fitting runs are slower. Use of cluster is recommended for Examples 1–4, which involve models and data reported in the literature (Monine et al., 2010; Kozer et al., 2013; Chylek et al., 2014b). Examples 5 and 6 are toy problems that can be run on a laptop.

Table 4.1 Parameter estimates[a,b] for Examples 1 and 2

Free parameter name[c]	Kozer et al. (2013)	BioNetFit (original model)[h]	BioNetFit (extended model)[i]
k_u (k_o)	6 s^{-1}	8.76 s^{-1}	30.3 s^{-1}
k_v (k_c)	1.6 s^{-1}	14.9 s^{-1}	0.660 s^{-1}
k_{cx} (kaf)	15.4 $nM^{-1}s^{-1}$ = 1.54×10^{-5} $(\#/cell)^{-1}s^{-1}$	70.4 $nM^{-1}s^{-1}$ —	— 2.20×10^{-6} $(\#/cell)^{-1}s^{-1}$
k_{cr} (kar)	8.89 s^{-1}	5.24 s^{-1}	0.275 s^{-1}
χ (chi_r)	4.37×10^{4} nM = 4.37×10^{10} #/cell	1.87×10^{4} nM —	— 1.06×10^{9} #/cell
— (alpha1)[d]	972 $(\#/\mu m^2)/nM$ = 9.72×10^{-4} $(\#/\mu m^2)/(\#/cell)$	911 $(\#/\mu m^2)/nM$ —	— 9.55×10^{-4} $(\#/\mu m^2)/(\#/cell)$
— (alpha2)[e]	820 $(\#/\mu m^2)/nM$ = 8.20×10^{-4} $(\#/\mu m^2)/(\#/cell)$	775 $(\#/\mu m^2)/nM$ —	— 7.91×10^{-4} $(\#/\mu m^2)/(\#/cell)$
— (alpha3)[f]	23.4 (a.u.)/nM = 2.34×10^{-5} (a.u.)/(#/cell)	22.7 (a.u.)/nM —	— 2.24×10^{-5} (a.u.)/(#/cell)
— (alpha4)[g]	21.0 (a.u.)/nM = 2.10×10^{-5} (a.u.)/(#/cell)	21.6 (a.u.)/nM —	— 2.25×10^{-5} (a.u.)/(#/cell)

[a]Estimates are reported using two different unit systems: 1) the unit system used by Kozer et al. (2013), wherein the unit for abundance is nM, and 2) the unit system used by Kozer et al. (2014) and recommended by Faeder et al. (2009) for (stochastic) simulations, wherein the unit for abundance is molecules per cell. The two unit systems are related by the following relation: 1 nM = 10^6 molecules per cell. The unit system for deterministic simulations is largely unimportant; the main issue is that parameter values be given in consistent units. However, stochastic simulation is not compatible with all unit systems. Stochastic simulation requires that the abundance of a chemical species in a reaction compartment (or a subvolume) be specified as the finite number of molecules in the compartment (or the subvolume). See Faeder et al. (2009) for more discussion of units. We use "#/μm^2" and "#/cell" as shorthand for number of EGFR clusters per square micron and number of molecules per cell. Clusters are taken to include EGFR monomers, dimers and higher-order oligomers (Kozer et al., 2013).

[b]Estimates are given for the nine free parameters in BioNetFit fitting runs. Not all model parameters were free to vary in fitting – for a complete description of model parameters and the values assigned to fixed parameters, see Kozer et al. (2013; 2014).

[c]Parameters are referred to by different names in the text of Kozer et al. (2013) and in the model-specification (.bngl) files for the original model and the extended model. Both names are listed for each parameter, except for scaling factors. The first name, if given, is the name used in the main text of Kozer et al. (2013). The second name, which is written using a monospaced font, is the named used in the .bngl files.

[d]This scaling factor relates a simulation output to a measurement reported in Fig. 2B in Kozer et al. (2013): alpha1 × Clusters = EGFR cluster density (clusters per square micron). Clusters are defined differently in the original and extended models. The difference arises because only EGFR monomers, dimers, trimers, linear tetramers and cyclic tetramers are considered in the original model. In contrast, the extended model does not impose a size limit on linear oligomers of EGFR.
[e]This scaling factor relates a simulation output to a measurement reported in Fig. 3B in Kozer et al. (2013): alpha2 × Clusters = EGFR cluster density (clusters per square micron). Clusters are defined differently in the original and extended models. The difference arises because only EGFR monomers, dimers, trimers, linear tetramers and cyclic tetramers are considered in the original model. In contrast, the extended model does not impose a size limit on linear oligomers of EGFR.
[f]This scaling factor relates a simulation output (the value of the observable pEGFR defined in Supplementary File S1 or S2, online at http://bionetfit.nau.edu) to a measurement reported in Fig. 2D in Kozer et al. (2013): alpha3 × pEGFR = relative level of EGFR phosphorylation (arbitrary units or a.u.). The definition of pEGFR is the same in Supplementary Files S1 and S2.
[g]This scaling factor relates a simulation output (the value of the observable pEGFR) to a measurement reported in Fig. 3D in Kozer et al. (2013): alpha4 × pEGFR = relative level of EGFR phosphorylation (arbitrary units or a.u.). The definition of pEGFR is the same in Supplementary Files S1 and S2.
[h]The simulator used in fitting was BioNetGen's default ODE solver (Faeder et al., 2009; Harris et al., submitted), which is CVODE in the SUNDIALS package (Hindmarsh et al., 2005). Kozer et al. (2013) used the same simulator. It should be noted that the fitting procedures used here and by Kozer et al. (2013) are different. For example, Kozer et al. (2013) did not use all data in fitting. We used BioNetFit to obtain a global fit.
[i]The simulator used in fitting was NFsim (Sneddon et al., 2011). This simulator implements a stochastic simulation algorithm that does not require enumeration of the individual reactions implied by the rules of a model.

Table 4.2. Parameter estimates[a] for Example 3

Free parameter name[b]	Monine et al. (2010)	BioNetFit
K_1 (K1)	0.467 nM^{-1}	0.313 nM^{-1}
K_2 (K2)	87.0 nM^{-1}	53.5 nM^{-1}
α (alpha)	0.816	0.832

[a]Estimates are reported for two parameters (the equilibrium association constants K_1 and K_2) of a model for interaction of a soluble Alexa-488-labeled trivalent ligand with bivalent cell-surface receptors and for the parameter of an associated measurement model (α). The dimensionless parameter α relates the fluorescence intensity of cell-associated ligand per cell (on average) to the fraction of the maximum amount of cell-associated ligand per cell, which is set by the total number of receptor sites per cell. See Monine et al. (2010) for additional information about the model and the parameters. The model, which is referred to as the 'TLBR model' by Monine et al. (2010), includes rules that define polymerization-like reactions. The simulator used in fitting was NFsim (Sneddon et al., 2011). This simulator implements a stochastic simulation algorithm that does not require enumeration of the individual reactions implied by the rules of a model. The data used in fitting are from Posner et al. (2007). See also Fig. 2(a) in Monine et al. (2010). Estimates are given for the three parameters that were free to vary in BioNetFit fitting runs. Not all model parameters were free to vary in fitting – for a complete description of model parameters and the values assigned to fixed parameters, see Monine et al. (2010). The quality of fit is illustrated in Figure 3. The parameter estimates found by BioNetFit lie within the confidence limits estimated by Monine et al. (2010). See Table 1 in Monine et al. (2010).
[b]Parameters are referred to by slightly different names in the text of Monine et al. (2010) and in the model-specification (.bngl) file (example3.bngl, which can be found in the 'example3' subdirectory of the 'examples' directory of the BioNetFit distribution). Both names are listed for each parameter. The first name is the name used by Monine et al. (2010). The second name, which is written using a monospaced font, is the named used in the .bngl file.

Table 4.3. Parameter estimates[a] for Example 4

Free parameter name	Chylek et al. (2014b)	BioNetFit
TCRtot	1.00×10^{5} molecules/cell	1.05×10^{5} molecules/cell
krl	2.00×10^{-3} s^{-1}	2.26×10^{-3} s^{-1}
kpLckZap	6 s^{-1}	8.19 s^{-1}
kpLckLck1	10 s^{-1}	7.66 s^{-1}
krZapTcr	0.25 s^{-1}	0.40 s^{-1}
kfl	7.45×10^{5} $M^{-1}s^{-1}$	5.07×10^{5} $M^{-1}s^{-1}$
krPagCsk	10 s^{-1}	21.76 s^{-1}
kfZapLat191	2.16×10^{7} $M^{-1}s^{-1}$	5.51×10^{7} $M^{-1}s^{-1}$
krPagLck	100 s^{-1}	60.80 s^{-1}
kfTcrFyn	3.00×10^{5} $M^{-1}s^{-1}$	8.85×10^{5} $M^{-1}s^{-1}$
kdpLck192	7 s^{-1}	27.67 s^{-1}
kpLckTcrz1	2 s^{-1}	1.11 s^{-1}
krPagPtp	100 s^{-1}	505.40 s^{-1}
KD_PagLck	8.33×10^{-6} M	4.58×10^{-6} M
cyt	0.1 (dimensionless)	0.580 (dimensionless)
CD28tot	1.00×10^{4} molecules/cell	6.41×10^{4} molecules/cell
kfZapLat132	7.20×10^{6} $M^{-1}s^{-1}$	4.66×10^{7} $M^{-1}s^{-1}$
KD_LckCd28	1.44×10^{-6} M	9.66×10^{-6} M
krLckPtp	4.00×10^{-3} s^{-1}	2.92×10^{-2} s^{-1}
krPlcgLat	1 s^{-1}	9.06 s^{-1}
krTcrFyn	1 s^{-1}	0.49 s^{-1}
kfLckPtp	3.50×10^{5} $M^{-1}s^{-1}$	1.57×10^{5} $M^{-1}s^{-1}$
kpLckItk1	3 s^{-1}	1.31 s^{-1}
kdpPag	66 s^{-1}	21.39 s^{-1}
krLcpItk	0.1 s^{-1}	0.029 s^{-1}
kdpLck394	60 s^{-1}	17.14 s^{-1}
KM_ZapLat132	3.19×10^{-5} M	9.01×10^{-6} M
kfZapLcp	1.44×10^{8} $M^{-1}s^{-1}$	3.12×10^{7} $M^{-1}s^{-1}$
kfPagCsk	2.40×10^{7} $M^{-1}s^{-1}$	4.74×10^{6} $M^{-1}s^{-1}$
kfLckCd28	2.50×10^{7} $M^{-1}s^{-1}$	4.47×10^{6} $M^{-1}s^{-1}$
krZapLcp	3 s^{-1}	0.52 s^{-1}
KM_Dok1Ptp	1.54×10^{-5} M	2.60×10^{-6} M
KM_PagPtp	2.08×10^{-5} M	3.10×10^{-6} M
kpWas	35 s^{-1}	1.47 s^{-1}

[a]Estimates are reported for 34 parameters of a model for early events in T-cell receptor (TCR) signaling (Chylek et al., 2014b). The simulator used in fitting was NFsim (Sneddon et al., 2011). This simulator implements a stochastic simulation algorithm that does not require enumeration of the individual reactions implied by the rules of a model. The data used in fitting are temporal

phosphoproteomic data generated using quantitative mass spectrometry (MS)-based proteomics (Chylek et al., 2014b). Estimates are given for the parameters that were free to vary in BioNetFit fitting runs. Not all model parameters were free to vary in fitting – for a complete description of model parameters and the values assigned to fixed parameters, see Chylek et al. (2014b). The quality of fit is illustrated in Figure 4.

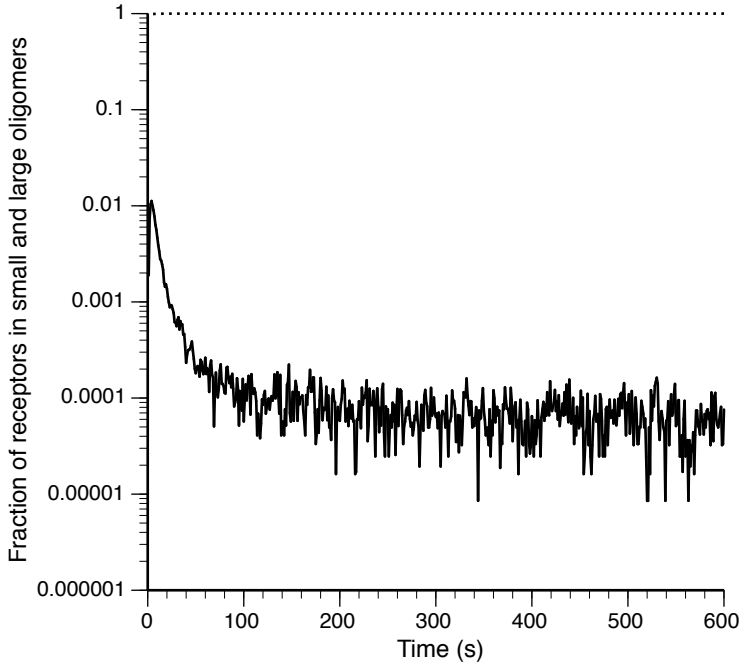


Figure 4.2. Extent of EGFR oligomerization according to the extended model for EGFR activation. This figure shows time courses of EGFR oligomerization after addition of EGF at a dose of 30 nM. NFsim was used to perform 300 replicate stochastic simulation runs (with the best-fit parameter values of Table 1) for a fraction of a cell (25%). Results were then averaged to obtain the results shown here. The dotted line indicates the amount of EGFR in small oligomers (monomers, dimers, trimers, linear tetramers, and cyclic tetramers). The solid line indicates the

amount of EGFR in higher-order linear oligomers (pentamers, hexamers, etc.). The only cyclic oligomer considered in the model is the cyclic tetramer (Kozer et al., 2013; 2014). A three-dimensional structural model of the cyclic tetramer having atomic resolution is shown in Fig. 3 of Chylek et al. (2014a). Large oligomers, meaning oligomers containing more than four receptors, are not predicted to be present at significant levels at steady state over a wide range of EGF doses (simulation results not shown). As illustrated here, large oligomers are predicted to form at low levels. At most, only ~1% of receptors are present in large oligomers and this level is transient.

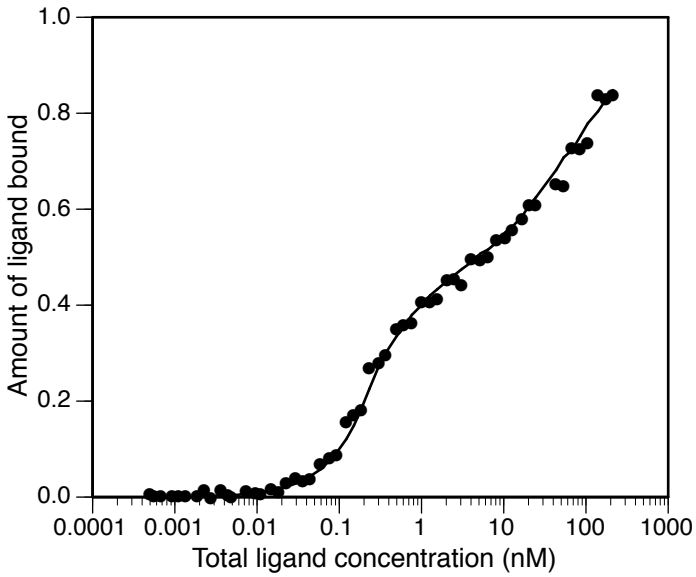


Figure 4.3. Quality of fit found by BioNetFit for Example 3. See the example3 subdirectory of the examples directory of the BioNetFit distribution for the .bngl, .exp, and .conf files needed to reproduce the fitting demonstration. The .bngl file defines a model for trivalent ligand interaction with a bivalent cell-surface receptor. This model is called the "TLBR model" by Monine et al. (2010). The .exp file contains flow cytometric data characterizing the amount of cell-associated Alexa-488-labeled ligand at equilibrium for different ligand doses; solid dots in the figure correspond to averaged measurements of fluorescence intensity per cell. The data points are from Posner et al. (2007); see also Panel a in Fig. 2 of Monine et al. (2010). The flow cytometric data have been scaled using the value estimated for α such that a value of 1 on the *y*-axis corresponds to saturation of receptor binding sites. Parameter estimates are summarized in Table 2. The curve in the figure represents averaged stochastic simulation results obtained using best-fit parameter values (Table 2) and NFsim (Sneddon et al., 2011): 18 simulation runs to steady state with *f* (fraction of cell) = 0.001.

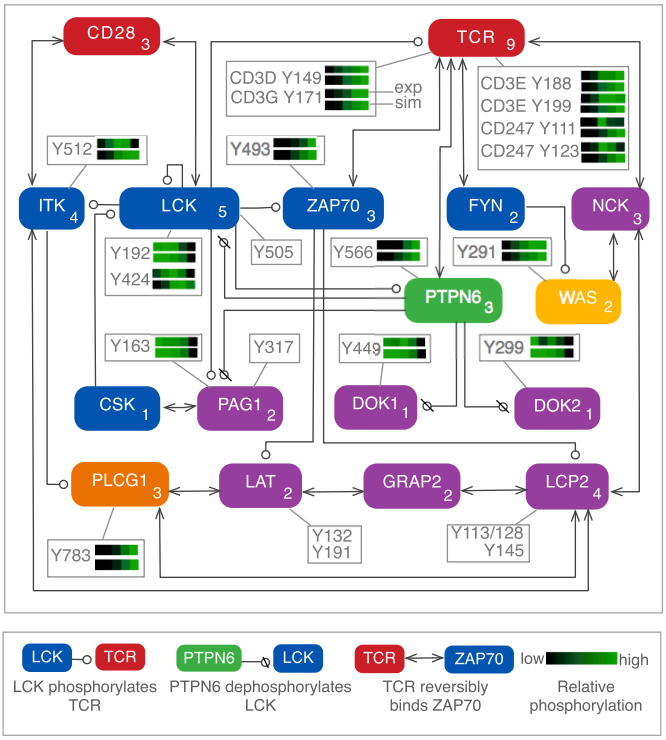


Figure 4.4. Model for TCR signaling. Proteins considered in a rule-based model for TCR signaling are represented by rounded boxes. Separate boxes indicate the phosphosites considered in the model. Sites detected in phosphoproteomic experiments are each associated with a pair of heatmaps, in which the upper heatmap reflects averaged experimental measurements of relative pTyr abundance and the lower heatmap reflects simulated phosphorylation levels at matching time points. The color scale of each heatmap is unique: black represents the lowest and green represents the highest level of phosphorylation for that site. Interactions are represented by arrows according to the conventions illustrated at bottom. The number in the lower right corner of a protein box represents the number of components of the protein (domains, motifs, and/or pTyr sites) considered in the model.

