ABSTRACT

DAVIS, XIN YANG. Computational Modeling of Cell Signaling Network Using Hill Function and Markov Chain Monte Carlo Methods. (Under the direction of M. Todd See and Roger L. McCraw).

Computational models have been successfully used by mathematicians, chemists and physicists, but rarely used by biologists. The reasons are probably related to the dynamics in biological systems and the uncertainty in a surrounding environment. Computational models can be very useful tools when they provide experimentally testable predictions. In this study, we developed a new mathematical modeling method using the Hill function (see glossary) and the Markov Chain Monte Carlo (MCMC) methods (see glossary). We modeled the epidermal growth factor receptor (EGFR) signaling transduction network with no experimental data available for the biochemical reactions. The model is based on the signaling directions and activation or inhibition information in the network. We also used this mathematical modeling method to model epidermal growth factor (EGF)-stimulated phosphoinositide transfer protein (PITP) to predict the structure of the signal transduction network at a systems level. The research was performed in two stages. First, we developed a new mathematical modeling method using the Hill function and MCMC methods, and modeled the EGFR signal transduction network. Experimental data of the response from extracellular-signal-regulated kinase (Erk) and Akt were used in fitting the EGFR model. Simulations were used to select parameters to generate family of solutions of responses from Erk and Akt based on regulatory activation or inhibition in the model. Transient response from Akt in the phosphoinositide 3-kinase (PI3K) pathway precedes the response from Erk in Mitogen -activated protein kinase (MAPK) pathway, and the response from Erk remains at its peak level for minutes. Statistical data analysis reveals the biological characteristics in the

signaling network model that indicate that cell signaling is mainly enzymatic regulation. Second, we developed a model EGF-stimulated PITP in EGFR signaling network with four different structures. From statistical data analysis and by comparing the simulation results, we observed some common properties among these models. Ras GDP and GTP conversion, and phosphatidylinositol (PI) and phosphatidylinositol 3, 4-biphosphate (PIP_2) conversion are more stable and balanced in the network for all four models. We simulated Ras, PI3K oncogenes and tumor suppressor phosphatase and tesin homolog (PTEN) and observed the dynamic behavior of the network at the system level. Observations of the behavior of this network can be related to cancer research. We made predictions that Model B and Model D would be more closely similar to experimental data. Simulation results demonstrated that our model, using this mathematical modeling method, can provide qualitative insights when information about chemical kinetics is unavailable for a complex signaling network.

Computational Modeling of Cell Signaling Network Using Hill Function and Markov Chain Monte Carlo Methods

by
Xin Yang Davis

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Quantitative Genetics

Raleigh, North Carolina

July 6, 2010

APPROVED BY:

_____	_____
M. Todd See Committee Chair	Roger L. McCraw Committee co-Chair
_____	_____
Vytas A. Bankaitis	Jason M. Haugh
_____	_____
Charles E. Smith	David A. Dickey

DEDICATION

This dissertation is dedicated to my mentors and my advisory committee. All of you have trained me to be a good scientist. I would like to express my deep gratitude to Dr. Roger McCraw and Dr. Vytas Bankaitis who made this research project possible. It has been a great experience and a happy journey. Most importantly it satisfies my curiosity.

BIOGRAPHY

Where there is a WILL, there is a way.

ACKNOWLEDGMENTS

I want to thank my mentors and committee members for your valuable advices and training. All of you are my role models for your dedication to research and training students. Thanks for your kindness that you always had time for my research project despite you all have a very busy schedule.

Dr. Roger McCraw introduced me to quantitative genetics, accepted me as his student even though the time was close to his retirement. As my mentor he encouraged me to find research projects interest me the most. He provided funding for my research and the international conferences I attended during my graduate school study.

Dr. Todd See was my committee chair when he had many responsibilities as the department head and professor at the same time. I benefited from his advices on how to improve my presentation skills.

Dr. Vytas Bankaitis allowed me to do research project in his lab despite his many responsibilities and very busy schedule. He mentored me on cell biology and advised me on research project on weekly basis. Because of working for him, I had the privilege to sit in many courses in UNC medical school and gained knowledge useful for my research project.

Dr. Jason Haugh taught me cell signaling modeling and spent many hours to advise me on my research project.

Dr. Charles Smith offered his office hours to me for asking questions and suggested many good modeling books for me to read.

Dr. David Dickey advised me on statistics, and always had time to answer questions.

Thanks to Dr. Gabriel Weinreb, Dr. Ken Jacobson, Dr. Tim Elston and Dr. Ju Youn Beak for their valuable discussions.

I want to thank professors at NCSU and UNC who allowed me to sit in their lectures and let me access their Blackboard for course materials. Their generosity provided me the opportunity to learn more knowledge that has greatly helped my research project and beyond.