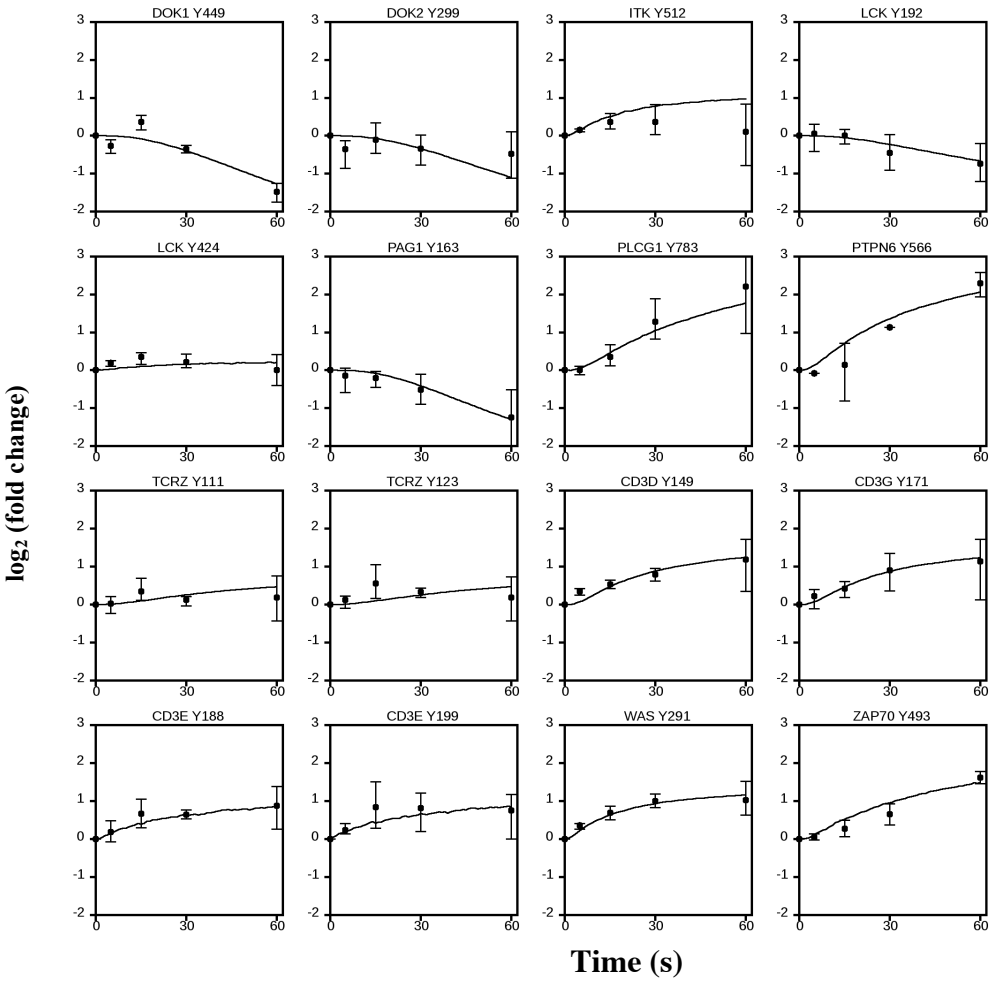


Figure 4.5. Quality of fit found by BioNetFit for Example 4. See the example4 subdirectory of the examples directory of the BioNetFit distribution for the .bngl, .exp, and .conf files needed to reproduce the fitting demonstration. The .bngl file defines a model for early events in T-cell receptor (TCR) signaling (Chylek et al., 2014b). The .exp file contains time courses of phosphorylation for various tyrosine residues measured using quantitative mass spectrometry (MS)-based proteomics (Chylek et al., 2014b). For each time course, the measurement at time t = 0 (baseline) is normalized. In other words, phosphorylation levels were measured relative the baseline level of phosphorylation in experiments. Simulation results were similarly normalized for comparison purposes in the fitting run. Parameter estimates are summarized in Table 3. The curves in the various figure panels represent averaged stochastic simulation results obtained using best-fit parameter values (Table 3) and NFsim (Sneddon et al., 2011): 500 simulation runs with f (fraction of cell) = 0.1.

References

Blinov,M.L. *et al*. (2004) BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, **20**, 3289–3291.

Blinov,M.L. *et al*. (2006) Graph theory for rule-based modeling of biochemical networks. *Lect. Notes Comput. Sci*., **4230**, 89–106.

Chylek,L.A. *et al*. (2013) Innovations of the rule-based modeling approach. In A. Prokop,A. and Csukás,B. (eds.), *Systems Biology: Integrative Biology and Simulation Tools*, Springer, Dordrecht, Vol. 1, pp. 273–300.

Chylek,L.A. et al. (2014a) Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. Wiley Interdiscip. Rev. Syst. Biol. Med., 6, 13–36.

Chylek,L.A. et al. (2014b) Phosphorylation site dynamics of early T-cell recep- tor signaling. Plos One, 9, e104240.

Chylek,L.A. *et al*. (2015) Modeling for (physical) biologists: an introduction to the rule-based approach. *Phys. Biol*. **12**, 045007.

Colvin,J. *et al*. (2009) Simulation of large-scale rule-based models. *Bioinformatics*, **25**, 910–917.

Colvin,J. *et al*. (2010) RuleMonkey: software for stochastic simulation of rule-based models. *BMC Bioinformatics*, **11**, 404.

Danos,V. et al. (2007) Scalable simulation of cellular signaling networks. Lect. Notes Comput. Sci., 4807, 139–157.

Danos,V. *et al*. (2007) Scalable simulation of cellular signaling networks. *Lect. Notes Comput. Sci*., **4807**, 139–157.

Elleman,T.C. *et al*. (2001) Identification of a determinant of epidermal growth factor receptor ligand-binding specificity using a truncated, high-affinity form of the ectodomain. *Biochemistry*, **40**, 8930–8939.

Faeder,J.R. et al. (2005) Rule-based modeling of biochemical networks. *Complexity*, **10**, 22–41.

Faeder,J.R. *et al*. (2009) Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol. Biol*., **500**, 113–167.

Harris,L.A. et al. (2015) BioNetGen 2.2: advances in rule-based modeling. Bioinformatics http://arxiv.org/pdf/1507.03572.pdf

Hindmarsh,A.C. *et al*. (2005) SUNDIALS: suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw*., **31**, 363–396.

Hogg,J.S. *et al*. (2014) Exact hybrid particle/population simulation of rule-based models of biochemical systems. *PLOS Comput. Biol*., **10**, e1003544.

Hucka,M. et al. (2003) The systems biology markup language (SBML): a me- dium for representation and exchange of biochemical network models. Bioinformatics, 19, 524–531.

Kleiman,L. *et al*. (2011) Rapid phospho-turnover by receptor tyrosine kinases impacts downstream signaling and drug binding. *Mol. Cell*, **43**, 723–737.

Kozer,N. et al. (2013) Exploring higher-order oligomerisation and phosphorylation—a combined experimental and theoretical approach. Mol. BioSyst., 9, 1849–1863.

Kozer,N. et al. (2014) Recruitment of the adaptor protein Grb2 to EGFR tetra- mers. Biochemistry, 53, 2594–2604.

Low-Nam,S.T. *et al*. (2011) ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding. *Nat. Struct. Mol. Biol*., **18**, 1244–1249.

Macdonald,J.L. and Pike,L.J. (2008) Heterogeneity in EGF-binding affinities arises from negative cooperativity in an aggregating system. *Proc. Natl. Acad. Sci. USA*, **105**, 112–117.

Monine,M.I. et al. (2010) Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. Biophys. J., 98, 48–56.

Posner,R.G. *et al*. (2007) Trivalent antigens for degranulation of mast cells. *Org. Lett*. **9**, 3551–3554.

Press,W.H. et al. (2007) Numerical Recipes: The Art of Scientific Computing, 3rd Edition. Cambridge University Press, Cambridge, UK.

Smith,J.E. and Eiben,A.E. (2008) Introduction to Evolutionary Computing. Springer, Berlin.

Sneddon,M.W. et al. (2011) Efficient modeling, simulation and coarse-grain ing of biological complexity with NFsim. Nat. Methods, 8, 177–183.

Stefan,M.I. et al. (2014) Multi-state modeling of biomolecules. PLOS Comput. Biol., 10, e1003844.

Whitley,D. (1994) A genetic algorithm tutorial Stat. Comput., 4, 65–85.

Yang,J. *et al*. (2008) Kinetic Monte Carlo method for rule-based modeling of biochemical networks. *Phys. Rev. E*., **78**, 031910.

Yang,J. and Hlavacek,W.S. (2011) The efficiency of reactant site sampling in network-free simulation of rule-based models for biochemical systems. *Phys. Biol*., **8,** 055009.

Chapter 5: Summary & Outlook

Ligand-induced receptor oligomerization followed by intracellular phosphorylation is a recurring theme of cell signaling in immune cells and beyond (Lemmon & Schlessinger, 1994). These processes are inherently complex. How tightly does the ligand bind to the receptor? How many phosphorylation sites does the receptor have, and what do these sites bind to? Do downstream events feed back onto earlier steps? The answers to these questions and others can determine how a cell turns an input into an output, which can in turn have decisive repercussions for the physiological state of the organism. However, to answer such questions we need to consider how these systems' abundant interconnections are influenced by quantitative elements, which can quickly confound human intuition (Tyson et al., 2003).

A model is a tool that amplifies our intuition. Models enable us to develop nontrivial hypotheses, which can then be tested experimentally to enable model refinement. Models can take many forms, from pictures to equations to computer programs. In our work, we have used rule-based models of chemical kinetics (Chylek et al., 2014a) to investigate signaling initiated by FcεRI, and linked these models to several types of experimental data.

We began by developing an interaction library for the FcεRI signaling network, which can be used to specify a variety of models (Chapter 1). This library combines previously developed models for receptor-mediated signaling with new components that have not previously been modeled. The library is visualized and annotated. We identified several different classes of network motifs (Alon, 2007) within the library, each of which might serve a specialized purpose.

To investigate the potential roles of these motifs in signaling (Chapter 2), we modeled circuits with and without positive feedback. We found that a positive feedback loop involving the adaptor protein Gab2 may enable some downstream responses to be robust to variation in upstream receptor phosphorylation. This outcome has been observed experimentally in studies using a set of different ligands (Sil et al., 2007), and thus our model is a proposed explanation for these results. We also developed a model for oscillations in intracellular Ca^{2+}, which is an important part of the response to antigen stimulation (Holowka et al., 2012). This model is based on the hypothesis that oscillations in Ca^{2+} dynamics result from oscillations in receptor phosphorylation caused by positive and negative feedbacks. Preliminary experimental data suggest that oscillations in receptor phosphorylation occur.

We then explored another aspect of signaling dynamics: how mast cells respond to inputs that vary over time (Chapter 3). Specifically, we assessed how cells respond to paired pulses of a stimulatory multivalent ligand, separated by a period of exposure to a non-stimulatory (quiescent) monovalent one. Experimentally, we found that short intervals of quiescence led to reduced degranultion responses during the second stimulation, whereas longer periods of quiescence led to enhanced responses. We developed a model encapsulating a proposed explanation for this phenomenon. Through an iterative process of modeling and experimental tests, we learned that a tug-of-war between positive signals from the tyrosine kinase Syk and negative signals from the lipid phosphatase Ship-1 lead to the observed differences. Furthermore, the adaptor protein Shc1 influences how the balance of positive and negative signals changes with time.

To facilitate the linking of models to experiments, we developed a new fitting tool (Chapter 4). In the past, it has been difficult to fit large models to experimental data, because

most conventional fitting programs are designed for models that are in the form of ordinary differential equations derived from a reaction network. Many cell signaling models cannot be translated into this form, and thus must be simulated using network-free methods. To make it possible to fit such models, we developed a fitting tool compatible with network-free simulators (Sneddon et al., 2011). This tool implements a genetic algorithm, which enables fitting of multiple parameters simultaneously, and also makes it possible to leverage parallel computing resources for greater efficiency. We demonstrated use of this tool using several test problems.

We also used this tool in a study of ligand-receptor interactions (Appendix 1). We developed a model for interactions of a trivalent ligand with a bivalent receptor, and parameterized it based on experimental data. The dissociation rate constant was estimated based on fluorescence quenching experiments, and other parameters were optimized to fit measurements of receptor aggregate size obtained through super-resolution imaging.

Finally, we also developed techniques for model visualization and annotation (Appendix 2) so that models can be represented in a succinct and accessible manner, and linked to the underlying biological knowledge that was used to construct each of the model's components.

Future Directions.

A next step would be to improve the ligand-receptor binding model (Appendix 1). The current model has some discrepancies in kinetics when compared to the aggregation data. These discrepancies could likely be reduced by implementing aggregate immobilization, which will make the process of aggregation more gradual. Immobilization would mean that receptor aggregates above a threshold size would be slow to bind, as well as slow to dissociate. Current general-purpose simulation tools are able to implement the former (ref?) but not the latter.

It may also be possible to develop a problem-specific code for simulating aggregation and immobilization; however, the downside of a problem-specific code is that it will be increasingly difficult to add in the steps of intracellular signaling. In contrast, a ligand-receptor binding model specified with general-purpose tools could be coupled to signaling components (e.g., the rules presented in Chapter 1) with relative ease. We want to achieve an understanding of how aggregation initiates signaling, and thus implementing immobilization with a general-purpose simulation tool will be the preferable solution.

In the two parts of Chapter 2, we considered two aspects of intracellular signaling: 1) how ligands that induce different levels of receptor phosphorylation can bring about similar levels of certain downstream readouts, and 2) how oscillations in Ca^{2+} might arise from oscillations in receptor phosphorylation. A future step could be to combine features of these two models, and to ask the question of whether different ligands induce different patterns of Ca^{2+} oscillations. (Accurate models for ligand-receptor interactions will be especially useful in this endeavor.) A study of this type will give us a better understanding of how various antigens influence a central feature of mast cell physiology.

Finally, with the rule library (Chapter 1) and fitting tool (Chapter 4) as groundwork, the door may be opened to development of models that can be matched to experimental data for a large set of outputs. It has already been demonstrated that models of chemical kinetics can be used to analyze multiplexed data, and that such analyses can yield insights and predictions distinct from what is obtainable through standard bioinformatic techniques (Chylek et al., 2014b; Stites et al., 2015). By taking advantage of these emerging approaches, we are well-positioned to further our understanding of how a multitude of molecular players are orchestrated during FcεRI signaling.

References

Alon, U. Network motifs: theory and experimental approaches. *Nat Rev Genet*. (2007) **8**, 450-61.

Chylek, L.A., Harris, L.A., Tung, C.S., Faeder, J.R., Lopez, C.F., Hlavacek, W.S. Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. *Wiley Interdiscip Rev Syst Biol Med*. (2014a) **6**, 13-36.

Chylek, L.A., Akimov, V., Dengjel, J., Rigbolt, K.T., Hu, B., Hlavacek, W.S., Blagoev, B. Phosphorylation site dynamics of early T-cell receptor signaling. *PLoS One*. (2014b) **9**, e104240.

Holowka, D., Calloway, N., Cohen, R., Gadi, D., Lee, J., Smith, N.L., Baird, B. Roles for Ca(2+) mobilization and its regulation in mast cell functions. *Front Immunol*. (2012) **3**, 104.

Lemmon, M.A., Schlessinger, J. Regulation of signal transduction and signal diversity by receptor oligomerization. *Trends Biochem Sci*. (1994) **19**, 459-463.

Sil, D., Lee, J.B., Luo, D., Holowka, D., Baird, B. Trivalent ligands with rigid DNA spacers reveal structural requirements for IgE receptor signaling in RBL mast cells. *ACS Chem Biol*. (2007) **2**, 674-684.

Sneddon, M.W., Faeder, J.R., Emonet, T. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat Methods*. (2011) **8**, 177-83.

Stites, E.C., Aziz, M., Creamer, M.S., Von Hoff, D.D., Posner, R.G., Hlavacek, W.S. Use of mechanistic models to integrate and analyze multiple proteomic datasets. *Biophys J*. (2015) **108**, 1819-1829.

Tyson, J.J., Chen, K.C., Novak, B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol*. (2003) **15**, 221-231.

Appendix 1: Binding properties of a structurally defined ligand

Structurally well-defined ligands are a useful tool in studying IgE receptor signaling because they help us understand what features of a ligand are conducive to cellular responses. Examples of such ligands are ones formed by conjugating DNP hapten groups to a rigid DNA scaffold consisting of double-stranded DNA (Sil et al., 2007). The scaffold can be considered rigid because the length of the arm structure is less than the persistence length of DNA (Brinker et al., 2009). These ligands have been studied in the past to evaluate how the distance between hapten groups, as defined by the number of DNA base pairs that make up the arms of the ligand, affects intracellular signaling. It was found that phosphorylation of the receptor and the adapter protein Lat, as well as degranulation, was strongly impacted by the distance between haptens. The most potent of these ligands was Y16, so named because it is a Y-shaped structure formed by three complimentary DNA strands with 16 nucleotide residues in each strand.

To enhance our understanding of how this ligand operates, we endeavored to integrate quantitative experimental data for receptor aggregation with computational models for ligand-receptor binding. This appendix summarizes the progress made towards these ends.

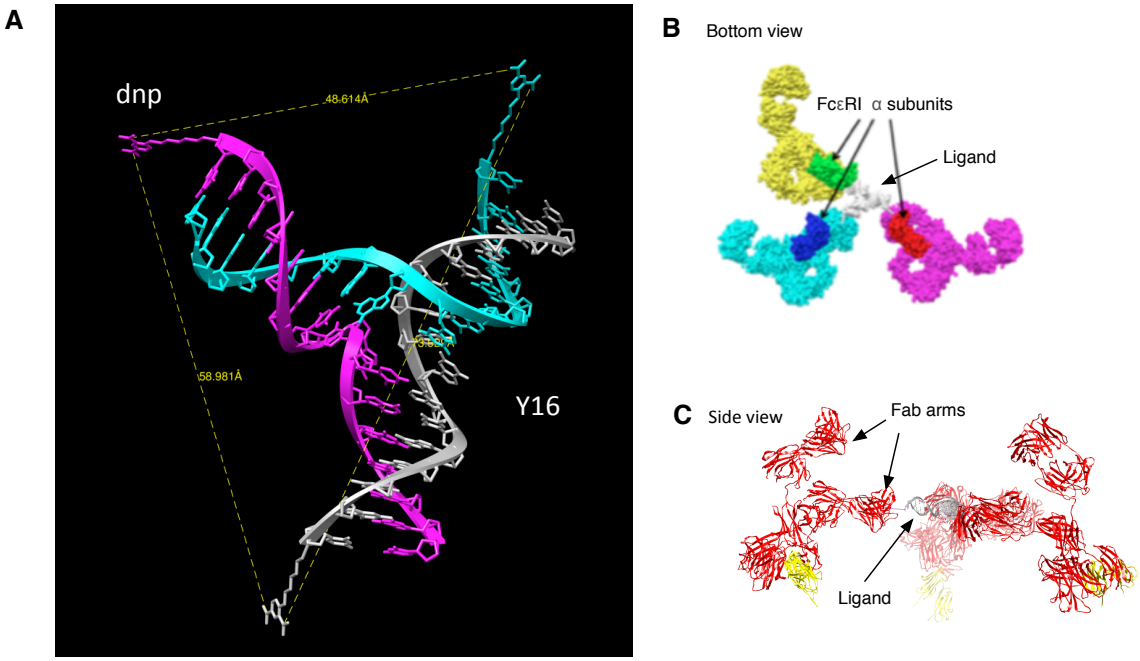


Figure A.1.1. Structural models of the Y16 trivalent ligand constructed using the motif binding geometry method (Tung & McMahon, 2012). A. An image of Y16, a ligand formed from three DNA strands whose ends are conjugated to DNP. The DNA junction structure is based on that of Thiviyanathan et al. (1999), PDB ID 1EKW. The distances between DNP haptens (measurable from the structural model, using PDB Viewer, http://spdbv.vital-it.ch/content.html) are 5.9 nm and 4.9 nm, which is within the range of 5.2±0.9 that was previously measured in FRET experiments (Sil et al., 2007). The distance of a single DNA arm, including the linker and DNP group, is approximately 3.7 nm, which is larger than the FRET-measured 2.7 nm. This discrepancy may arise from the flexibility of the linkers connecting the DNA arm to the DNP group, which is not accounted for in this static model. B. A bottom view of Y16 bound to three IgE-alpha complexes, as seen from below the cell membrane. C. Side view, showing the Fab arms of IgE. Credit: Chang-Shung Tung (LANL)

Fluorescence measurements for receptor dissociation

We first performed kinetic measurements to ascertain a dissociation rate constant. These experiments were performed by monitoring bulk fluorescence via a fluorimeter (Erickson et al., 1987). Cells were sensitized with IgE labeled with flourescein isothiocyanate (FITC). The fluorescence of FITC is quenched in the presence of DNP. Thus, saturation of IgE binding sites can be detected when adding increasing amounts of a DNP ligand causes no further quenching.

Quenching is caused not just by the DNP that is in direct contact with the IgE, but also by other DNP groups that are in the vicinity. Thus, a trivalent ligand, with three DNP groups, quenches FITC to a greater extent than a monovalent ligand that contains only one. In the presence of an excess of monovalent DCT, the monovalent hapten displaces the previously bound Y16 (Fig. 2). This displacement event causes an increase in fluorescence. The change in fluorescence can be transformed into a curve showing the fraction of IgE sites bound as a function of time (Posner et al., 1991). The decay curve can be fit by an exponential function, which yields an estimate of the rate constant for dissociation of a DNP site in Y16 from anti-DNP IgE (Fig. 3). The equation used for fitting is the following:

$$B(t) = B(0)*e^{-kt}$$

Where *B(t)* is the number of bound sites at time *t*, and *B(0)* is the number of bound sites at *t*=0. The parameter that is optimized in fitting is *k*, which is the dissociation rate constant. Optimization was performed using the *nlinfit* function of Matlab, and confidence intervals were obtained using bootstrapping (Press, 2007). This model assumes that a single dissociation rate constant is sufficient to characterize the data.

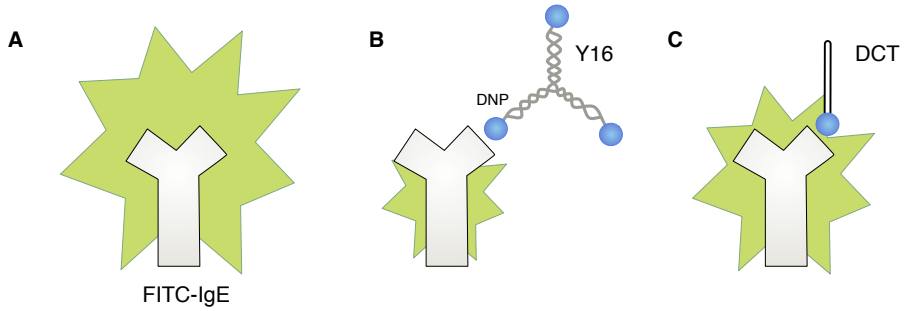


Figure A.1.2.: In dissociation experiments, FITC-labeled IgE is at its maximum fluorescence when fully unbound. Its fluorescence is quenched in the presence of Y16, due to this ligand's

three DNP groups. Its fluorescence is quenched to a lesser extent in the presence of DCT, which has only one DNP group.

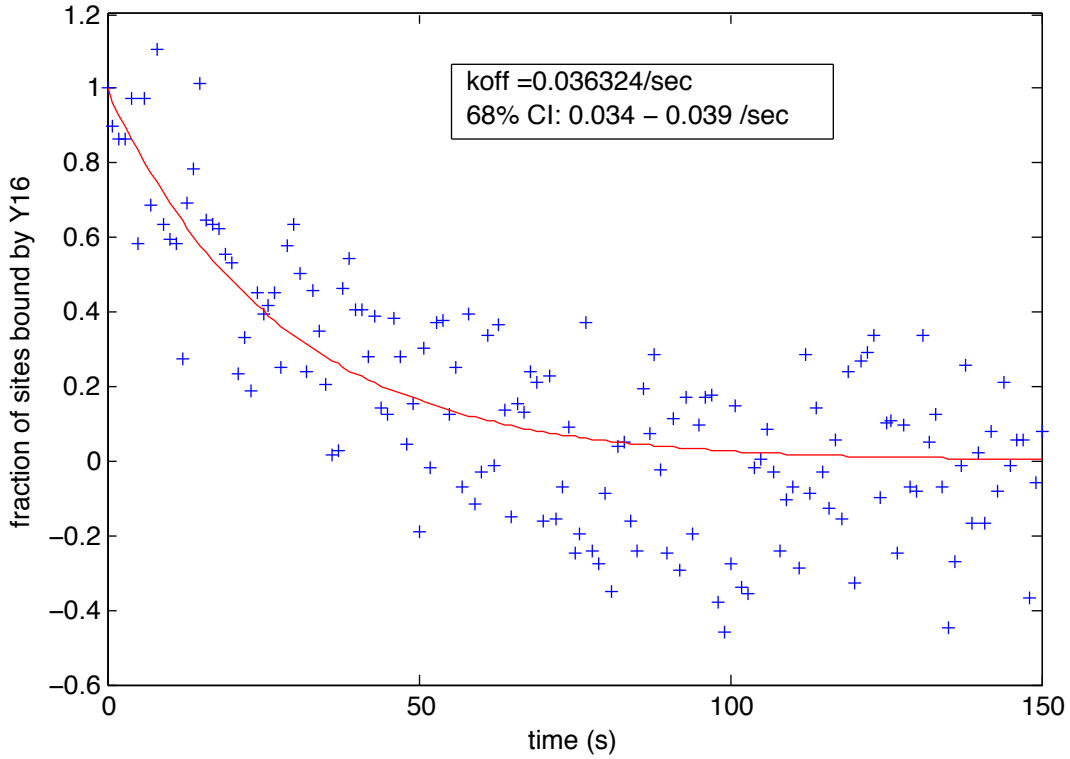


Figure A.1.3. An example time course of dissociation of Y16-DNP from FITC-anti-DNP IgE. At time 0, FITC quenching was maximal due to all receptor sites being bound by Y16-DNP, which was at a concentration of 2.1 nM. At this point, a 50-fold excess (in terms of DNP sites) of monovalent DCT was added. The resulting climb in fluorescence was transformed into the above plot for fraction of binding sites occupied by Y16 as a function of time. The data was fit to a single-exponential curve to obtain an estimate of the dissociation rate constant (koff).

Development and parameterization of a ligand-receptor binding model

Using a rule-based approach, we developed a model for interactions between a trivalent ligand and a bivalent receptor. The molecule type definitions, meaning the molecules included in the model and their constituent parts, are shown in Fig. 4. The rules of the model are shown in Fig. 5.

The model that we used to fit these data consisted of rules for 1) ligand-receptor binding, 2) receptor cross-linking, and 3) formation of cyclic receptor dimers. The parameters that were fit were the equilibrium constants and rate constants for each interaction.

Parameter fitting was performed using BioNetFit (Thomas et al., 2015). The dissociation rate constant was constrained to the range of values that had been estimated experimentally from fluorescence data (a mean of 0.05 /s, and standard deviation of 0.047 /s). The equilibrium constants were fitted freely. The forward rate constants were determined by the ratio of equilibrium constants to the dissociation rate constant (which we assume took only a single value). The experimental data used in fitting were stochastic optical reconstruction microscopy (STORM) measurements, which can characterize the average number of receptor per aggregate as a function of time (Shelby et al., 2013). The best-fit parameters that the software was able to obtain are shown in Table 1, and a plot of simulated vs. measured aggregate size is shown in Fig. 6. In previous studies (Sil et al., 2007), a single equilibrium constant had been estimated based on fluorescence titration experiments, with the value of 6.67e8 /M. Although this value is close to our fitted K value of 6.78e8 /M, our fit may not be unique, an issue that could be explored in more detail by obtaining a bootstrap confidence interval (Press, 2007).

Table A.1.1. Summary of parameter estimates.

Parameter name	Description	Value
K	Equilibrium constant for ligand capture	6.76e8 /M

Kx	Equilibrium constant for receptor crosslinking	5.92e11 /M
J	Equilibrium constant for cyclic dimer formation	6.44e1 (dimensionless)
koff	Dissociation rate constant	0.041 /s

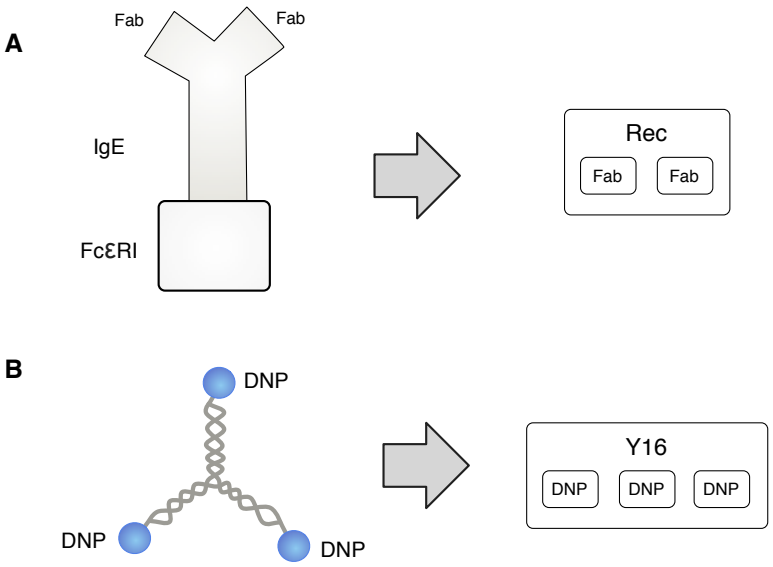


Figure A.1.4. Molecule type definitions used in the model of trivalent ligand-bivalent receptor interactions. A. The IgE-FcRI complex, the receptor, is represented as a molecule with two identical binding sites that are both named Fab, for the Fab arms of the IgE antibody that interact with antigen. In BioNetGen language, the molecule type definition is `Rec(Fab,Fab)`. B. Y16, the trivalent DNP ligand, is represented as a molecule with three identical sites for interaction with the receptor. In BioNetGen Languge, the definition is `Y16(DNP,DNP,DNP)`.


A
Rec
Fab
+
Y16
DNP
DNP
DNP
Rec
Fab
Y16
DNP
DNP
DNP
B
Rec
Fab
Y16
DNP
DNP
Rec
Fab
Y16
DNP
DNP
C
Rec
Fab
Y16
DNP
Rec
Fab
Y16
DNP
D
Rec
Fab
Fab
Y16
DNP
DNP
DNP
Y16
DNP
DNP
Rec
Fab
Fab
E
F
G

Figure A.1.5. Rules for the trivalent ligand-bivalent receptor model. A. Binding of the ligand to the receptor. On the left-hand side of the rule, the ligand is specified to have three unbound DNP sites, because it is freely diffusing in the solution and not tethered to the cell surface. The receptor needs to have a free binding site for the interaction to proceed. The state of the other binding site is assumed not to affect this interaction, and thus is omitted from the rule. B. Receptor crosslinking. This rule describes binding to Y16 to a receptor when Y16 is already tethered to the cell surface by having bound another receptor (indicated by the box connected to one DNP of Y16). C. Dissociation of a ligand-receptor bond. D-F: Rules for formation of cyclic dimers. These separate rules need to be specified to exclude formation of a sterically unlikely multicyclic aggregate, which is shown in panel G.

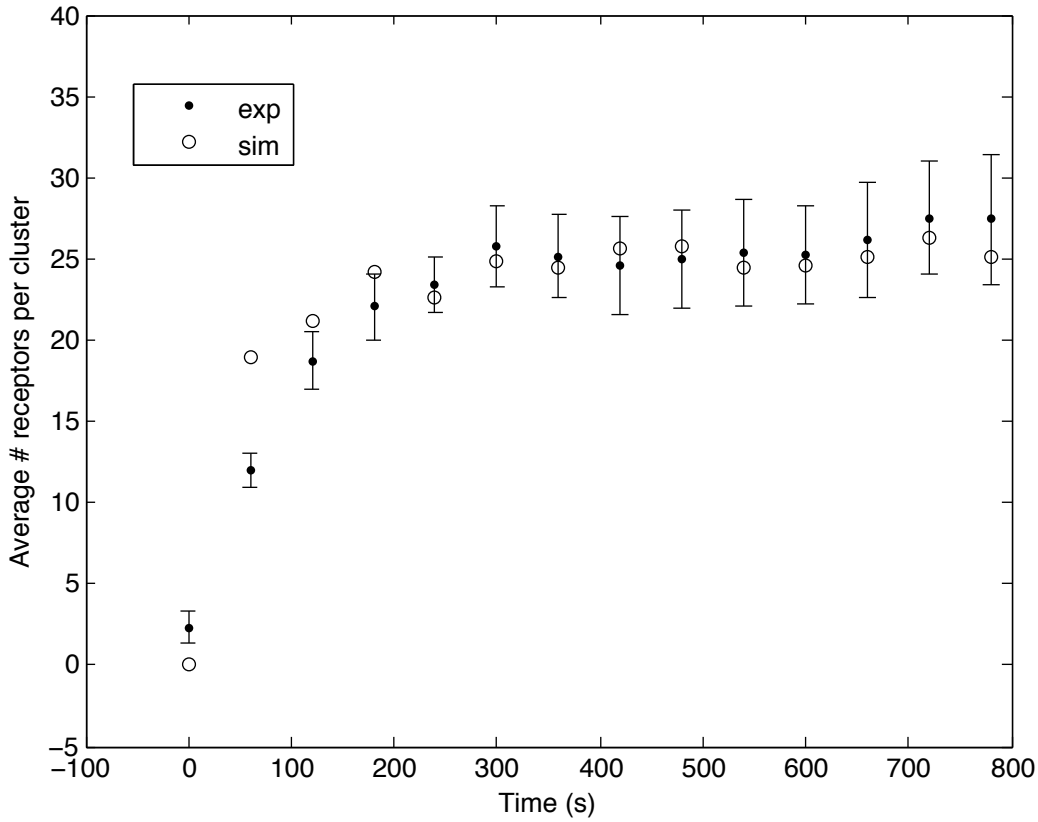


Figure A.1.6. STORM measurements (black circles with error bars) and simulation results (open circles) from the fitted model for average number of receptors per aggregate. Credit for experimental data: Eshan Mitra.

Future directions

The resulting fit shows good agreement once aggregation seems to reach a steady-state. However, the early kinetics of aggregation are faster in the simulation than in the experiment. This discrepancy is likely due to the model not being able to capture the slowing of diffusion of receptor aggregates. This problem could be resolved through implementation of aggregate immobilization, reflecting a reduction in diffusion rate. Immobilization would be implemented as a reduction in the association rate constant (e.g., to 10% of the original value) for two aggregates above a threshold size, paired with an equal reduction in the rate constant for dissociation of an aggregate into two aggregates that are both above the threshold size.

Forward reactions can be immobilized (i.e., forced to occur at a reduced rate) if the aggregate reactants are above a certain size. However, this immobilization cannot be enforced for dissociation, which is necessary to preserve equilibrium constants and capture the physical reality of how immobilized aggregates are likely to move on the cell surface. Modification of the NFsim simulation code (Sneddon et al., 2011) will be necessary to enable immobilization of receptors during dissociation.

An alternative approach would be to develop a problem-specific rule-based modeling code for this problem (Monine et al., 2010), and/or to use spatial modeling (Kochanczyk and Lipniacki, 2012). However, the drawback of such approaches is that extending the model to account for intracellular signaling will be an increasingly difficult challenge, whereas adding intracellular signaling to a BioNetGen model will be more straightforward (see Chapter 1 for examples of rules for signaling that can be coupled to rules for ligand-receptor binding). Thus, the most promising direction for achieving agreement between simulated and experimental time

courses of aggregate formation will be correcting BioNetGen's implementation of aggregation immobilization.

References

Brinkers, S., Dietrich, H.R., de Groote, F.H., Young, I.T., Rieger, B. The persistence length of double stranded DNA determined using dark field tethered particle motion. *J Chem Phys*. (2009) **130**, 215105.

Erickson, J., Goldstein B, Holowka D, Baird B. The effect of receptor density on the forward rate constant for binding of ligands to cell surface receptors. *Biophys J*. (1987) **52**, 657-62.

Kochanczyk, M., Lipniacki, T. SpatKin: Spatially-resolved rule-based modeling of biochemical systems on the membrane. Abstract at the Sixth q-bio conference, August 9th, 2012.