Dr. Jeffrey Thorne introduced me to systems biology, which led me to do this research project at UNC medical school.

Dr. Miles Engell helped me to get into her Biology class. Her class was a good foundation in biology for me and enabled me to continue on learning Cell Biology, Neurobiology, and completing this research project in cell signaling.

The administrative personnel in the UNC Department of Cell and Developmental Biology were very helpful in arranging project meetings, and providing parking coupons.

Thank you ALL!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Supplement

GLOSSARY

Markov Chain Monte Carlo method - simulate direct draws from some complex, nonstandard multivariate distributions of interest [Chib et al. 1995].

Markov Chain - a sequence of random variables generated by the Markov process which is defined by its transition probabilities. The transition probability is a conditional distribution function that represents the probability of moving from the current point to a future point for the same parameter [Walsh 2004].

Monte Carlo - sampling uses Bayesian inference that is based on random sampling. It was introduced by Metropolis [Metropolis et al. 1953] at Los Alamos National Laboratory.

Hill function - in 1910, British physiologist Archibald Vivian Hill formulated the Hill function to describe the sigmoidal O_2 binding curve of hemoglobin [Hill 1910].

Hill coefficient - the Hill coefficient *n* quantifies the binding cooperativeness resulted from the interactions between binding sites. A coefficient of 1 indicates non-cooperative; a coefficient greater than 1 indicates positive cooperative binding; a coefficient less than 1 indicates negative cooperative binding [Hill 1910].

EGFR - epidermal growth factor receptor (EGFR) is a plasma membrane receptor protein, binds to EGF extracellular signaling molecule. The binding activates the receptor, and activates the intracellular signaling pathway.

EGFR dimer - two activated EGFR form an EGFR dimer.

Dimer auto-phosphorylation - once a receptor dimer is formed, it gains tyrosine kinase activity and can auto-phosphorylate on several tyrosine residues. Proteins Shc, Grb_2, PTP-1B

and Gab_1 are recruited by binding directly to tyrosine-phophorylated receptors [Birtwistle et al. 2007].

Grb_2 - an adaptor protein and links receptor with the Ras-MAPK signaling pathway. It is recruited by binding directly to tyrosine phosphorylated EGFR dimer.

Shc - an adaptor protein that contains both PTB and SH2 domains and becomes phosphorylated on tyrosine in response to many different extracellular stimuli. It is recruited by binding directly to tyrosine phosphorylated EGFR dimer, and inhibited by PTP-1B.

PTP-1B - protein tyrosine phosphatase-1B (PTP-1B) is a negative regulator of insulin signaling. It is activated by tyrosine phosphorylated EGFR dimer, and inhibits Shc, tyrosine phosphorylated EGFR dimer, and Gab_1.

Gab_1 - Grb_2 associated binding protein 1 (Gab_1), is secondary recruit and binding to Shc SH3 domains of Grb_2. It is also activated by Grb_2. Gab_1 can also be recruited to the membrane via its PH domain binding to PIP_3 [Holgado-Madruga et al. 2003].

Sos_2 - son-of-sevenless (also called Ras-GEF, guanine nucleotide exchange factor) is an exchange factor, stimulates the inactive Ras protein to replace its bound GDP by GTP, which activates RasGDP to relay the signal downstream. SOS_2 is secondary recruit, and binds to Src SH3 domains of Grb_2.

RasGDP - Ras protein in GDP-bond form and inactive. It becomes active when exchanges GDP for a GTP molecule in response to extracellular signals.

RasGTP - Ras protein which relays signals from cell-surface receptors to downstream singling pathway. It is a molecular switch which operates in intracellular signaling pathway and cycling between two conformational states, active when bound with GTP and inactive

when bound with GDP. RasGTP is converted back to RasGDP by RasGAP. RasGTP activates Raf_1, and has link to PI3K.

RasGAP - Ras GTPase-activating protein is a regulatory protein and controls GTP-binding proteins. It deactivates the proteins by hydrolysis of bound GTP. RasGAP is activated by tyrosine phosphorylated EGFR dimer and Gab_1.

Raf1 - (also called MAPKKK), a protein kinase, and receiving an activating signal directly from Ras, it phosphorylates and activates MEK (MAPKK).

MEK - (also called MAPKK), a protein kinase. It is phosphorylated and activated by Raf1, and in turn activates MAPK.

Erk - extracellular-signal-regulated kinase, (also called MAPK), a protein kinase, activated by Raf1 through MEK. Since the activation of Erk by MEK follows a distributed mechanism [Zha and Zhang, 2001] therefore use a full, mass action description for Erk activation. In the distributed mechanism, the substrate dissociates and rebinds the enzyme between catalysis steps and can lead to potential ultrasensitivity and bistability [Markvich et al. 2004, Ferrel et al. 1998].