Posner, R.G., Erickson, J.W., Holowka, D., Baird, B., Goldstein, B. Dissociation kinetics of bivalent ligand-immunoglobulin E aggregates in solution. *Biochemistry*. (1991) **30**, 2348-56.

Press, W.H. Numerical Recipes, 3rd Edition: The Art of Scientific Computing. (2007) Cambridge University Press.

Shelby, S.A., Holowka, D., Baird, B., Veatch, S.L. Distinct stages of stimulated FcεRI receptor clustering and immobilization are identified through superresolution imaging. Biophys J. (2013) 105, 2343-54.

Sil D, Lee J, Luo D, Holowka D, Baird B. Trivalent ligands with rigid DNA spacers reveal structural requirements for IgE receptor signaling in RBL mast cells. ACS Chem Biol (2007) 2:674–84.

Sneddon, M.W., Faeder, J.R., Emonet, T. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat Methods*. (2011) **8**, 177-83.

Thiviyanathan, V., Luxon, B.A., Leontis, N.B., Illangasekare, N., Donne, D.G., Gorenstein, D.G. Hybrid-hybrid matrix structural refinement of a DNA three-way junction from 3D NOESY-NOESY. *J Biomol NMR*. (1999) **14**, 209-21.

Thomas, B.R., Chylek, L.A., Colvin, J., Sirimulla, S., Clayton, A.H., Hlavacek, W.S., Posner, R.G. BioNetFit: a fitting tool compatible with BioNetGen, NFsim and distributed computing environments. *Bioinformatics*. (2015) **pii**, btv655.

Tung, C.S., McMahon, B.H. A structural model of the E. coli PhoB dimer in the transcription initiation complex. *BMC Struct Biol*. (2012) **12**, 3

Yang, J., Hlavacek, W.S. The efficiency of reactant site sampling in network-free simulation of rule-based models for biochemical systems. *Phys Biol*. (2011) **8**, 055009.

Appendix 2: Visualizing rule-based models [1]

Rule-based modeling provides a means to represent cell signaling systems in a way that captures site-specific details of molecular interactions. For rule-based models to be more widely understood and (re)used, conventions for model visualization and annotation are needed. We have developed the concepts of an extended contact map and a model guide for illustrating and annotating rule-based models. An extended contact map represents the scope of a model by providing an illustration of each molecule, molecular component, direct physical interaction, post-translational modification, and enzyme-substrate relationship considered in a model. A map can also illustrate allosteric effects, structural relationships among molecular components, and compartmental locations of molecules. A model guide associates elements of a contact map with annotation and elements of an underlying model, which may be fully or partially specified. A guide can also serve to document the biological knowledge upon which a model is based. We provide examples of a map and guide for a published rule-based model that characterizes early events in IgE receptor (FcεRI) signaling. We also provide examples of how to visualize a variety of processes that are common in cell signaling systems but not considered in the example model, such as ubiquitination. An extended contact map and an associated guide can document knowledge of a cell signaling system in a form that is visual as well as executable. As a tool for

[1] This work was published as: **Chylek LA**, Hu B, Blinov ML, Emonet T, Faeder JR, Goldstein B, Gutenkunst RN, Haugh JM, Lipniacki T, Posner RG, Yang J, Hlavacek WS. (2011) Guidelines for visualizing and annotating rule-based models. *Molecular BioSystems* **7**, 2779 - 2795 (A top-ten most accessed article of June 2011)

model annotation, a map and guide can communicate the content of a model clearly and with precision, even for large models.

Introduction

Cellular responses to environmental changes and signals are mediated by cell signaling systems. A cell signaling system is composed largely of a network of interacting proteins, which are responsible for information processing. A typical signaling protein contains multiple functional components. The components found in signaling proteins include catalytic domains (1,2), modular protein interaction domains (3), linear motifs (4), and sites of post-translational modification (5). Understanding the functional roles of protein components or sites is critical for a thorough understanding of cell signaling, because protein interactions generally depend on site-specific details. For example, many protein-protein interactions are modulated by tyrosine phosphorylation (6). A large amount of information is available about the site-specific details of protein interactions. There is a need to be able to use this information to make predictions about system behaviors. In other words, we need mathematical/computational models to better understand cell signaling, which is complex (7,8).

With recent developments in simulation methodology (9–13), rule-based modeling (14), discussed in detail below, now offers a viable approach for studying large numbers of protein interactions with consideration of site-specific details. Here, with the goal of making this modeling approach more accessible, we demonstrate how rule-based models can be better visualized and annotated, which is important for modeling efforts that aim to comprehensively

capture the molecules and interactions involved in an entire cell signaling system or set of systems. A large, detailed model is of limited use unless it is presented in an understandable manner. The proteins and interactions included in a model, as well as the justification for modeling assumptions, should be communicated clearly and precisely if a model is to be understood, critically evaluated, and reused. To enable clear communication of rule-based models, we introduce the concept of an extended contact map, which serves to illustrate the scope of a rule-based model. We also introduce the concept of an associated model guide. A model guide attaches rules, which are formal representations of interactions, to arrows in an extended contact map. It also attaches molecule type definitions, which are formal representations of molecules, to boxes in a map. A map and a guide that annotates a complete model together provide a visual and executable means to document information about the site-specific details of molecular interactions in a cell signaling system. We expect that the concepts presented here should be useful for modelers as well as others interested in applying systems approaches to the study of cell signaling.

A rule-based model for early events in FcεRI signaling

We will use the model of Goldstein et al. (22) and Faeder et al. (23), an early application of the rule-based modeling approach, to exemplify the basic conventions of an extended contact map and a model guide. Here, we provide an overview of this model, which we will refer to as the FcεRI model.

The FcεRI model (22,23) is composed of 19 rules in total, and it captures early events in IgE receptor (FcεRI) signaling, which triggers allergic reactions. The receptor is composed of an α chain, a β chain, and a homodimer of two disulfide-linked γ chains. The extracellular portion of the α chain binds the Fc portion of IgE (54); the interaction is long lived (55). The β and γ chains each contain an immunoreceptor tyrosine-based activation motif (ITAM) (56), a linear motif. Signaling is initiated when a multivalent antigen or other receptor crosslinking reagent bridges two receptors. In the model, receptor crosslinking is taken to be mediated by a chemically crosslinked dimer of IgE. Following receptor aggregation, the kinase Lyn, which constitutively interacts with the β chain, phosphorylates the β and γ ITAMs in neighboring receptors. As a result, the receptor can recruit Lyn and Syk, a second kinase involved in FcεRI signaling, through phosphorylation-dependent interactions. Syk is phosphorylated via two mechanisms: Lyn phosphorylates tyrosine residues in the linker region, and Syk trans-phosphorylates tyrosine residues in the activation loop of the kinase domain of a neighboring copy of Syk. In the model, all phosphorylation events are reversed by unspecified phosphatases, which are assumed to be available in excess. The model is based on several additional assumptions. For example, some tyrosine residues are treated as a single unit, i.e., lumped together as a virtual phosphorylation site.

Currently available methods for visualization of rule-based models

Models of biochemical processes, including rule-based models, are often easier to understand if they are visualized. Recently, efforts have been made to standardize visual representations of biochemical systems and models of biochemical systems. These efforts have culminated in

Systems Biology Graphical Notation (SBGN) (57). SBGN provides three sets of notational conventions, called languages, for various types of visualizations. Among these, the Process Description (PD) language can be used to visualize a biochemical reaction network or a model of such a network. Diagrams made using the PD language or the earlier related conventions of process diagrams (58) are available that illustrate fairly large reaction networks (59–61). For example, the diagram of Caron et al. (61) accounts for 964 species and 777 reactions. Unfortunately, this network is small compared to some of the reaction networks underlying rule-based models (14,62). Some rule-based models can be converted to conventional models and visualized using methods developed for such models, including SBGN, but there are many rule-based models that for all intents and purposes do not have conventional counterparts. It is especially for these cases that new visualization methods are needed. Below, we briefly review three visualization methods that have been used specifically for rule-based models and we discuss their limitations.

Graphical representation of individual rules

A rule can be visualized via the graphical conventions of Faeder et al. (48). These conventions are used in Fig. 1 to illustrate a rule in the FcεRI model (22,23) that characterizes binding of Lyn to the phosphorylated β chain of FcεRI. The conventions of process diagrams (58) may also be used to illustrate individual rules (63). The approach of Fig. 1 is only adequate for illustrating one rule or a few rules. Individually illustrating every rule in a large model (i.e., a model composed of a large number of rules) will result in a diagram that is locally comprehensible but globally incomprehensible. Thus, illustration of individual rules is impractical for communicating the content of a large model.

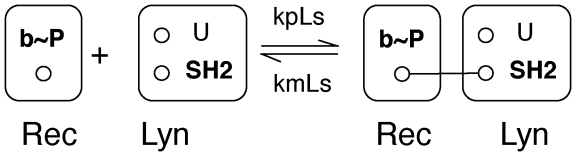


Figure A.2.1. Visualization of a rule. A molecule or part of a molecule is represented as a set of vertices, which represent molecular components. Vertices are labeled and may also be attributed. Here, the vertex b has the attribute P. Bonds are represented as edges. The reaction center defined by this rule is highlighted by bold vertex labels.

Contact maps

Danos et al. (62) introduced contact maps, which facilitate static analysis of rules (44,47). Contact maps are also useful for visualization purposes. (The term 'contact map' should not be confused with the term 'protein contact map,' which is used in structural biology64.) A contact map, which is a type of site graph, can be derived unambiguously from a rule-based model. A contact map identifies the molecules, the components of molecules, the possible internal states of components, and the possible bonds between components that are included in a model. Software tools are available for constructing a contact map automatically from a BNGL- or Kappa-encoded specification of a rule-based model50 (http://www.rulebase.org). A contact map for the FcεRI model (22,23) is shown in Fig. 2.

The contact map of Fig. 2 is derived directly from the FcεRI model (22,23). Thus, it reflects modeling assumptions, and fails to convey certain information about FcεRI signaling that was used in model specification. For example, the kinases responsible for phosphorylation events are not identified in Fig. 2. Typically, in a rule-based model, catalysts are not explicitly represented in rules, so contact maps generally will not reveal enzyme-substrate relationships. The graphical

representation of molecules in Fig. 2 conforms to the underlying graphical formalism of BNGL (49). In this formalism, only molecules and molecular components (i.e., only one layer of parent-child relationships) can be represented, even though molecular components can contain subcomponents. As a result, as discussed in detail by Lemons et al. (65), structural relationships among the functional components and subcomponents of signaling proteins can be obscured. Explicit representation of enzyme-substrate relationships and structural relationships are generally not necessary for simulation purposes (15), but omitting these types of details from an illustration of a model, such as that of Fig. 2, can hide the biological knowledge underlying a model specification.

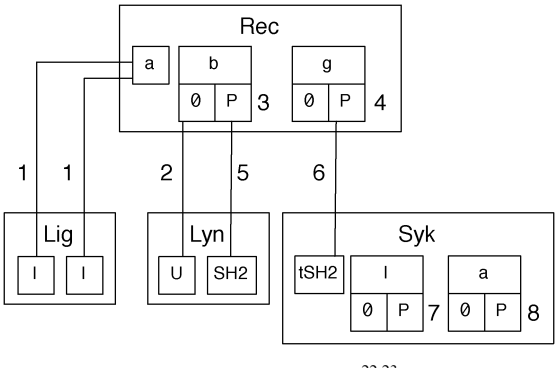


Figure A.2.2. A contact map for the FceRI model. A contact map shows the molecules, components, and direct binding interactions that are specified in a model. An outer box represents a molecule type. Inner boxes represent components. The possible internal states of compo- nents are also represented by boxes below component boxes. Binding partners are connected by lines. This contact map indicates, for example, that the g component of Rec has two possible states (0 and P), and when in the P state, g may bind the tSH2 component of Syk. This component represents the two tandem SH2 domains of Syk. The labels 1, 2, 5 and 6 refer to sets of rules in the FceRI model that characterize association/dissociation reactions. The labels 3, 4, 7 and 8 refer to sets of rules that characterize internal state change reactions, i.e., phosphorylation reactions in the FceRI model.

Molecular interaction maps

Kohn et al. (66) proposed conventions for representing a system marked by combinatorial complexity in the form of a molecular interaction map (MIM). Such a MIM can be used to visualize rules (14). It should be noted that MIM-like diagrams can be specified using the Entity Relationship (ER) language of SBGN (57). A MIM provides a visualization of a biological system by using boxes to represent molecules and a variety of symbols and lines/arrows to represent different types of interactions and influences. A MIM for the FcεRI model (22,23) is shown in Fig. 3. Annotation of this MIM is provided in Appendix S2. The main purpose of Appendix S2 is to explain our use of MIM notation, i.e., why we used MIM notations as we did in our attempt to provide a MIM that accurately reflects the FcεRI model (22,23). The conventions of a MIM call for the representation of a molecule only once so that all interactions involving a molecule can be traced to a common origin. This feature of a MIM, which is highly desirable as it avoids the need to represent every chemical species that can be populated, as in a conventional reaction scheme, is shared by a contact map. Unlike the situation for contact maps, software is not available for drawing MIMs automatically from model specifications. A MIM is a handcrafted illustration, although MIM construction is aided by a PathVisio (67) plugin (http://discover.nci.nih.gov/mim).

Interactions illustrated in a MIM by lines/arrows fall into two categories: direct interactions, or reactions, and contingencies, which characterize how interactions/reactions affect one another. In other words, a MIM depicts molecular interactions as well as the way in which interactions are affected by the context in which they take place. For example, the MIM of Fig. 3 shows that the SH2 domain of Lyn interacts with the phosphorylated β ITAM of FcεRI (see the arrow labeled '5,' which depicts a reaction), and it also shows that this interaction is mutually exclusive with

binding of the unique domain of Lyn to unphosphorylated β (see the pair of inhibition arrows between the arrows labeled ‘2’ and ‘5’).

The conventions of Kohn et al. (66) do not allow for the explicit representation of molecular substructures and site-specific details of molecular interactions. As shown in Fig. 3, boxes are used to represent molecules, and molecular components are represented using plain text inside molecule boxes. Components are not assigned their own boxes, and there is no provision for subcomponents. Thus, structural relationships can be difficult to visualize. For example, it is difficult to visually suggest that the ‘activation loop,’ identified as a site of phosphorylation in Fig. 3, is located within the PTK domain of Syk. Furthermore, interaction arrows and glyphs for post-translational modifications terminate at the edge of a molecule box, which makes it difficult to identify the components responsible for an interaction or the components affected by post-translational modifications. As demonstrated in Fig. 3, arrows can be positioned to suggest which components are responsible for an interaction, but nevertheless, with respect to representation of interactions at the level of molecular components, the conventions of Kohn et al. (66) are somewhat imprecise and less precise than the conventions of Danos et al.62 (cf. Figs. 2 and and3).3). On the other hand, a MIM provides a clearer picture of the enzymes responsible for post-translational modifications than a contact map (cf. Figs. 2 and and3).3). The conventions used to draw the MIM of Fig. 3 date back to 2006. An update of these conventions recently became available (http://discover.nci.nih.gov/mim), which allows for better representation of molecular components and site-specific details of molecular interactions (69). The updated conventions introduce ‘entity feature’ glyphs, which essentially allow boxes to be used to represent molecular components, not just whole molecules. These conventions differ from those

that we will recommend below and do not specifically address visualization or annotation of rule-based models.

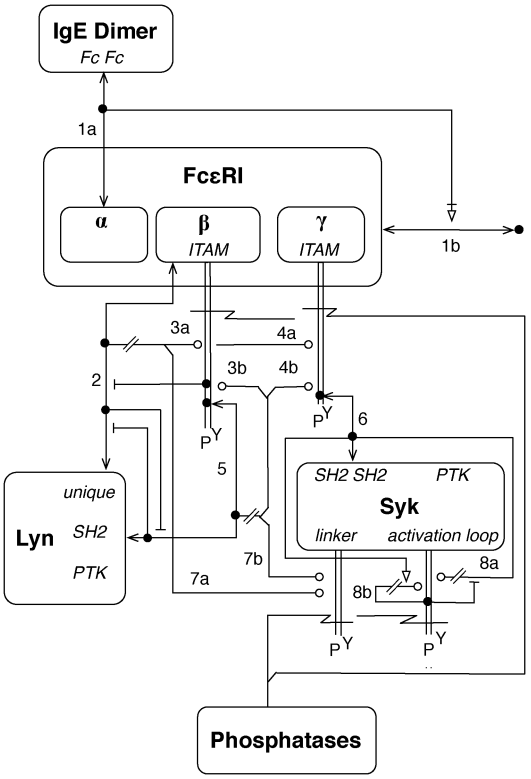


Figure A.2.3. A molecular interaction map for the FceRI model. A MIM illustrates molecules, reactions, and contingencies. A line that begins and ends with an arrowhead represents noncovalent binding, whereas a pair of parallel lines with no arrowheads represents covalent binding. An open circle indicates enzymatic catalysis, an open triangle indicates stimulation, and a jagged line indicates cleavage of a covalent bond. For additional information about MIM conventions, see Kohn et al (66,68). Note that this MIM is intended to be read using the combinatorial interpretation of MIM notations (66).

Results and Discussion

Having briefly reviewed the background material presented above, we are now prepared to introduce the concept of an extended contact map, which combines features of a plain, model-derived contact map (Fig. 2) with features of a MIM (Fig. 3). Our intention is to provide a means

to visualize site-specific details of molecular interactions in cell signaling systems as well as to provide a means to illustrate and annotate rule-based models, which typically account for such details.

One can view the conventions proposed here as a tuning of the established MIM and contact map conventions of Kohn et al. (66) and Danos et al. (62) to make these conventions more useful for visualization of (large) rule-based models, protein substructures and site-specific details of protein interactions. Our notations are largely consistent with MIM conventions, but there are differences. For example, we introduce nesting of boxes to better represent protein substructures, and we propose the linking of maps to rules, and vice versa. Importantly, because rules are powerful tools for concisely and precisely representing contextual constraints on molecular interactions, we deemphasize the visualization of contextual aspects of interactions.

Below, we first provide an overview of the basic principles of an extended contact map and we then present several example visualizations. These examples serve to elaborate the concept of an extended contact map and to illustrate how various cell signaling processes can be visualized within the framework of an extended contact map. Finally, we discuss the concept of a map guide, which can be associated with an extended contact map to document additional information about the molecules and molecular interactions visualized in the map, particularly the contextual dependencies of the interactions. A map guide can also be used to specify and annotate an executable rule-based model encompassing the molecules and interactions visualized in a map. The model specification may be partial or complete. If a guide serves to annotate a model, it can

be referred to as a model guide. The conventions presented here can be used to visualize and annotate an existing model or to depict a set of interactions before they are formalized as rules.

Basic features of an extended contact map

An extended contact map for the FcεRI model (22,23) is shown in Fig. 4. A guide for the map of Fig. 4 is excerpted in Fig. 10 and will be discussed later. The guide lists and annotates proteins and interactions included in the model. Arrows in the map are numbered to correspond to sections in the guide. Each section includes a summary of available knowledge about an interaction and a set of rules. The rules in a set are related, in that they share a common reaction center. In other words, the rules in a set describe the same interaction, but in different contexts. In general, if an interaction depicted in a map occurs in more than one contextual setting, then a rule can be provided for each contextual setting of interest.

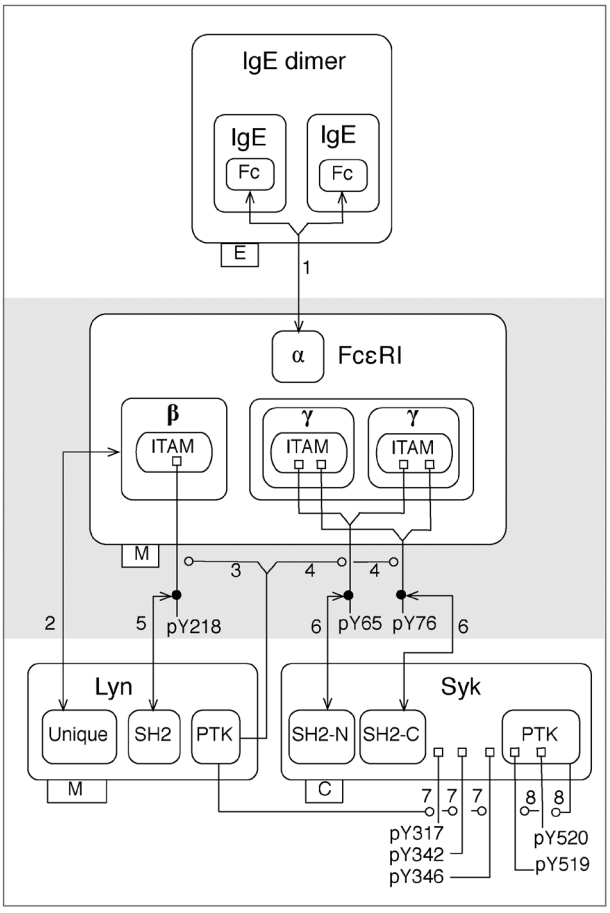


Figure A.2.4. An extended contact map for the FceRI model (22,23). Molecules are represented with nested boxes. A direct physical interaction is represented by an arrow that begins and ends with an arrowhead. An enzyme–substrate relationship is represented by an arrow that begins at an enzyme or catalytic subunit box and terminates with an open circle, which identifies the substrate. Tags attached to lower left corners of molecule boxes identify compartmental locations: E stands for extracellular, M stands for plasma membrane (the subunits of FceRI are transmembrane proteins and Lyn is anchored to the inner leaflet of the plasma membrane), and C stands for cytoplasmic (Syk is able to freely diffuse in the cytoplasm). Phosphatase activity is not depicted in this map, as phosphatases are considered only implicitly in the model (22,23). Note that Arrow 5 corresponds to the rule of Fig. 1. The labels 1–8 next to arrows refer to sections of the model guide. Each section includes a summary of available knowledge about an interaction and a set of rules that formally characterize the interaction. Note that the numbered arrows in this diagram correspond to the numbered lines/arrows in the diagrams of Fig. 2 and 3.

The map of Fig. 4 has three layers, which are indicated with shading. The concept of layers is based on the conventions of Kohn et al. (70) The top layer includes a depiction of an IgE dimer,

the receptor crosslinking reagent that initiates signaling in the FcεRI model (22,23). The second layer contains FcεRI, which is the only molecule in the model to interact with IgE. The third layer contains the kinases Lyn and Syk, which interact with FcεRI. In general, the idea is to organize molecules in a layout to reflect the causality of events in cell signaling. A molecule or set of molecules is chosen as the starting point of the signaling process and is placed at a certain location in a map (e.g., at the top), which defines the first layer. The second layer contains molecules that interact with the molecules in the first layer, the third layer contains molecules that interact with molecules of the second layer, and so on. This layout is not strictly a representation of causality or information flow, which is better represented with a path (25) or story (62). A path or story (i.e., a minimal path) can be used to guide the numbering of arrows and the layering of an extended contact map.

As can be seen in Fig. 4, nested boxes are used to represent molecules (all proteins in this example) and their component and subcomponent parts. These nested boxes correspond to hierarchical graphs. Lemons et al.65 have recently proposed conventions that allow such graphs to be used to annotate rule-based models. (Incidentally, these conventions are consistent with the related representational formalism of Yang et al. (71).) In Fig. 4, components of a protein are arranged linearly with the most N-terminal component at the left and the most C-terminal component at the right, a recommended convention consistent with many diagrammatic representations of proteins. The use of nested boxes allows for explicit representation of the structural relationships among the components and subcomponents of a molecule. For example, the β and γ chains of the receptor are shown to have multiple levels of internal structure: each contains an ITAM, which each contains a tyrosine residue that is a substrate of Lyn. We

generally recommend that a protein be depicted in a map only once. A complex can be depicted if the complex is treated as an indivisible unit in a model. In Fig. 4, the γ chain is depicted twice, because the two γ chains are covalently coupled to each other by disulfide bonds and are constituent components of a multimeric protein (FcεRI), which is treated as an indivisible molecular entity in the FcεRI model (22,23).

Two types of interactions are illustrated in the map of Fig. 4: direct physical interactions marked by reversible binding and enzyme-substrate interactions marked by covalent bond formation. A direct physical interaction is represented by a line that begins and ends with an arrowhead. The arrows labeled 1, 2, 5, and 6 in Fig. 4 represent direct physical interactions. For example, Arrow 2 indicates that the unique domain of Lyn interacts with the β chain of FcεRI. At the time that the FcεRI model (22,23) was originally formulated it was unclear how the unique domain of Lyn interacts with the β chain specifically. Accordingly, the arrow from the unique domain is terminated at the border of the β chain instead of extending further. An enzyme-substrate interaction that results in formation of a covalent bond is represented by an arrow that begins at an enzyme or catalytic domain box and terminates with an open circle at a modification flag, which identifies the modification (i.e., the covalent bond formed) and the substrate. The arrows labeled 3, 4, 7, and 8 in Fig. 4 represent enzyme-substrate interactions. For example, Arrow 3 indicates that Lyn catalyzes phosphorylation of tyrosine 218 in the β ITAM of FcεRI.

As can be seen in Fig. 4, flags are attached to molecule boxes to indicate sites of post-translational modifications. A flag represents a covalent bond between a protein and a functional group (e.g., phosphate) or a small protein, such as ubiquitin. We later demonstrate how a similar

notation can be used to represent covalent bonds in general. Post-translational modification flags have three parts: the 'base' of the flag is a small square that represents an amino acid residue in a polypeptide chain, the 'pole' is a line that represents a covalent bond, and the 'flag' itself is a text label. The text label of a modification flag (e.g., pY218) is used to identify the type of modification (e.g., 'p' represents phosphorylation) and the location of the modification (e.g., the single-letter amino acid code and number of a residue within a polypeptide chain). If a direct physical interaction depends on a post-translational modification, the arrow representing this interaction may originate/terminate at a modification flag, where a solid dot is placed as a point of origin/termination, in accordance with the conventions of Kohn et al. (68). For example, the SH2 domain of Lyn interacts with phosphorylated tyrosine 218 in the β chain of FcεRI; Arrow 5 connects the SH2 domain box of Lyn to a dot on the pY218 modification flag. If an unmodified amino acid must be represented, it is simply drawn as a component, i.e., absence of a modification flag indicates absence of modification. If modification of an amino acid residue inhibits, rather than enables, an interaction, an inhibition arrow originating from a dot on the flag for this modification and terminating at the appropriate interaction arrow may be used to represent the negative effect of the modification on interaction. Flags in maps will tend to correspond to internal states of components of proteins included in a model, and flags will tend to be connected to arrows representing rules that define internal state changes.

It may be useful to point out how Fig. 4 differs from Figs. 2 and and 3. Figure 4 contains information not shown in Fig. 2. This missing information in Fig. 2 is information that cannot be directly derived from the BNGL-encoded specification of the FcεRI model (22,23). As mentioned above, explicit representation of catalysts is usually missing in BNGL-encoded rules,

and this model is no exception. Thus, enzyme-substrate relationships are not revealed in Fig. 2, whereas such relationships are revealed in Fig. 4. This is one reason why we refer to Fig. 4 as an extended contact map. Another example of information provided in Fig. 4 beyond that provided in Fig. 2 is identification of the individual sites of phosphorylation within the linker region and activation loop of the PTK domain of Syk. When an extended contact map is used to illustrate a model, we recommend that the map illustrate the biological knowledge underlying the model specification, i.e., the information available to the modeler and considered in model formulation. Comparison of an extended contact map and the corresponding model-derived contact map can then reveal how biological knowledge of a cell signaling system has been translated into a formal specification of a model for the system.

Visually, some of the differences between Fig. 3 (a MIM) and Fig. 4 (an extended contact map) may seem superficial. However, Fig. 4 introduces conventions that are essential for the consideration of molecular substructure and site-specific details of molecular interactions, most prominently nested boxes for the representation of structural relationships. Another key difference is that Fig. 3 contains information about the contextual dependencies of molecular interactions that is not represented in Fig. 4. For example, binding of an IgE dimer to the α chain of FcεRI is indicated to be a prerequisite for receptor dimerization in Fig. 3, but not in Fig. 4. In fact, Fig. 4 does not explicitly show that FcεRI dimerizes, although this can be inferred. Another example of context depicted in Fig. 3 but not in Fig. 4 is the case of the rightmost phosphorylation glyph attached to Syk. The arrows terminating and originating at this glyph are intended to indicate that Syk trans-phosphorylates a second copy of Syk in a dimeric receptor complex and that the rate of phosphorylation is enhanced when the first copy of Syk is

phosphorylated in its activation loop. A MIM tends to emphasize the contextual constraints on interactions rather than the component parts of molecules responsible for interactions. The opposite holds true for an extended contact map. We recommend a minimal representation of contextual information in an extended contact map because it is difficult to represent this type of information in the form of a diagram without sacrificing precision and/or readability. Thus, for example, avidity effects such as those considered in the model of Barua et al. (27) would not be depicted in a map. In our experience, visualization of contextual dependencies tends to result in an overloaded diagram, especially in the case of large models. Our position is that a rule is usually the best way of capturing the contextual dependencies of an interaction. Therefore, we suggest that interaction arrows in an extended contact map be cross-referenced to a list of rules. As noted above, the interaction arrows of Fig. 4 are labeled 1–8 and these labels correspond to sections of an associated guide (see Fig. 10 for an excerpt) where rules representing the interactions are listed and annotated.

A MIM can serve as a stand-alone summary of available biological knowledge. An extended contact map can also serve the same purpose. However, we recommend that a map always be accompanied by a guide containing rules for interactions. The guide need not fully specify a model. For example, a guide containing rules but omitting rate laws for rules, which are required for simulations, can still be useful, because rules are suitable for providing details that are not easily captured in a map. A MIM can be supplemented with annotation (for example, see Kohn et al.70). What is different here is that we are proposing that the annotation associated with an extended contact map include formal elements of an executable rule-based model, especially rules. It should be noted that rules, because they are formal representations of interactions, are

more easily associated with arrows in a map, which also are representations of interactions, than the formal elements of a conventional model. In an ODE-based model, for example, multiple terms in multiple equations are typically required to capture the effects of a single interaction(14,62). For an example of a MIM for which a corresponding conventional model is available, see Kim et al. (72).

We have now introduced the basic features of an extended contact map by way of example. Below, we give additional guidance about the representation of molecules and molecular interactions before introducing several additional simple examples, which illustrate cell signaling processes and types of molecules that are not included in Fig. 4 but that are commonly found in cell signaling systems.

General guidelines for representation of molecules

As described above, proteins in an extended contact map are represented with nested boxes that correspond to hierarchical graphs, and sites of post-translational modifications are marked with modification flags. Recommended box and flag glyphs are summarized in Fig. 5. The components of a protein are ordered from N-terminal to C-terminal. When this type of ordering is not possible, as with separate polypeptide chains in a multimeric protein, individual polypeptides may be arranged in a way that reflects their physical organization. For example, in the case of a multimeric cell-surface receptor (e.g., FcεRI), a mostly extracellular subunit (e.g., FcεRIα) may be placed above other mostly cytoplasmic subunits (FcεRIβ and γ2).

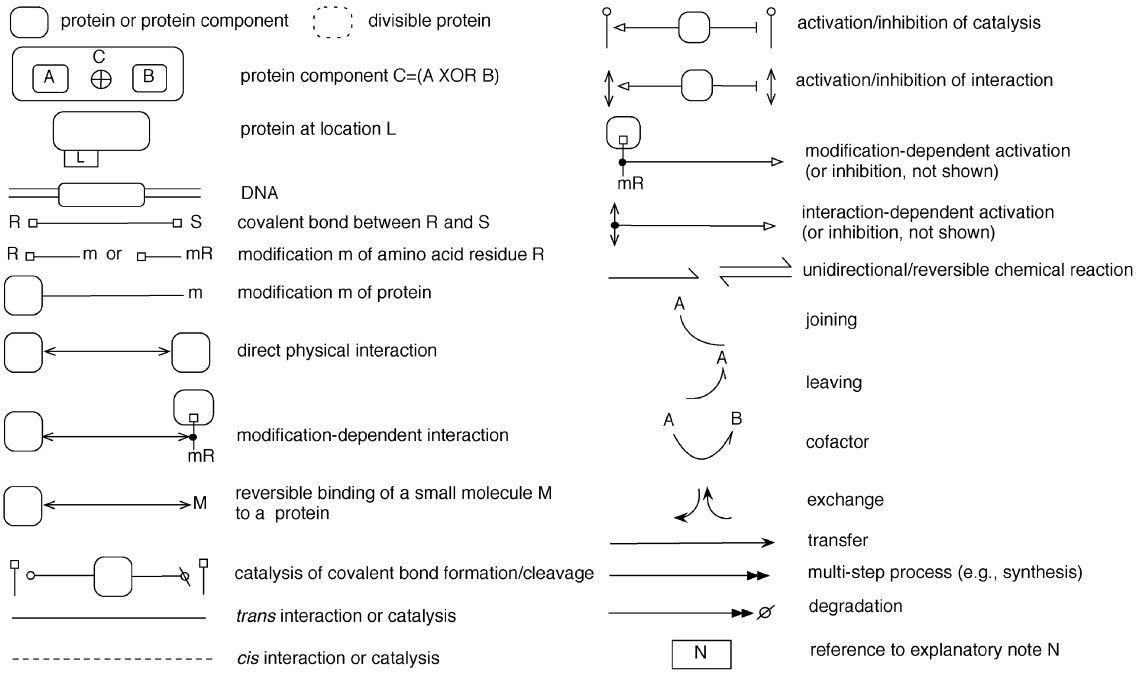


Figure A.2.5. Boxes, flags, arrows and other symbols that can be used to draw an extended contact map. All of these glyphs are used in example diagrams that follow; see the main text for discussion. The label 'm' represents a post-translational modification. A separator, such as ':', can be inserted between the two parts of a modification flag label. Thus, 'mR' and 'm:R' are both acceptable forms of a modification flag label. For post- translational modifications commonly involved in cell signaling,73 we recommend the following labels: Ac for acetylation, Me for methylation, OH for hydroxylation, p or P for phosphorylation, and Ub for ubiquitination. The 'joining,' 'leaving,' and 'cofactor' arrows are intended for use in combination with chemical reaction arrows. Note that the joining and leaving arrows together equal a cofactor arrow. An OmniGraffle stencil of these symbols is available at the BioNetGen web site (http://bionetgen.org).

To maintain compactness of a diagram, we recommend that only components of interest (e.g., domains, motifs, and amino acid residues that are included in a model) be shown in a map. For a map illustrating a BNGL-encoded rule-based model, the representation of molecules should reflect the BNGL molecule type definitions (15) of the model. A more complete annotation of known molecular substructure can be included in a map/model guide if desired. In addition, a molecule is generally only shown once in an extended contact map, with the exception of molecules that are represented using plain text (see below) and molecules that are present in

multiple copies in a complex (e.g., the γ chains of FcεRI). To avoid redundancy in the depiction of post-translational modifications, we recommend that the line segments (i.e., the poles) of modification flags attached to repeating component boxes in a map be consolidated so that they emerge from a molecule box as a single line. An example of this practice is shown in Fig. 4; see the pY65 and pY76 flags.

In addition to representing protein substructure, an extended contact map can provide other information about a protein, namely its location(s) and products of proteolytic cleavage. To indicate the possible compartmental locations of a protein, one can attach a compartment tab to a molecule box. Labels within the tab represent different compartments. A label need not be included for a compartmental location that can be inferred. For example, the compartment tab of the Syk molecule box in Fig. 4 contains the label 'C' (cytosolic) but not 'M' (plasma membrane). This is because membrane association of Syk can be inferred by the association of Syk with FcεRI, a membrane protein. If a model includes rules for translocation of proteins, such as the rule of Eq. (2), a tab can be associated with multiple labels to indicate all of the compartments in which a protein can be found, and the compartment tab can also be associated with a set of translocation rules, which can be listed and annotated in a model guide. To indicate that a protein is divisible (i.e., cleaved by the action of a protease into two or more smaller proteins), one can use a dotted molecule box to represent the protein. This also applies to the representation of divisible components. However, a dotted box should only be used when the protein fragments that result from proteolytic cleavage are relevant for understanding the system depicted in a map. One would not use a dotted molecule box to simply indicate that a protein is degraded.

Here, we emphasize visualization of proteins, but an extended contact map can also include other types of macromolecules, such as DNA, as well as small-molecule compounds, such as lipids, drugs, and metabolites. We recommend that boxes be reserved for macromolecules and we recommend that small-molecule compounds be represented using plain text.