PI 3-Kinase - phosphoinositide 3-kinase is plasma-membrane-bound enzyme. Gab_1 binds to PI3K as a crosstalk in two different pathways.

PIP_2 - phosphatidylinositol 4,5-biphosphate is a phosphorylated inositol phospholipid. It is a substrate and binds to PI3K to produce PIP_3. It is cleaved by PLC_r to make IP_3 and DAG.

PI 3,4,5 P_3 - phosphatidylinositol (3,4,5) P_3 is a phosphoinositide produced by PIP_2 when phosphorylated by PI3K. PIP_3 binds to Gab1 via Gab1 PH domain to recruit Gab1 to the membrane.

PTEN - phosphatase and tensin homolog is a tumor suppressor. It is a negative regulator to regulate PIP_3 and prevent PIP_3 accumulation in the plasma membrane, therefore inhibiting PI3K signaling.

Akt - (also called PKB) is a protein, converted by PIP_3 from inactive form to active form.

PITP - phosphoinositide transfer protein can bind and exchange one molecule of phosphatidylinositol (PI) and facilitate the transfer of these lipids between different membrane compartments. PITP is critical regulator of phosphoinositides in cellular compartments, and participate in signal transduction and in membrane traffic. Dysfunction of PITP may lead to neurodegeneration diseases.

PI4K - PtdIns 4-OH kinase is an enzyme, activated by $PITP_a$ and converts PI to PIP_2.

PI – phosphatidylinositol is a lipid at cell membrane and in metabolic process. It is phosphorylated to form PIP.

Sac - suppressor of actin is an enzyme, and converts PIP_2 back to PI.

PLC_r - phospholipase C is a plasma-membrane-bound enzyme that cleaves PIP_2 to produce IP_3 and DAG.

IP_3 - inositol 1,4,5-trisphosphate is a small intracellular mediator. It diffuses through the cytosol and release Ca^{2+} from the endoplasmic reticulum (ER) by binding to and opens IP_3-gated Ca^{2+} - release channels (IP_3 receptors) in the ER membrane.

DAG - diacylglycerol is a small intracellular mediator, embedded in the plasma membrane.

REFERENCES

Alberts, B. et al. (2008) Molecular Biology of THE CELL, Fifth Edition.

Ben-Shlomo I, Yu Hsu S, Rauch R et al. (2003) Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Sci STKE* 187:RE9.

Berridge MJ, Bootman MD & Roderick HL (2003) Calcium signaling dynamics, homeostasis and remodeling. *Nature Rev Mol Cell Biol* 4:517-529.

Birtwistle, MR et al. 2007. Ligand-dependet responses of the ErbB signaling network: experimental and modeling analyses. *Mol Systems Biol* 3:144.

Bourne HR (1995) GTPases: a family of molecular switches and clocks. *Philos Trans R Soc. Lond B Biol Sci* 349:283-289.

Bradshaw RA & Dennis EA (eds) (2003) Handbook of Cell Signaling. *Elsevier*: St. Louis.

Bretscher, Otto (2005). *Linear Algebra with Applications, 3rd ed.* Upper Saddle River NJ: Prentice Hall.

Burns ME & Baylor DA (2001) Activation, deactivation, and adaptation in vertebrate photoreceptor cells. *Annu Rev Neurosci* 24:779-805.

Chen, S., Cowen, C.F.N, and Grant, P.M. (1991). *IEEE TRANSACTIONS ON NEURAL NETWORKS,* Vol. 2, No. 2.

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings Algorithm. *The American Statistician, Vol.* 49, No. 4, 327-335.

Dard N & Peter M (2006) Scaffold proteins in MAP kinase signaling: more than simple passive activating platforms. *BioEssays* 28:146-156.

Dong, C. et al. (2002) MAP kinases in the immune response. *Annu Rev Immunol* 20:55.

Downward J (2004) PI 3-kinase, Akt and cell survival. *Semin Cell Dev Biol* 15:177-182.

Ferrell JE. Jr. (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 14:140-148.

Frenkel, D. (2004) Introduction to Monte Carlo Methods. *NIC Series, Vol. 23,* 29-60.

Gelman, A., and D. B. Rubin. (1992). Inferences from iterative simulation using multiple sequences (with discussion). *Statistical Science* 7: 457 - 511.

Geyer, C. J. 1992. Practical Markov chain Monte Carlo (with discussion). *Stat. Sci.* 7: 473–511.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* 57:97-109.

Metropolis, N. et al (1953) Equations of state calculations by fast computing machines. *J. of Chemical Physics* 21:1087-1091.