General guidelines for representation of molecular interactions

Interactions among molecules are visualized with arrows in an extended contact map (Fig. 5). The same set of interactions can also generally be represented with rules, and thus an arrow in a map can be linked to one or more rules in a model. This connection is made through a model guide: arrows in a map are numbered, and rules and sections in a model guide are numbered to correspond with arrows. An arrow may correspond to more than one rule if a set of rules share a reaction center. A reaction center is defined as the set of vertices (components) that undergo modification in a graph-rewriting operation defined by a rule (15). When a reaction center is common to multiple distinct rules, it means that the rules are representing a common interaction that takes place in multiple contexts. Rules that share a common reaction center can be mapped to a single interaction arrow in an extended contact map and the contextual differences need not be captured in the map, as these differences are accounted for in the rules themselves.

It is important to note that arrows are drawn as specifically as possible; in other words, they extend as many layers into the molecule as available knowledge allows, but not further. If an exact binding site is not known, an arrow is terminated at an outer layer and may even terminate at the outermost border of a molecule box. To accommodate space limitations in a map, arrows

may branch. As seen in Fig. 4, a catalysis arrow from Lyn branches to show phosphorylation of the β and γ chains. When an arrow branches, a short diagonal segment or pair of diagonal segments can be introduced, which helps identify the box from which the arrow originates (see Arrows 1, 3 and 4 in Fig. 4). A catalytic arrow can be broken and extended to point to multiple modifications flags (see Arrows 4, 7 and 8 in Fig. 4). If an arrow crosses a modification flag that it does not affect, it may be drawn continuously or broken into segments; breaking of a line into segments is a stylistic option that does not affect the meaning of an arrow. Recommended arrows are summarized in Fig. 5. Unless otherwise noted, all arrows drawn with solid lines should be assumed to depict trans interactions; cis interactions are depicted with dotted lines. This convention can be reversed if convenient, e.g., in a case where most arrows in a map represent cis interactions. A reversal of the convention should be duly noted.

Example visualizations of common cell signaling processes

We now demonstrate how the conventions described above can be used to represent various biochemical processes found in cell signaling systems (Figs. 6–8).

Protein synthesis and interaction of a transcription factor with a DNA binding site

According to the central dogma of molecular biology, protein synthesis consists of two basic steps: transcription of DNA into mRNA, and translation of mRNA into a polypeptide76. These steps may be regulated in many ways and additional steps may be involved in de novo protein synthesis; however, we are often only interested in the relationship between a gene and its protein product. In this case, one can use a shorthand notation to indicate synthesis of a protein

encoded by a gene (Fig. 6A). A double-headed arrow points from a molecule box for a gene to a molecule box for a protein to represent the multistep process of transcription/translation. The double arrowhead is intended to suggest that steps are not shown. DNA is represented as a pair of parallel lines, and boxes for genes, promoters and other regulatory elements are embedded within these lines. This example also shows binding of a transcription factor (TF) to DNA and indicates that this interaction stimulates transcription/translation. A solid dot placed on the DNA-TF interaction arrow serves as a point of origin for an activation arrow. In general, a dot is placed on an arrow when it is necessary for another arrow to begin or end at that point. A similar combination of symbols could be used to represent other synthetic processes.

Proteolysis and protein degradation

Cells routinely degrade proteins: unnecessary or misfolded proteins are dismantled, and protein degradation is used to regulate the rates of biochemical reactions. Much protein degradation takes place in proteasomes (76). In an extended contact map, degradation can be simply depicted as a double-headed arrow pointing from the degraded protein to a 'null' symbol (Fig. 6B). Proteases catalyze cleavage of peptide bonds between amino acids. This process has a role in protein degradation as well as in regulation of enzymatic activity. For example, caspase signaling involves caspase-catalyzed cleavage of caspase proteins, which liberates enzymatic subunits to assemble into active caspase enzymes (77). The uncleaved form of a protein may be represented with a dotted border, indicating that it is divisible (Fig. 6C). The proteins that result from the cleavage event are represented within this box. They are connected by a solid line with squares at either end, representing a covalent bond. A 'no' arrowhead points from the catalytic domain of a protease to the covalent bond, indicating that the bond is cleaved. A more elaborate example of

representation of a proteolytic cascade is provided in Fig. S4, which depicts proteolytic cleavage of complement component C3 to C3d (78–80). This figure illustrates how a proteolytic cascade that results in cleavage of a protein at multiple sites can be represented in an extended contact map.

Allosteric regulation of a metabolic reaction

Allosteric regulation occurs when an effector molecule alters an enzyme's activity by binding to a site on the enzyme that is distinct from the active site. The result may be either an increase or decrease in catalytic activity. An example of an enzyme controlled by allosteric regulation is phosphofructokinase-1 (PFK-1). This enzyme catalyzes a key, irreversible step in the glycolysis pathway, and it is a central point of regulation. For example, PFK-1 is positively regulated by fructose-2,6-bisphosphate (76). In an extended contact map, allosteric regulation of enzymatic activity by a small-molecule effector is represented as follows. A direct physical interaction arrow is drawn between the enzyme and effector. An activation or inhibition arrow then originates from the interaction arrow and points to the catalysis arrow between the enzyme and substrate (Fig. 6D). We generally discourage the use of activation and inhibition arrows because they tend to be ambiguous, but they are useful for representing allosteric regulation. In this example, plain text is used to represent metabolites, rather than boxes, to make a distinction between small molecules and macromolecules. If a material component considered in a model is not treated as a structured object (i.e., a graph) in a model, it and the reactions in which it participates can be represented using conventional means for representing biochemical reaction networks.

Dephosphorylation

Representation of phosphorylation is demonstrated in Fig. 4. The reverse process, dephosphorylation, is the enzyme-catalyzed removal of a phosphate group from an amino acid residue. Dephosphorylation can be just as important as phosphorylation in regulating protein interactions and catalytic activities. Unregulated basal dephosphorylation by unspecified phosphatases can be omitted from an extended contact map, as in Fig. 4, because it would necessitate an additional arrow for every phosphorylated residue, making the map less readable. However, it is sometimes significant that a specific phosphatase acts on a specific substrate. For example, dephosphorylation of the C-terminal regulatory tyrosine in the kinase Lck by SHP-1 prevents the formation of an intramolecular bond, which regulates Lck kinase activity (81). As in the MIM of Fig. 3, Kohn and co-workers use a jagged line to represent dephosphorylation (68). As an alternative that is more compact and more consistent with our notation for catalysis of covalent bond formation, we suggest depicting dephosphorylation (and more generally cleavage of a covalent bond) with a 'no' symbol (Fig. 6F). In the case of lipids (e.g., dephosphorylation of phosphatidylinositol (3,4,5)-trisphosphate by PTEN (82)), dephosphorylation can be represented as a standard chemical reaction with a catalysis arrow pointing from the enzyme to the reaction (Fig. 6E). We also use this example to demonstrate an interaction between a lipid and a protein: PIP3 binds the pleckstrin homology (PH) domain of PDK1, recruiting PDK1 to the plasma membrane (83).

Transport

An extended contact map does not aim to illustrate transport or trafficking between compartments, but a map can be used to indicate compartmental locations of molecules.

Compartments and transport between compartments can be represented explicitly using cBNGL (51). The names of the compartments in which a molecule can be found can be included in an extended contact map in the form of a tag attached to a molecule box. In Fig. 6G, two location labels, 'Cyt' and 'Nuc,' are included within a single location tab attached to a molecule box for NF-κB. The tag indicates that NF-κB is considered to have two possible compartmental locations. A location tag can be associated with a rule, such as the rule of Eq. (2), to clarify details about trafficking between compartments. In the case of Fig. 4, molecules are considered that are found in three compartmental locations, and all the molecules are represented in the same map. In more complicated cases, it may be convenient to draw separate maps for separate compartments. Note that compartmental locations that can be inferred from interactions need not be included in a map. For example, the location tag attached to the Syk molecule box in Fig. 4 only indicates that Syk is cytoplasmic. It can be inferred that Syk is membrane associated when it interacts with FcεRI, so a membrane location label is not included in the Syk location tag.

Association

The extended contact map of Fig. 4 demonstrates how direct physical interactions between protein binding partners (see Arrows 1 and 2) and phosphorylation-dependent interactions (see Arrows 5 and 6) can be represented. Interactions that depend on other types of post-translational modifications can be represented in the same way as a phosphorylation-dependent interaction. A direct physical interaction between a protein and DNA can be represented as shown in Fig. 6A. A direct physical interaction between a protein and a small molecule can be represented as shown in panels D and E of Fig. 6. If two proteins are associated indirectly via an unknown linker, the boxes representing the proteins can be connected via a direct physical interaction

arrow and the arrow can be attached to a note tag, a rectangle enclosing a reference to a note of explanation.

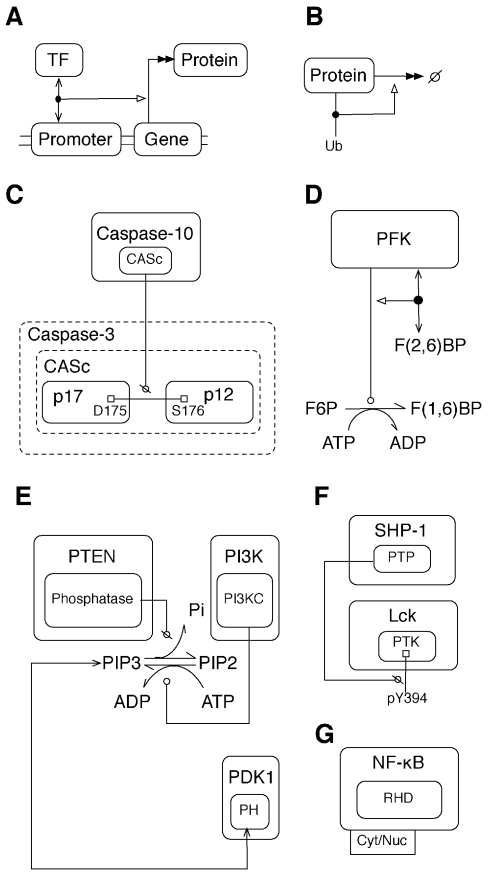


Figure A.2.6. Visualization of various cell signaling processes. (A) Stimula- tion of transcription and translation by transcription factor (TF) binding to DNA. (B) Stimulation of protein degradation by ubiquitination. (C) Proteolysis. Dotted box lines identify the parts of the caspase-3 polypeptide chain affected by the proteolytic action of caspase-10. Cleavage of the indicated peptide bond breaks the chain itself and the CASc segment. (D) Allosteric regulation of a metabolic reaction, conversion of F6P (fructose 6-phosphate) to F(1,6)BP (fructose 1,6-bisphosphate). The effector is F(2,6)BP (fructose 2,6-bisphosphate). (E) Lipid phosphorylation and dephosphorylation, and PH domain interaction with phosphatidylinositol (3,4,5)-trisphos- phate (PIP3). The small-molecule metabolites ATP and ADP and inorganic phosphate (Pi) are shown to participate in the reactions between phosphatidylinositol (4,5)-bisphosphate (PIP2) and PIP3 to illustrate how such details can be represented in a map if desired. However, such details would normally be omitted for simplicity. (F) Protein dephosphorylation. (G) A protein (NF-kB) that traffics between the cytoplasm (Cyt) and nucleus (Nuc).

Conjugation and transfer: ubiquitin and ubiquitin-like proteins

Ubiquitin is a small protein that may be covalently coupled to copies of itself and to other proteins. Ubiquitination (Ub) tags proteins for degradation and serves various other functions (84).

Representation of ubiquitination can be similar to representation of phosphorylation: a catalysis arrow can point from an enzyme to a substrate, where the type of modification ('Ub' for ubiquitination) and the location of the modification are specified. However, unlike phosphorylation, multiple enzymes are involved in the ubiquitination process: an E1 activating enzyme, an E2 conjugating enzyme, and an E3 ligase. Ubiquitin is bound to a cysteine residue in the active site of E1, transferred to the active site of E2, and then bound to the target substrate in a reaction catalyzed by E3 (86). Representation of ubiquitination in an extended contact map may vary. A detailed representation of ubiquitination includes all three enzymes and the target substrate. Arrowheads representing catalysis of covalent bond formation and cleavage can be used to implicitly represent transfer of Ub from one protein to the next (Fig. 7A). These reactions result in transfer of Ub, which can be alternatively represented with a transfer arrow, as depicted in Fig. 7B. Note that the arrowheads used for binding and transfer arrows are similar but distinct. See Fig. 5. Further note that Figs. 7A and 7B need not represent different models; the two diagrams could represent the same set of rules. In Fig. 7A, the dotted arrow from E1 indicates that an E1 enzyme removes ubiquitin from itself, rather than from a second E1 molecule. A more specific representation of ubiquitination in the style of Fig. 7A is shown in Fig. 7C. In some cases, specific residues in ubiquitin or ubiquitin-like proteins may be of interest. In the example of Fig. 7D, a specific glycine residue in the ubiquitin-like protein Atg12 is shown to form

covalent bonds with specific residues in Atg7, Atg5, and Atg10 (87). In addition, activation arrows point from catalytic arrows to transfer arrows, which represent the sequential transfer of Atg12 from Atg7 to Atg10 to Atg5. The activation arrows, which emerge from dots on the catalytic arrows, are intended to indicate that enzyme-catalyzed cleavage and formation of the indicated covalent bonds serve to transfer Atg12. Dashed borders for the molecules containing Atg12 indicate that these entities are divisible. Note that Fig. 7D illustrates how the styles of Figs. 7A and 7B can be combined. Lastly, it is worth mentioning that monoubuiqitination can be distinguished from polyubiquitination (i.e., formation of a ubiquitin chain86) in the label of a modification flag. For example, the label 'UbnK' can be used to represent a chain of n ubiquitin molecules.

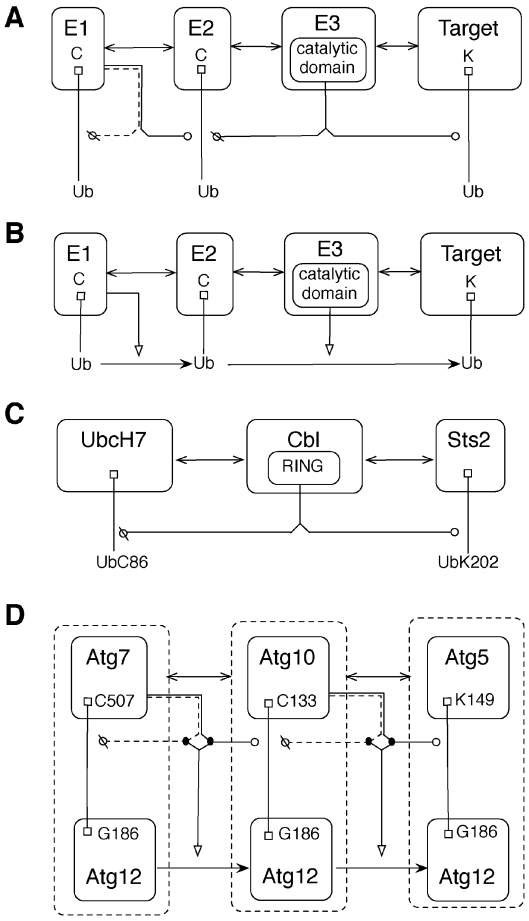


Figure A.2.7. Visualization of ubiquitin and ubiquitin-like protein conjugation and transfer reactions. (A) A general representation of ubiquitination that includes E1, E2, and E3 enzymes and a target substrate. E1 catalyzes cleavage of ubiquitin from itself and formation of a covalent bond between ubiquitin and E2. E3 catalyzes cleavage of ubiquitin from E2 and formation of a covalent bond between ubiquitin and the target. (B) In an alternative representation, ubiquitination is represented by stimulated transfer of ubiquitin from one molecule to the next. (C) A specific representation of ubiquitination showing only an E2 (UbcH7), an E3 (Cbl), and a target (Sts2), which is ubiquitinated at K202 (85). (D) A specific glycine residue in the ubiquitin-like protein Atg12 is shown to form covalent bonds with Atg7, Atg5, and Atg10. Activation arrows point from catalytic arrows to transfer arrows. Dashed borders surrounding molecules containing Atg12 indicate that these entities are divisible.

Exchange: Ras

GTPases in the Ras family of proteins are hydrolase enzymes that bind and act on guanosine triphosphate (GTP) to yield guanosine diphosphate (GDP). In cell signaling, GTPases function as

switches, being ‘on’ when bound to GTP (i.e., able to bind an effector) and ‘off’ when bound to GDP (i.e., unable to bind an effector). Transitions between these two states are mediated by GTPase activating proteins (GAPs), which stimulate a GTPase's intrinsic catalytic activity thereby accelerating the rate at which GTP is converted to GDP, and guanine nucleotide exchange factors (GEFs), which facilitate exchange of GDP for GTP by loosening the binding of a GTPase to both GTP and GDP. GTP is at a higher concentration than GDP in cells and is more likely to bind an empty binding site. HRas is a GTPase that is acted upon by p120RasGAP, a GAP, and by Sos1, a GEF88. In Fig. 8, HRas is drawn with a branched interaction arrow pointing to GTP and GDP. A unidirectional chemical reaction arrow from GTP to GDP represents the conversion of GTP to GDP. A cis (dashed) catalytic arrow from HRas to the reaction arrow indicates that HRas catalyzes the cleavage of a covalent bond and converts GTP to GDP. Exchange of GDP for GTP is represented with a special exchange glyph consisting of a pair of bent arrows. An activation arrow from the p120RasGAP-HRas interaction arrow indicates that RasGAP stimulates GTPase activity. An activation arrow from the HRas-Sos1 interaction arrow pointing to the exchange glyph indicates that Sos1 stimulates GTP/GDP exchange. As depicted in Fig. 8, interaction between HRas and the REM domain of Sos1 allosterically activates GEF activity (89). The HRas molecule that allosterically activates Sos1 is distinct from the HRas molecule affected by the GEF activity of Sos1, and GDP- and GTP-loaded HRas have different allosteric effects, but these distinctions are not made in an extended contact map. Instead, rules in an associated model guide would clarify the mechanism depicted in the map. As depicted in Fig. 8, the GTP-bound form of HRas is able to bind Raf-1 (90). The dependence of this interaction on GTP loading is indicated by the activation arrow extending from a solid dot on the GTP-HRas interaction arrow to the interaction arrow between HRas and Raf-1. The diagram

of Fig. 8 contains a number of activation arrows. As mentioned earlier, we generally discourage the use of activation and inhibition arrows, but Fig. 8 provides an example of where these arrows are useful for representing allosteric regulation.

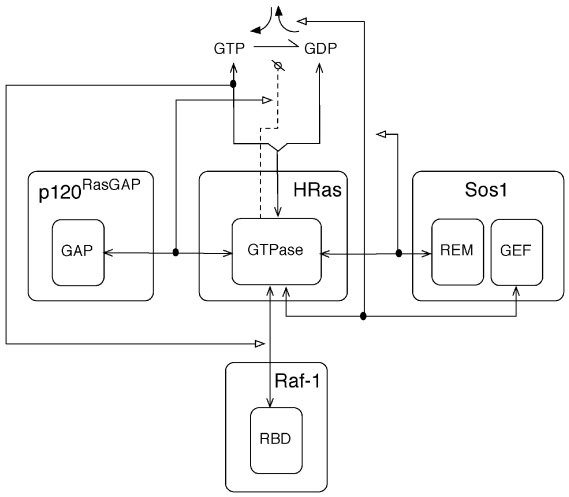


Figure A.2.8. Visualization of Ras regulation. HRas, which has GTPase activity, is drawn with branched arrows pointing to GTP and GDP to indicate that HRas contains a single binding site for the guanine nucleotides. The reaction arrow between GTP and GDP represents the transition from one bound state of GTPase to another. A catalysis arrow is drawn from the GTPase domain of HRas to the GTP-to-GDP reaction arrow to represent the intrinsic catalytic activity of HRas. The arrow is dotted to indicate that it represents a cis interaction (i.e., the GTPase acts on a GTP molecule that is bound to itself). A stimulation arrow is drawn from the GTPase–GAP interaction arrow to the catalysis arrow to indicate GAP-mediated upregulation of GTPase activity. Exchange of GDP for GTP is represented with a pair of bent arrows, and an activation arrow indicates that Sos1 stimulates exchange. Sos1 is allosterically activated by HRas binding to the REM domain.

Example visualizations of miscellaneous molecule types

We will now demonstrate how various molecule types not yet considered may be represented (Fig. 9).

Divisible proteins

All proteins are divisible, i.e., their peptide bonds may be cleaved. However, in some models it is relevant to track the cleavage of a particular protein. In such cases, a special notation for divisible proteins is useful. A protein that may be cleaved is represented with a dotted molecule box, which encloses the fragments that result from cleavage. A divisible protein, caspase-3, is visualized in Fig. 6C. Caspase-3 is cleaved by the action of caspase-10, which allows the p17 and p12 components of the CASc domain of caspase-3 to assemble into an active caspase77. A representation of complement component C3 is given in Supplemental Figure S4 (ESI).

Alternate subunits: APC/C

Many enzymes are multimeric proteins. An example is APC/C, a cullin-RING domain E3 ubiquitin ligase, the specificity of which is determined by a regulatory subunit. The regulatory subunit can be either Cdh1 or Cdc20 (91). In Fig. 9A, a component box is introduced for a regulatory subunit in which the two possible components are included, separated by an XOR symbol, indicating that only one may be associated with core APC/C at a time.

Sites of multiple modifications: Histone H3

Histone modification regulates chromatin structure. As depicted in Fig. 9B, lysine 9 in histone H3 may be modified in two possible ways, by acetylation and by methylation. The balance between these two modifications may influence gene regulation over the course of the cell cycle (92).

Homodimer: EGFR

Binding of epidermal growth factor (EGF) to the EGF receptor (EGFR) leads to formation of EGFR dimers. As depicted in Fig. 9C, receptors dimerize via ectodomain interactions (93). Note that the arrow in Fig. 9C represent a trans interaction.

Overlapping linear motifs: CD3ε

The CD3ε chain of the T-cell receptor (TCR) contains a proline-rich sequence (PRS) and an ITAM that overlap. In the region of overlap there is a tyrosine residue (Y188), which is a substrate of kinases and phosphatases. As part of the ITAM, Y188 is phosphorylated during TCR signaling. Phosphorylation of Y188 inhibits binding of the PRS to SH3 domains in interaction partners, and binding of the PRS inhibits phosphorylation of Y18894. Thus, it is relevant to show that the PRS and ITAM overlap. In Fig. 9D, the PRS and ITAM are represented as overlapping boxes with Y188 located in the overlapping region. The two component boxes can be distinguished by using box lines that differ in shading (as shown) or color. In complicated cases, it may be necessary to explain overlaps in a note or map/model guide.

Discontinuous binding sites: biotin and streptavidin

Binding sites may be composed of parts of distinct components of a protein or protein complex, and there are various possibilities for how such binding sites and their interactions can be represented in an extended contact map. For example, the four biotin binding sites in a streptavidin tetramer are formed by residues of adjacent monomers that interact as functional dimers (95). In Fig. 9E, the interaction of biotin with a streptavidin monomer is shown to be activated by a neighboring monomer. This diagram can be considered nonstandard. In such a case, a reference to an explanatory note can be included in a diagram. Here, ‘N’ is a label that

refers to the explanatory note ‘adjacent monomers form biotin binding sites.’ In general, a rectangle enclosing a label can be introduced to clarify aspects of map by providing a reference to a note of explanation.

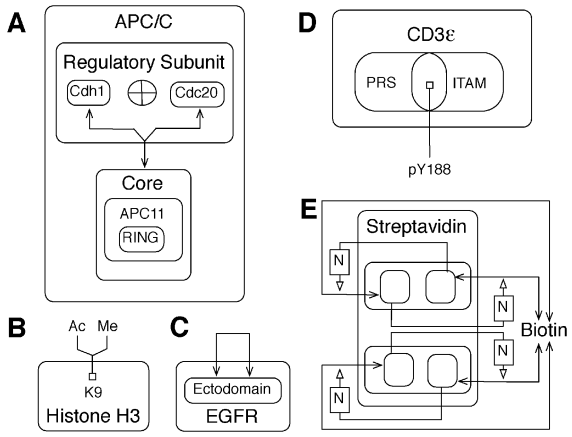


Figure A.2.9. Visualization of various molecule types. (A) A multimeric protein with two possible regulatory subunits. The two possible regulatory subunits of APC/C, Cdh1 and Cdc20, are shown as boxes within a ‘Regulatory Subunit’ box, separated with an ‘XOR’ (exclusive or) symbol to indicate that only one of these proteins may associate with the core at a time. (B) A protein containing a site that can be modified in multiple ways. (C) A protein that dimerizes. (D) A protein with overlapping linear motifs. Boxes of overlapping components are distinguished by different shades (as shown) or colors. The PRS box is gray and the ITAM box is black. (E) Discontinuous biotin binding sites of streptavidin. The use of activation arrows in this diagram can be considered nonstandard, so a note box is included to provide a note of explanation. The letter ‘N’ serves to label an explanatory note, ‘adjacent monomers form biotin binding sites.’ In general, we recommend that explanatory notes be labeled with letters, although longer labels are acceptable. Note that the monomeric subunits of streptavidin, a dimer of dimers, are not labeled, nor are the constituent dimers.

Basic features of a map/model guide

An extended contact map can be associated with a map guide or a model guide. A map guide complements an extended contact map by providing annotation about molecules and interactions

visualized in a map. A model guide goes beyond a map guide by attaching formal elements of a rule-based model, molecule type definitions and rules, to boxes and arrows. We recommend that a model guide be organized so that sections in the guide correspond to blocks of a BioNetGen input file (15). A model guide essentially serves as a specification of a rule-based model, although the specification need not be complete. It can serve to annotate not only an extended contact map but also the underlying model illustrated by the map. We recommend that rules in a model guide be specified using BNGL15 because of the availability of various BNGL-compatible software tools (11–13,15–17,41,42,45,50). However, any language for specifying rule-based models could be used.

A guide may contain representations of molecules in the form of BNGL molecule type definitions (15,49). A molecule type definition includes a list of internal states for all components that have internal states, as well as locations for components if one is using cBNGL51. A guide may also contain additional information that is not included in an extended contact map, such as links to online resources (e.g., UniProt (96), Pfam (97), and Phopho.ELM (4)), a narrative summary of available information about a protein, and estimates of protein copy numbers. A guide can include diagrams of complete domain structures of proteins in the form of domain graphs (98) and/or diagrams that define the compartmental locations of molecules. Such diagrams can be included in a guide to provide a more complete picture of individual proteins. As discussed previously, only components of interest are included in protein representations in an extended contact map; a protein may contain other elements, but depicting all of them in a map is discouraged, in part because the practice would tend to make maps difficult to read. In the case of a map used to illustrate a model, protein representations should reflect the components

considered in the formulation of the model. Consistent with the conventions of Kohn et al. (66), a modification flag in a map only indicates the modified state of an amino acid residue, even though a residue may also have an unmodified state. An unmodified state may be specified in a guide if desired. In rule-based models, post-translational modifications are often represented using internal states, which are simply variable attributes associated with vertices of graphs. The value of an attribute associated with a particular modification state is arbitrary. Thus, it can be useful to specify a mapping of modification states of an amino acid residue (including an unmodified state) to the values of the corresponding internal state attribute in a model. Figure 10A shows annotation for Syk. In Fig. 10B, a diagram of Syk is shown with embedded annotation for the molecule and individual components (e.g., the SH2 domains of Syk are identified as protein interaction domains), possible internal states ('0' for unmodified and 'P' for phosphorylated), and compartmental location ('cytoplasmic').

A map guide also serves to annotate the interactions represented by rules. Each interaction arrow in an extended contact map corresponds to either a rule or a set of rules in which all rules contain a common reaction center. An interaction annotation, such as that shown in Fig. 10C, has three parts: a summary of available information about an interaction, including citations from the primary literature; the rules used to model the interactions and/or to summarize the contextual dependencies of the interactions; and an explanation of the rules, including modeling assumptions. If a guide describes a fully specified model, rules will be associated with rate laws and estimates of parameters in the rate laws.

Typically, rules contain contextual information, but every interaction in an extended contact map can be trivially associated with a context-free rule. Thus, every extended contact map corresponds to a set of rules that comprise an executable model composed of context-free rules. A context-free rule is one in which all components are part of a reaction center. Consider the rules of Eqs. (1) and (2), which include contextual components: U and RHD, respectively. If these contextual components are omitted, the rules of Eqs. (1) and (2) become context-free rules.

An extended contact map (e.g., Fig. 4) and a model guide (e.g., Fig. 10) capture more details about a biological system than a BNGL-encoded specification of a model for the or a plain model-derived contact map (e.g., Fig. 2). As discussed previously, explicit representations of enzyme-substrate interactions are often omitted from rules, which is reflected in a model-derived contact map. In contrast, enzyme-substrate relationships are shown in an extended contact map. For example, Lyn-mediated phosphorylation of the linker region in Syk is shown in Fig. 4 but not in Fig. 2. The reason for extra details being included in an extended contact map is that these details are considered in the formulation of a model. If information is collected by a modeler and used to formulate a model, the information should not be lost or separated from a model specification simply because model simulations do not require the explicit incorporation of the information into the formal elements of a model. In addition, an extended contact map and model guide elucidate modeling assumptions. For example, the BNGL-encoded specification of the FcεRI model (22,23), contains a number of modeling assumptions, such as the lumping together of multiple tyrosine residues in the linker region of Syk as a single component, l. Accordingly, a l component appears in Fig. 2, without information about the tyrosine residues that are phosphorylated. In contrast, Fig. 4 identifies three tyrosine residues in the linker region that are

phosphorylated during signaling. Fig. 4 also identifies specific tyrosine residues in the activation loop of the PTK domain of Syk and in the β and γ ITAMs of the receptor that are not shown in Fig. 2. As illustrated by these examples, an extended contact map and a model-derived contact map can be compared to reveal the assumptions of a model.

A guide can be used to specify and annotate a rule-based model, and an extended contact map can be used to illustrate the model. The map provides an extended description of the model, one that goes beyond that provided by the formal model specification. Fig. 4 is more detailed than Fig. 2, which is derived directly from the model and is therefore representative of the formal model specification. Although Fig. 4 is more detailed than Fig. 2, Fig. 4 is restricted in scope to the same molecules, molecular components, post-translational modifications, and interactions considered in the FcεRI model (22,23). Consider dephosphorylation. Phosphatases play an important role in regulating FcεRI signaling (99) but no specific phosphatases are included in the model. Instead, unspecified phosphatases are assumed to be available in excess. Accordingly, no phosphatase is shown in Fig. 4. Similarly, phosphorylation and dephosphorylation of a C-terminal tyrosine residue of Lyn is important for regulating Lyn activity and FcεRI signaling (99), but this residue is not included in the model. Rather a certain fraction of total Lyn is assumed to be in active form, a form in which the C-terminal regulatory tyrosine is not phosphorylated. As a general guideline, we suggest that an extended contact map be drawn to reflect the biological knowledge that underlies the model being illustrated by the map.

A

1. Name of molecule

Syk (spleen tyrosine kinase)

2. Summary of essential information about molecule

Syk is the prototypical member of the Syk-family of protein tyrosine kinases [18]. It plays an important role in FcεRI signaling [19]. Syk contains two tandem SH2 domains, a PTK domain, and multiple sites of phosphorylation, including Y317, Y342 and Y346 in the so-called linker region and Y519 and Y520 in the activation loop of the PTK domain. [...]

3. Molecule type definition in BNGL with brief explanation of definition

In the model (model.bngl, ESI), Syk (`Syk`) is assigned the following molecule type definition:

```
Syk(tSH2,l~0~P,a~0~P)
```
(4)

The component `tSH2` represents the tandem SH2 domains of Syk, the component `l` represents tyrosine residues in the linker region of Syk, and the component `a` represents tyrosine residues in the activation loop of the PTK domain in Syk. The components `l` and `a` each have two possible internal states: `0` (not phosphorylated) and `P` (phosphorylated). The model (model.bngl, ESI) implicitly accounts for extracellular, plasma membrane and cytoplasmic compartments. Free Syk is located in the cytoplasmic compartment.

4. Hierarchical graph representation of molecule with embedded annotation

In the extended contact map (Fig. S1, ESI), Syk is represented by a simplified form of the following annotated hierarchical molecule type graph: [...]

C

1. Brief Description

Syk transphosphorylates Y519 and Y520 in the activation loop of Syk.

2. Explanatory Note

Full activation of Syk, needed for degranulation of RBL cells, requires transphosphorylation at the two activation loop tyrosines, Y519 and Y520 [26-28].

3. Rules

Phosphorylation of Syk by Syk is modeled using the following rules:

```
Lig(l!1,l!2).Syk(tSH2!3,a~0).
Rec(a!2,g~P!3).Rec(a!1,g~P!4).
Syk(tSH2!4,a~0) ->
Lig(l!1,l!2).Syk(tSH2!3,a~0).
Rec(a!2,g~P!3).Rec(a!1,g~P!4).
Syk(tSH2!4,a~P) pSS
```
(8a)

```
Lig(l!1,l!2).Syk(tSH2!3,a~P).
Rec(a!2,g~P!3).Rec(a!1,g~P!4).
Syk(tSH2!4,a~0) ->
Lig(l!1,l!2).Syk(tSH2!3,a~0).
Rec(a!2,g~P!3).Rec(a!1,g~P!4).
Syk(tSH2!4,a~P) pSSs
```
(8b)

4. Comments on Rules

The above rules characterize phosphorylation of Syk by Syk. Syk activity is modulated by phosphorylation of tyrosine residues in its activation loop. Phosphorylated Syk (Eq. 8a) is more active than unphosphorylated Syk (Eq. 8b) [...]

B

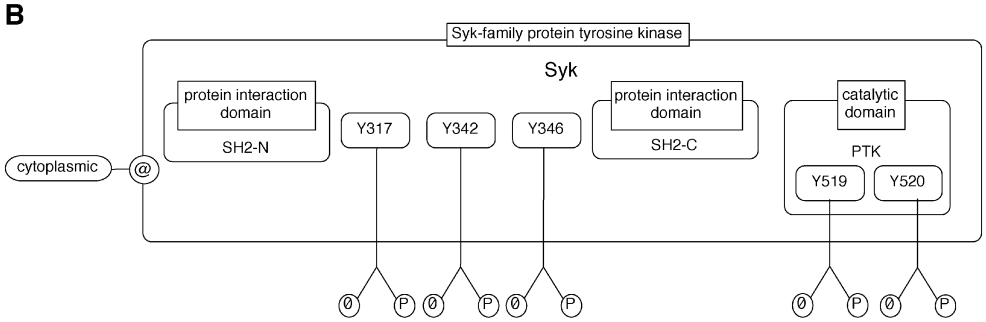


Figure A.2.10. Excerpts from a model guide. (A) A protein annotation includes the name of a protein, a molecule type definition in BNGL, and a summary of relevant information from the literature. Other information may be included as well, such as a UniProt accession number (http://uniprot.org) and experimental data. (B) Illustration of Syk with embedded annotation. This panel demonstrates how an ad hoc illustration of a molecule can be included in a guide to supplement a standardized representation of a molecule in a map. (C) An interaction annotation includes a brief description of the interaction, a listing of all rules that characterize the interaction, and an explanation of modeling assumptions.

Tools for drawing maps

The diagrams presented above were handcrafted using a general-purpose drawing tool, OmniGraffle (The Omni Group, Seattle, WA). OmniGraffle is only available for the Mac platform. Comparable software available on the Windows platform includes Microsoft Visio. Files can be exchanged between OmniGraffle and Microsoft Visio using the Microsoft Visio XML file format. We provide no software for automatically drawing an extended contact map for a given set of rules or for automatically writing context-free rules for a given map. The requirement for manual construction of a map should not be onerous but there are potential pitfalls. For example, a map could be drawn incorrectly so that it is not entirely consistent with an underlying model as intended, or during the process of model development, map and guide updates could fall significantly out-of-sync. However, our goal has been to present a set of standards that are easy to follow and, if followed, should facilitate the understanding and reuse of rule-based models.

To provide software for automatically drawing an extended contact map, we will first need to formalize the relationship between a model and a map and then extend one of the languages for specifying rule-based models (e.g., BNGL or Kappa). These languages do not currently provide a satisfactory means for encoding all of the information that one may wish to visualize in an extended contact map. For example, the catalyst responsible for a reaction represented by a rule is not usually discernible from the rule specification alone. An extension of BNGL could perhaps be introduced to allow for the identification of catalysts and enzyme-substrate interactions in the form of metadata attached to rules or to incorporate the hierarchical graphs of Lemons et al. (65) for more natural representation of structural relationships. The development of software for

drawing extended contact maps, such as the software available for drawing contact maps like that of Fig. 2, is beyond the intended scope of the work presented here, which is focused primarily on establishing guidelines for visualizing and annotating rule-based models.