Hill AV (1910) The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *Proc Of The Phys Society*.

Holgado-Madruga et al. (2003), Gab1 is an integrator of cell death versus cell survival signals in oxidative stress, *Mol Cell Biol* 13 (23) : 4471-84.

Hudmon A & Schulman H (2002) Structure-function of the multifunctional Ca^{2+}/calmodulin-dependent protein kinase II. *Biochem J* 364:593-611.

Kolch, W. et al. (2005) When kinases meet mathematics: the systems biology of MAPK signalling. *FEBS Letters* 579: 1891–1895.
Liu, Y. and Bankaitis, V.A. (2010) Phosphoinositide phosphatases in cell biology and disease. *Progress in Lipid Research* **49:**201-217.

Luttrell LM (2006) Transmembrane signaling by Gprotein-coupled receptors. *Methods Mol Bio* 332:3-49.

Mitin N, Rossman KL & Der CJ (2005) Signaling interplay in Ras superfamily function. *Curr Biol* 15:R563-574.

Moghal, N. and Sternberg, P.W. (1999) Multiple positive and negative regulators of signaling by the EGF-receptor. Curr. Opin. Cell Biol. 11, 190-196.

Mullschleger S, Leowith R & Hall MN (2006) OR signaling in growth and metabolism. *Cell* 124:471-484.

Pawson T (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* 116:191-203.

Pierce KL, Premont RT & Lefkowitz RJ (2002) Seven-transmembrane receptors. *Nature Rev Mol Cell Biol* 3:639-650.

Raftery, A. E., and S. Lewis. 1992b. Comment: One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo. *Stat. Sci.* 7: 493–497.

Reddy, C.C. et al. (1994) Proliferative response of fibroblasts expressing internalization-deficient epidermal growth factor (EGF) receptors is altered via differential EGF depletion effect. *Biotechnol. Prog.* 10, 377-384.

Rhee SG (2001) Regulation of phosphoinositide-specific phospholipase C. *Annu Rev Biochem* 70:281-312.

Roberts, G.O. et al (1994) Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *Technical Report,* University of Cambridge.

Robishaw JD & Berlot CH (2004) Translating G protein subunit diversiy into functional specificity. *Curr Opin Cell Biol* 16:206-209.

Roskoski R Jr (2004) Src protein-tyrosine kinase structure and regulation. *Biochem Biophys Res Commun* 324:1155-1164.

Qi M & Elion EA (2005) MAP kinase pathways. *J Cell Sci* 118:3569-3572.

Schlessinger J (2000) Cell signaling by receptor tyrosine kinases. *Cell* 103:211-225.

Schwartz MA & Madhani HD (2004) Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. *Annu Rev Genet* 38:725-748.

Science's Signal Transduction Knowledge Environment (Stke): www.stke.org

Seet BT, Kikic I, Zhou MM & Pawson T (2006) Reading protein modifications with interaction domains. *Nature Rev Mol Cell Biol* 7:473-483.

Shaw RJ & Cantley IC (2006) Ras, PI(3)K and mTOR signaling controls tumour cell growth. *Nature* 44:424-430.

Shaywitz AJ & Greenberg ME (1999) CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annu Rev Biochem* 68:821-861.

van der Geer, P. et al. (1994) Receptor proteintyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell Biol.* 10, 251–337.

Walsh, B. (2004) Markov Chain Monte Carlo and Gibbs Sampling. *Lecture Notes for EEB 581.*

Wassarman DA, Therrien M & Rubin GM (1995) The ras signaling pathway in Drosophila. *Curr Opin Genet Dev* 5:44-50.

Wells, A. (1999) EGF receptor. *Int. J. Biochem. Chell Biol.* 31, 637-643.

Wetzels, R. et al. (2009) Bayesian Inference Using WBDev: A Tutorial for Social Scientists.

Chapter One **Background and Motivation**

INTRODUCTION

Computational models have been successfully used by mathematicians, chemists and physicists, but are rarely used by biologists. The reason is probably due to the uncertainty of dynamic biological systems interacting with the surrounding environment. Computational models can be very useful tools when they provide experimentally testable predictions. Modeling complex intracellular signal transduction networks presents challenges. Models based on kinetics of chemical reactions most likely contain a large number of parameters, and rate constants for many parameters are unknown because of the difficulty of experimentally determining the rate constants for individual steps in these networks. [Brown et al. 2004]. It would be even more difficult when the signaling network involves a large number of protein-protein interactions.