Conclusions

Large rule-based models are on the horizon. The motivation to develop such models derives in part from the need for analysis tools, such as models, to interpret molecular properties of cancer cells and to guide the treatment of patients on the basis of molecular profiling data (100). As models become larger, it will become increasingly important that models of cell signaling systems be documented and communicated in an understandable way. For the purpose of clear communication of complex information, diagrams have generally proven to be valuable. Readability is essential and weighs against diagrams overloaded with details.

The visualization and annotation guidelines recommended here for rule-based models are likely to aid modelers in three specific ways: 1) in specification of a model, 2) in communication and evaluation of a model, and 3) in reuse of models. As a starting point for modeling, an extended contact map can provide a way of summarizing and assembling information about interactions of interest before the formal elements of a model are specified. A map also provides an outline for organizing the elements of a model. In fact, a map can be used to organize the work of model specification and model annotation: sections in a guide corresponding to elements of a map can be completed one by one using appropriate parts of Fig. 10 as templates. Model communication and evaluation are aided because a map and guide together provide documentation of the basis

for a model. In the hands of a reviewer, a map should be especially useful. A map identifies what molecules and interactions are included in a model. The accompanying guide explains how these molecules and molecular interactions are modeled. If one is an expert on a particular molecule or is concerned about representation of a particular interaction, one can use a map and guide to quickly identify the parts of a model that should be scrutinized. Finally, model reuse is facilitated in part because biological knowledge and modeling assumptions are clearly delineated in a guide. Many parts of a guide, perhaps especially the parts related to biological knowledge, can likely be reused if a model is revised and/or extended, easing the burden of model specification and documentation for modelers who wish to build on the work of others. In fact, because a model specification is divided/organized into units (the sections of a model guide), new models can be quickly built through composition of these units. These benefits are perhaps meager for small models but they should be invaluable for large models and more apparent as more models become available.

We expect that the ideas presented here will be immediately useful for the visualization of (large) rule-based models, as well as for more general-purpose visualization of cell signaling systems when one is concerned about protein substructures and site-specific details of protein interactions. Models can be evaluated more efficiently when their contents can be visualized and their connections to biological knowledge can be identified. A map and associated guide provide an effective way of making these connections for rule-based models. We have attempted to anticipate the needs of those who wish to build large rule-based models of cell signaling systems, considering the visualization of an array of molecule types and molecular interactions found in cell signaling systems (see Fig. 4 and Figs. 6–9). Also, to help ensure serviceable

recommendations, we have leveraged the notational conventions of Kohn and co-workers (66,72). However, at present, the development of large models is not routine, and the guidelines presented here may require modification at some point. In the immediate future, we are dedicated to using and testing these guidelines in our modeling efforts.

References

1. J. M. Bradshaw, Cell Signalling, 2010, 22, 1175–1184. 2
2. M. A. Lemmon and J. Schlessinger, Cell, 2010, 141, 1117–1134.
3. T. Pawson and P. Nash, Science, 2003, 300, 445–452.
4. C. M. Gould, F. Diella, A. Via, P. Puntervoll, C. Gemund, S. Chabanis-Davidson, S. Michael, A. Sayadi, J. C. Bryne, C. Chica, M. Seiler, N. E. Davey, N. Haslam, R. J. Weatheritt A.Budd, T.Hughes, J.Pas, L.Rychlewski, G.Trave, R.Aasland, M. Helmer-Citterich, R. Linding and T. J. Gibson, Nucleic Acids Res., 2010, 38, D167–D180.
5. C. T. Walsh, S. Garneau-Tsodikova and G. J. Gatto Jr., Angew Chem., Int. Ed., 2005, 44, 7342–7372.
6. T. Hunter, Curr. Opin. Cell Biol., 2009, 21, 140–146. 7 B. N. Kholodenko, Nat. Rev. Mol. Cell Biol., 2006, 7, 165–176. 8 B. N. Kholodenko, J. F. Hancock and W. Kolch, Nat. Rev. Mol.
7. B. N. Kholodenko, Nat. Rev. Mol. Cell Biol., 2006, 7, 165–176.
8. B. N. Kholodenko, J. F. Hancock and W. Kolch, Nat. Rev. Mol. Cell Biol., 2010, 11, 414–426.
9. V. Danos, J. Feret, W. Fontana and J. Krivine, Lect. Notes Comput. Sci., 2007, 4807, 139–157.

10. J. Yang, M. I. Monine, J. R. Faeder and W. S. Hlavacek, Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys., 2008, 78, 031910. 11
11. J. Colvin, M. I. Monine, J. R. Faeder, W. S. Hlavacek, D. D. Von Hoff and R. G. Posner, Bioinformatics, 2009, 25, 910–917.
12. J. Colvin, M. I. Monine, R. N. Gutenkunst, W. S. Hlavacek, D. D. Von Hoff and R. G. Posner, BMC Bioinf., 2010, 11, 404.
13. M. W. Sneddon, J. R. Faeder and T. Emonet, Nat. Methods, 2011, 8, 177–183.
14. W. S. Hlavacek, J. R. Faeder, M. L. Blinov, R. G. Posner, M. Hucka and W. Fontana, Sci. STKE, 2006, 2006, re6.
15. J. R. Faeder, M. L. Blinov and W. S. Hlavacek, Methods Mol. Biol., 2009, 500, 113–167.
16. M. L. Blinov, J. R. Faeder, B. Goldstein and W. S. Hlavacek, Bioinformatics, 2006, 20, 3289–3291.
17. J. R. Faeder, M. L. Blinov, B. Goldstein and W. S. Hlavacek, Complexity, 2005, 10, 22–41.
18. D. Endy and R. Brent, Nature, 2001, 409, 391–395.
19. W. S. Hlavacek, J. R. Faeder, M. L. Blinov, A. S. Perelson and B. Goldstein, Biotechnol. Bioeng., 2003, 84, 783–794.
20. B. N. Kholodenko, Nat. Rev. Mol. Cell Biol., 2006, 7, 165–176.
21. B. J. Mayer, M. L. Blinov and L. M. Loew, J. Biol., 2009, 8, 81.
22. B. Goldstein, J. R. Faeder, W. S. Hlavacek, M. L. Blinov, A. Redondo and C. Wofsy, Mol. Immunol., 2002, 38, 1213–1219.
23. J. R. Faeder, W. S. Hlavacek, I. Reischl, M. L. Blinov, H. Metzger, A. Redondo, C. Wofsy and B. Goldstein, J. Immunol., 2003, 170, 3769–3781.

24. M. L. Blinov, J. R. Faeder, B. Goldstein and W. S. Hlavacek, Biosystems, 2006, 83, 136–151.

25. J. R. Faeder, M. L. Blinov, B. Goldstein and W. S. Hlavacek, Syst. Biol., 2005, 2, 5–15.

26. D. Barua, J. R. Faeder and J. M. Haugh, Biophys. J., 2007, 92, 2290–2300.

27. D. Barua, J. R. Faeder and J. M. Haugh, J. Biol. Chem., 2008,

28. 283, 7338–7345. 28 D. Barua, J. R. Faeder and J. M. Haugh, PLoS Comput. Biol., 2009, 5, e1000364.

29. L. M. Heiser, N. J. Wang, C. L. Talcott, K. R. Laderoute, M. Knapp, Y. Guan, Z. Hu, S. Ziyad, B. L. Weber, S. Laquerre, J. R. Jackson, R. F. Wooster, W. L. Kuo, J. W. Gray and P. T. Spellman, Genome Biol., 2009, 10, R31.

30. A. Nag, M. I. Monine, J. R. Faeder and B. Goldstein, Biophys. J., 2009, 96, 2604–2623.

31. A. Nag, M. I. Monine, M. L. Blinov and B. Goldstein, J. Immunol., 2010, 185, 3268–3276.

32. A. Nag, J. R. Faeder and B. Goldstein, IET Syst. Biol., 2010, 4, 334–347.

33. M. K. Malleshaiah, V. Shahrezaei, P. S. Swain and S. W. Michnick, Nature, 2010, 465, 101–105.

34. M. I. Monine, R. G. Posner, P. B. Savage, J. R. Faeder and W. S. Hlavacek, Biophys. J., 2010, 98, 48–56.

35. J. C. J. Ray and O. A. Igoshin, PLoS Comput. Biol., 2010, 6, e1000676.

36. G. R. Smith and D. P. Shanley, PLoS One, 2010, 5, e11092.

37. L. Lok and R. Brent, Nat. Biotechnol., 2005, 23, 131–136.

38. S. S. Andrews, N. J. Addy, R. Brent and A. P. Arkin, PLoS Comput. Biol., 2010, 6, e1000705.

39. M. Meier-Schellersheim, X. Xu, B. Angermann, E. J. Kunkel, T. Jin and R. N. Germain, PLoS Comput. Biol., 2006, 2, e82.

40. M. Koschorreck and E. D. Gilles, BMC Syst. Biol., 2008, 2, 91.

41. I. I. Moraru, J. C. Schaff, B. M. Slepchenko, B. M. L. F. Morgan, A. Lakshminarayana, F. Gao, Y. Li and L. M. Loew, IET Syst. Biol., 2008, 2, 352–362.

42. A. Mallavarapu, M. Thomson, B. Ullian and J. Gunawardena, J. R. Soc., Interface, 2009, 6, 257–270.

43. M. Lis, M. N. Artyomov, S. Devadas and A. K. Chakraborty, Bioinformatics, 2009, 25, 2289–2291.

44. J. Feret, V. Danos, J. Krivine, R. Harmer and W. Fontana, Proc. Natl. Acad. Sci. U. S. A., 2009, 106, 6453–6458.

45. G. Gruenert, B. Ibrahim, T. Lenser, M. Lohel, T. Hinze and P. Dittrich, BMC Bioinf., 2010, 11, 307.

46. J. F. Ollivier, V. Shahrezaei and P. S. Swain, PLoS Comput. Biol., 2010, 6, e1000975.

47. R. Harmer, V. Danos, J. Feret, J. Krivine and W. Fontana, Chaos, 2010, 20, 037108.

48. J. R. Faeder, M. L. Blinov and W. S. Hlavacek, Proceedings of the 2005 ACM Symposium on Applied Computing, Santa Fe, NM, 2005, pp. 133–140.

49. M. L. Blinov, J. Yang, J. R. Faeder and W. S. Hlavacek, Lect. Notes Comput. Sci., 2006, 4230, 89–106.

50. B. Hu, G. M. Fricke, J. R. Faeder, R. G. Posner and W. S. Hlavacek, Bioinformatics, 2009, 25, 1457–1460.

51. L. A. Harris, J. S. Hogg and J. M. Faeder, Proceedings of the 2009 Winter Simulation Conference, 2009, pp. 908–919.

52. M. D. Jacobs and S. C. Harrison, Cell, 1998, 11, 749–758.

53. E. O'Dea and A. Hoffmann, Cold Spring Harb. Perspect. Biol., 2010, 2, a000216.

54. S. C. Garman, B. A. Wurzburg, S. S. Tarchevskaya, J. P. Kinet and T. S. Jardetzky, Nature, 2000, 406, 259–266.

55. A. Kulczycki Jr. and H. Metzger, J. Exp. Med., 1974, 140, 1676–1695.

56. J. C. Cambier, J. Immunol., 1995, 155, 3281–3285.

57. Le Novere, M. Hucka, H. Mi, S. Moodie, F. Schreiber, Sorokin, E. Demir, K. Wegner, M. I. Aladjem, M. Wimalaratne, F. T. Bergman, R. Gauges, P. Ghazal, Kawaji, L. Li, Y. Matsuoka, A. Villeger, S. E. Boyd, Calzone, M. Courtot, U. Dogrusoz, T. C. Freeman, Funahashi, S. Ghosh, A. Jouraku, S. Kim, F. Kolpakov, Luna, S. Sahle, E. Schmidt, S. Watterson, G. Wu, I. Goryanin, B. Kell, C. Sander, H. Sauro, J. L. Snoep, K. Kohn and Kitano, Nat. Biotechnol., 2009, 27, 735–741.

58. Kitano, A. Funahashi, Y. Matsuoka and K. Oda, Nat. Biotechnol., 2005, 23, 961–966.

59. K. Oda, Y. Matsuoka, A. Funahashi and H. Kitano, Mol. Syst. Biol., 2005, 1, 2006.0010.

60. K. Oda and H. Kitano, Mol. Syst. Biol., 2006, 2, 2006.0015.

61. E. Caron, S. Ghosh, Y. Matsuoka, D. Ashton-Beaucage, M. Therrien, S. Lemieux, C. Perreault, P. P. Roux and H. Kitano, Mol. Syst. Biol., 2010, 6, 453.

62. V. Danos, J. Feret, W. Fontana, R. Harmer and J. Krivine, Lect. Notes Comput. Sci., 2007, 4703, 17–41.

63. M. L. Blinov, J. Yang, J. R. Faeder and W. S. Hlavacek, Nat. Biotechnol., 2006, 24, 137–138.

64. L. Holm and C. Sander, Science, 1996, 273, 595–603.

65. N. W. Lemons, B. Hu and W. S. Hlavacek, BMC Bioinf., 2011, 12, 45.

66. K. W. Kohn, M. I. Aladjem, S. Kim, J. N. Weinstein and Y. Pommier, Mol. Syst. Biol., 2006, 2, 51.

67. M. L. van Iersel, T. Kelder, A. R. Pic, K. Hanspers, S. Coort, B. R. Conklin and C. Evelo, BMC Bioinf., 2008, 9, 399.

68. K. W. Kohn, M. I. Aladjem, J. N. Weinstein and Y. Pommier, Mol. Biol. Cell, 2006, 17, 1–13.

69. A. Luna, E. I. Karac, M. Sunshine, L. Chang, R. Nussinov, M. I. Aladjem and K. W. Kohn, BMC Bioinf., 2011, 12, 167.

70. K. W. Kohn, M. I. Aladjem, J. N. Weinstein and Y. Pommier, Cell Cycle, 2009, 8, 2281–2299.

71. J. Yang, X. Meng and W. S. Hlavacek, IET Syst. Biol., 2010, 4, 453–466.

72. S. Kim, M. I. Aladjem, G. B. McFadden and K. W. Kohn, PLoS Comput. Biol., 2010, 6, e1000665.

73. Y. L. Deribe, T. Pawson and I. Dikic, Nat. Struct. Mol. Biol., 2010, 17, 666–672.

74. J. Krivine, V. Danos and A. Benecke, Lect. Notes Comput. Sci.,

75. 2009, 5643, 17–32. 75 J. A. P. Sekar and J. R. Faeder, Methods Mol. Biol, in press.

76. D. L. Nelson and M. M. Cox, Lehninger Principles of Biochemistry, 5th Edition, W.H. Freeman, 2008.

77. R. C. Taylor, S. P. Cullen and S. J. Martin, Nat. Rev. Mol. Cell Biol., 2008, 9, 231–241.

78. S. K. Law and A. W. Dodds, Protein Sci., 1997, 6, 263–274.

79. A. Sahu and J. D. Lambris, Immunol. Rev., 2001, 180, 35–48.

80. B. J. Janssen, E. G. Huizinga, H. C. Raaijmakers, A. Roos, M. R. Daha, K. Nilsson-Ekdahl, B. Nilsson and P. Gros, Nature, 2005, 437, 505–511.

81. G. G. Chiang and B. M. Sefton, J. Biol. Chem., 2001, 276, 23173–23178.

82. F. Vasquez and P. Devreotes, Cell Cycle, 2006, 5, 1523–1527.

83. B. Miao, I. Skidan, J. Yang, A. Lugovsky, M. Reibarkh, K. Long, T. Brazell, K. A. Durugkar, J. Maki, C. V. Ramana, B. Schaffhausen, G. Wagner, V. Torchilin, J. Yuan and A. Degterev, Proc. Natl. Acad. Sci. U. S. A., 2010, 107, 20126–20131.

84. J. D. Schnell and L. Hicke, J. Biol. Chem., 2003, 278, 48–56.

85. D. Hoeller, N. Crosetto, B. Blagoev, C. Raiborg, R. Tikkanen, S. Wagner, K. Kowanetz, R. Breitling, M. Mann, H. Stenmark and I. Dikic, Nat. Cell Biol., 2006, 8, 163–169.

86. R. J. Deshaies and C. A. Joazeiro, Annu. Rev. Biochem., 2009, 78, 399–434.

87. Z. Yang and D. J. Klionsky, Curr. Top. Microbiol. Immunol., 2009, 335, 1–32.

88. J. Colicelli, Sci. STKE, 2004, 2004, re13.

89. J. Gureasko, O. Kuchment, D. L. Makino, H. Sondermann, D. Bar-Sagi and J. Kuriyan, Proc. Natl. Acad. Sci. U. S. A., 2010, 107, 3430–3435.

90. X. F. Zhang, J. Settleman, J. M. Kyrakis, E. Takeuchi-Suzuki, S. J. Elledge, M. S. Marshall, J. T. Bruder, U. R. Rapp and J. Avruch, Nature, 1993, 364, 308–312.

91. J. A. Pesin and T. L. Orr-Weaver, Annu. Rev. Cell Dev. Biol., 2008, 24, 475–499.

92. E. Nicolas, C. Roumillac and D. Trouche, Mol. Cell Biol., 2003, 23, 1614–1622.

93. 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, Mol. Cell, 2003, 12, 541–552.

94. T. Kesti, A. Ruppelt, J. H. Wang, M. Liss, R. Wagner, K. Tasken and K. Saksela, J. Immunol., 2007, 179, 878–885.

95. P. C. Weber, D. H. Ohlendorf, J. J. Wendoloski and F. R. Salemme, Science, 1989, 243, 85–88.

96. E. Jain, A. Bairoch, S. Duvaud, I. Phan, N. Redaschi, B. E. Suzek, M. J. Martin, P. McGarvey and E. Gasteiger, BMC Bioinf., 2009, 10, 136.

97. R.D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. Sonnhammer, S. R. Eddy and A. Bateman, Nucleic Acids Res., 2010, 38, D211–D222.

98. J. Ren, L. Wen, X. Gao, C. Jin, Y. Xue and X. Yao, Cell Res., 2009, 19, 271–273.

99. J. Rivera, N. A. Fierro, A. Olivera and R. Suzuk, Adv. Immunol., 2008, 98, 85–120.

100. P. K. Kreeger and D. A. Lauffenburger, Carcinogenesis, 2010, 31, 2–8.