Our research objective was to model a signal transduction network using coarse-grained computational modeling methodology. We believe that our research is the first attempt to study the Epidermal Growth Factor (EGF) receptor signal transduction network including all three major signaling pathways [von der Geer, P., 1994, Schlessinger, J., 2000] in the intracellular signal transduction network. The pathways involved are the phosphatidylinositol 3-kinase (PI3K) pathway, Mitogen activated protein kinase (MAPK) pathway, and phospholipids C (PLC_r) pathway. In this study, our goals are to understand the dynamic behavior of epidermal growth factor (EGF) receptor signal transduction network by developing mathematical models to make experimentally testable predictions, and to explain novel experimental results. These predictions were validated by experimental data. The EGF

receptor (EGFR) is a useful test case for modeling cell signaling networks because EGF is in different cell types, and antibodies and reagents are available for testing [Wiley et al. 2003]. The most intensively studied signaling pathway is the MAPK pathway.

We developed a new mathematical modeling method, a coarse-grained systems biology modeling method, using the Hill function (see glossary) and the Markov Chain Monte Carlo (MCMC) methods (see glossary). We modeled the epidermal growth factor receptor (EGFR) signaling transduction network with no experimental data available for the biochemical reactions. The model is based on the signaling directions and activation or inhibition information in the network. We also used this mathematical modeling method to model EGF-stimulated phosphoinositide transfer protein (PITP) to predict the structure of the signal transduction network at a systems level. The limitation is that some detailed information is not captured by the model.

Concerning the constraints speed of simulations and the difficulty of experimentally testing the predictions from the model, building a simple model that can explain complex biological processes is our ultimate goal. Our models made perturbations on phosphorylates PI (4,5) P_2 to simulate a mutational study and explain the outcome of novel experimental data. Our focus is to investigate protein-protein interactions in the EGFR signal transduction network at the system level.

In this chapter, the intracellular signal transduction network and systems biology are reviewed. Both intracellular signal transduction network and systems biology provide background for understanding the modeling effort on EGFR signal transduction network at a

systems level. Next, the outline of research is highlighted, and finally the major results of this research are summarized.

INTRACELLULAR SIGNAL TRANSDUCTION NETWORK

Cell communication is mediated by extracellular signal molecules through cell signaling pathways or cell signal transduction network. Some signals travel over long distances to cells far away via axons, and some signals are shared by immediate neighbors through gap junctions. Signal molecules can be proteins, small peptides, amino acids, or many other kinds of molecules. Most extracellular signal molecules are hydrophilic and, therefore, are not able to enter the plasma membrane of the target cell. They bind to specific cell-surface receptors in the cell membrane and the activated receptors generate signals inside the target cell [Bradshaw et al. 2003].

There are different ways that signal proteins process and relay signals in the signaling network. Seven ways the protein may function are: (1) *relay* the signal to the next signaling protein in the pathway; (2) act as *a scaffold* to bring more signaling proteins together that then can interact quickly and efficiently; (3) *transform* or *transduce* the signal into different forms for the signal stimulating a cell response; (4) *amplify* the signals received by producing large amounts of a small intracellular mediator or by activating many copies of a downstream signaling protein; (5) *integrate* signals received from different signaling pathways then relaying a signal onward; (6) *spread* the signal from one pathway to another and create branch or cross links; it may *anchor* signals in the pathway; (7) *modulate* the activity of other signaling proteins and thereby regulate the strength of signaling along a pathway [Alberts et al. 2008].

Each cell is programmed to respond to specific combinations of extracellular signal molecules. They respond differently at different times and locations. Apoptosis is programmed cell death that is caused by deprived signals. Different cells respond to different extracellular signal molecules, and different types of cells respond differently to the same extracellular signal molecule. The response to extracellular signals from cells can be transient with duration of seconds to minutes or even hours, or oscillating, or remaining at steady state. How does a cell interpret a specific combination of signals and make the decision to divide, to grow, or to differentiate? Answering this question poses a great challenge in cell biology, signaling modeling, and simulation.

When extracellular signal molecules bind to specific cell-surface receptors in the cell membrane, the receptors activate downstream signals inside the target cell. The intracellular signal molecules relay signals to their effectors in the cell interior, thereby changing the behavior of the cell. Consider epidermal growth factor and its receptor, and three major signaling pathways in the intracellular signal transduction network, namely PI3K pathway, MAPK pathway and PLC_r pathway. Each pathway may follow many different branches, depending on cell type and stimuli. Different branches may form different pathways that lead to the same effectors due to cross links between pathways. Signaling and regulatory pathways consist of genes, proteins and signaling molecules, interconnected as a complex network of interactions.

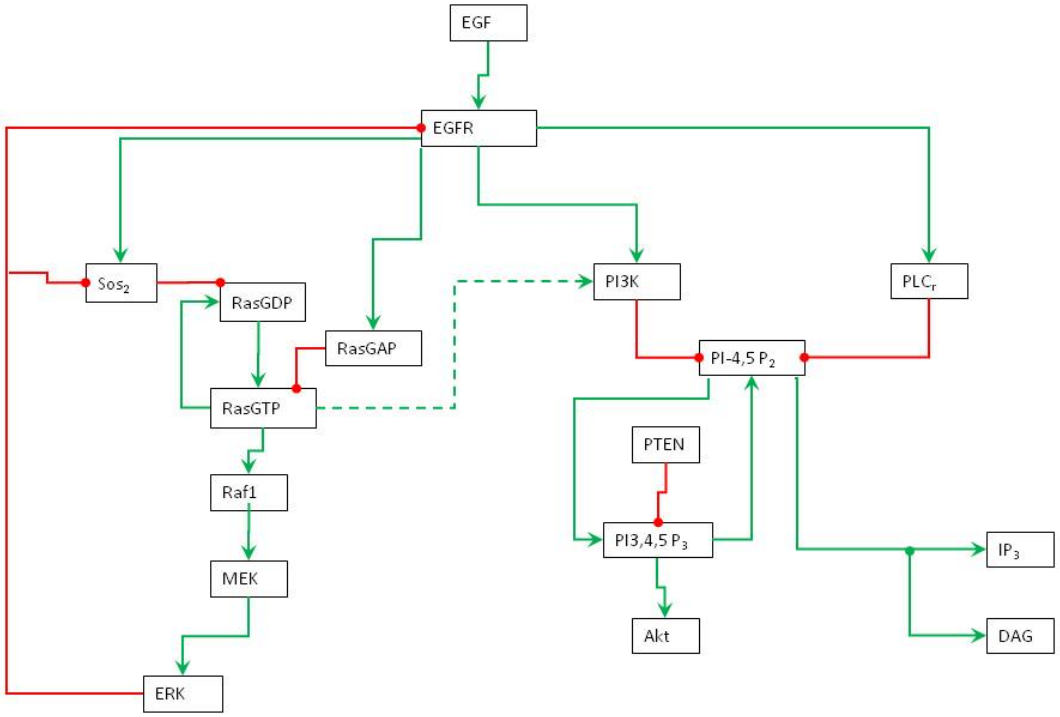

Figure 1: Schematic diagram of EGFR signal transduction network. Green line indicates activation, red line indicates consumption or inhibition.

Epidermal Growth Factor and Receptor

Epidermal growth factor is present in various cell types and stimulates cell growth, proliferation, survival, or differentiation, and functions as an inductive signal in development. EGF is an extracellular signaling molecule and is hydrophilic, therefore, it is not able to enter the plasma membrane of the target cell. It binds to EGF receptor and the activated EGF receptor generates signals inside the target cell.

The EGF receptor is an enzyme-coupled membrane receptor and it is activated by binding to EGF or other homologous ligands. The EGFR family consists of four receptors: ErbB1/EGFR/HER1, ErbB2/Neu/HER2, ErbB3/HER3, and ErbB4/HER4. EGFR has an important role in human diseases. An excess of ErbB1 and ErbB2 can cause cancer in humans. EGF receptors also play crucial roles in propagating signals that regulate cell proliferation, differentiation, motility, and apoptosis [Holbro, 2004]. EGF receptors are pursued as therapeutic targets. The EGFR system has been used in studies of oncogenesis, mitogen-activated-protein-kinase signaling pathways, and many others [Carpenter et al. 2000]. It will continue to be the source of cell signaling studies involved in development and disease [Wells et al. 1999, Moghal et al. 1999]. The EGFR system has been a useful test case for modeling since it is present in different cell types and high-quality antibodies are widely available [Wiley et al. 2003].

EGFR mutant with normal kinase activity but reduced internalization rates causes increased cell proliferation [Reddy et al. 1994]. The ligand-receptor complexes number remained the same for cells expressing mutant or wild-type receptors, but the number of

ligand-receptor complexes remained at higher levels longer for the mutant receptors [Albert et al. 2008,].

PI3-Kinase Pathway

PI3-kinase (PI3K) pathway is involved in cell survival and growth. In this pathway EGF binds to EGF receptor which activates PI3K, in turn PI3K phosphorylates PI (4,5) P_2 to produce the plasma membrane lipid second messenger phosphoinositide-3,4,5-trisphosphate PI (3,4,5) P_3. Phosphatase and tensin homolog (PTEN) negatively regulates PI (3,4,5) P_3 and converts PI (3,4,5) P_3 back to PI (4,5)P_2. Then PI (3,4,5) P_3 swaps its role as an enzyme and activates Akt from inactivated form to activated form Akt. Akt phosphatase dephosphorylates Akt and converts Akt back to inactivated Akt. In the process of phosphorylation and dephosphorylation of Akt, Akt is neither consumed nor produced. This process is important in cell signaling because cell signaling is all about regulation of enzymatic reaction. Activated Akt phosphorylates various target proteins including p70s6 kinase and Bad. Activated Bad releases an inhibitory protein to block apoptosis and promote cell survival which is important in cancer research. In the nucleus, p70s6 kinase activates JUN. PTEN inhibits PI (3,4,5) P_3. Activated Akt inhibits upstream PI3K via a negative feedback loop. PI3K also is an effector of RasGTP in the MAP-kinase pathway that is cross linked between RasGTP and PI3K. Akt also called protein kinase B, is important in mammalian cellular signaling and plays a key role in control of cell survival and growth. Akt inhibits the apoptosis process to allow cell survival, and induces protein synthesis to generate tissue growth and skeletal muscle hypertrophy. Akt was selected for modeling and experimental

measurements in part because of its importance in this signaling pathway and the convenience of obtaining readouts [Albert et al. 2008, Downward 2004, Shaw et al. 2006].

MAP-Kinase Pathway

The MAP-kinase (MAPK) pathway is one of the most intensely studied signaling pathways. It was one of the first connections between extracellular cues and changes in gene expression that was mapped in molecular detail, and it turned out to be involved in the control of a bewildering number of cellular processes including fundamental functions such as cell proliferation, survival, motility, and differentiation [Kolch et al. 2005].

In this pathway EGF binds to EGF receptor thereby activating guanine nucleotide exchange factors (GEFs). Son-of-sevenless (Sos_2) is specific for Ras protein and functions as an enzyme because it speeds up the transformation process of converting the inactive Ras protein to replace its bound GDP by GTP. Ras functions as a molecular switch in two distinct conformational states, active when GTP is bound and inactive when GDP is bound. Ras GTPase-activating proteins (GAPs) induce activated Ras to inactivate itself by hydrolyzing its bound GTP to GDP. Because of the transient activation of Ras, short-lived signaling events are converted into long-lasting events that can relay the signal to the nucleus and alter gene expression, thereby stimulating cells to proliferate or differentiate. The MAP kinase module (Raf_1, MEK, Erk) is used to effect this process. RasGTP phosphorylates downstream Raf_1 (MapKKK), Raf_1 activates MEK (MAPKK), and MEK activates Erk (MAPK). While Raf_1 and MEK have a very restricted set of substrates, Erk features more than 70 substrates including nuclear transcription factors [Kolch et al. 2005]. RasGTP also phosphorylates PI3K pathway. Ras is an oncogene contained in about 25-30% of cancer cells. Therefore, Ras

plays an important role in cancer research and drug discovery, one of many reasons why this signaling pathway is extensively studied [Albert et al, 2008, Mitin et al. 2005, Bourne 1999].

Erk phosphorylates gene regulatory proteins FOS and JUN in the nucleus. These result in complex changes in gene expression and cell behavior. Erk also phosphorylates and inactivates Raf_1 and SOS_2, providing negative feedback loops to help shut off the MAP kinase module. Extracellular signals usually activate Erk transiently and the duration of activation influences the response. In a neural precursor cell line, Erk responds to EGF that peaks at 5 minutes and then rapidly declines, but the same cell line responds to NGF (Nerve Growth Factor) with hours of activation that the cells differentiate into neurons. Phosphorylation of Erk has been shown to be an important event in promoting T cell survival and proliferation [Dong et al. 2002]. Similarly to Akt, Erk was selected for modeling and experimental measurement in part because of its importance and the convenience of obtaining readouts.

Phospholipids C_r Pathway

Phospholipids C (PLC_r) pathway is involved in cell migration. In this pathway EGF receptor binding complex activates PLC_r which cleaves PI (4,5)P_2 to produce diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3), hence this pathway branches into two pathways. When IP_3 reaches the endoplasmic reticulum (ER), it opens IP_3-gated Ca^{2+} -release channels (also called IP_3 receptors) in the ER membrane and hence Ca^{2+} is released from ER to cytosol. DAG and Ca^{2+} bind to each site of protein kinase C (PKC). Activated PKC phosphorylates target proteins, including JUN, in the nucleus [Albert et al. 2008, Rhee 2001].

Ca^{2+} is an important ion because it functions as a ubiquitous intracellular mediator. Many extracellular signals can trigger an increase in Ca^{2+} concentration in the cytosol. It is present in egg cells, muscle cells, and nerve cells. Ca^{2+} also can act as a signal because its concentration is very low in the cytosol and is high in the lumen of ER and sarcoplasmic reticulum (SR) in the muscle. A large gradient tends to drive Ca^{2+} from plasma membrane and the ER into the cytosol and may increase Ca^{2+} concentration 10-20 fold. In order to keep Ca^{2+} concentration low in the cytosol in resting cells, Ca^{2+} is actively pumped from the cytosol to the outside of cell or into the ER and mitochondria. Ca^{2+} responding to extracellular signals can oscillate in part because of a combination of positive and negative feedback loops by Ca^{2+} on IP_3 receptors. The response also can be transient [Albert et al. 2008, Berridge 2005, Berridge et al. 2003].

SYSTEMS BIOLOGY

System biology is a new field of study that aims to understand interactions of genes, proteins and biochemical reactions at a system level. Components of the biological system are genetic, metabolic and signal transduction pathways. The biological system is very complex due to its dynamics and interaction within an environment of uncertainty. The systems may be cells, organisms, or human beings. A reductionist approach of study of basic components such as pathways is limited in its capacity to translate the effect of perturbations in these pathways to the cell as a whole. It is often said that a biological system in its entirety presents more than a sum of all its parts.

Systems biology of necessity, involves multi-disciplinary research that requires collective efforts from multiple research areas: molecular biology to understand regulatory

relationship of genes, and interactions of proteins in the signal transduction and metabolism pathways; computer science to model the biological system, and find simulation results consistent with experimental data; analysis of the dynamics of the system to understand how stimuli and external perturbation affect system behavior; and technologies for high throughput and precision measurements [Kitano, 2000].

Systems biology was pursued by scientific community. In 1933, Walter Cannon proposed a concept of "homeostasis" [Cannon, 1933], a homeostatic system of dynamic equilibria controlled by a regulatory mechanism, and focused at the physiological level. The goal of systems biology is to fundamentally transform the traditional practice of medicine to predictive, preventive, and personalized medicine [Hood et al. 2004]. By examining an individual's genomic makeup and protein markers, a physician can make predictions about the probability of having a particular disease and current status, and is able to provide guidance for preventive treatments, or customized therapeutic drugs. It is well recognized that many mechanisms involved in complex diseases such as cancer cannot be understood on the basis of molecular parts. For example, Ras is found in 25-30% of cancer cells and has been a cancer drug target for over two decades, but no cure has been found yet, perhaps due to lack of system-level understanding of cellular dynamics.

Exploring how protein-protein interactions change the dynamics of cellular signal transduction network is at the heart of computational systems biology and is the purpose of our research as well.

OUTLINE OF RESEARCH

The goal of the research presented in this dissertation was to develop a mathematical modeling method using the Hill function and the Markov Chain Monte Carlo methods, and use this method to model EGFR signal transduction network and model PITP in EGFR signal transduction network. The research was performed in two stages presented in chapters 4 and 5 following the literature review in Chapter 2 (computational modeling methods) and in Chapter 3 (parameter selection methods):

Modeling Cell Signal Transduction Network Using Hill Function and Markov Chain Monte Carlo Methods (Chapter 4). The purpose of this study was to develop a mathematical modeling method using the Hill function and the Markov Chain Monte Carlo methods, and model epidermal growth factor receptor signal transduction network. Experimental data of response from Erk and Akt was used to fitting model.

Modeling PITP Using Hill Function and Markov Chain Monte Carlo Methods (Chapter 5). The purpose of this study was to model four different PITP structures in EGFR signal transduction network using the Hill function and the Markov Chain Monte Carlo methods, and make predictions on which model is most likely close to experiment. Experimental data of response from Erk and Akt was used to fitting model.

Future direction (Chapter 6).

REFERENCES

Alberts, B. et al. (2008) Molecular Biology of THE CELL, Fifth Edition.

Aldridge B.B. (2006) Physicochemical modeling of cell signaling pathways. *Nat Cell Biol* 8: 1195–1203.

Baker MD, Wolanin PM & Stock JB (2006) Signal transduction in bacterial chemotaxis. *BioEssays* 28:9-22.

Ben-Shlomo I, Yu Hsu S, Rauch R et al. (2003) Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Sci STKE* 187:RE9.

Berridge MJ (2005) Unlocking the secrets of cell signaling. *Annu Rev Physiol* 67:1-21.

Berridge MJ, Bootman MD & Roderick HL (2003) Calcium signaling dynamics, homeostasis and remodeling. *Nature Rev Mol Cell Biol* 4:517-529.

Bourne HR (1995) GTPases: a family of molecular switches and clocks. *Philos Trans R Soc. Lond B Biol Sci* 349:283-289.

Bradshaw RA & Dennis EA (eds) (2003) Handbook of Cell Signaling. *Elsevier*: St. Louis.

Brown, K.S. et al. (2004) The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys. Biol.* 184-195.

Burns ME & Baylor DA (2001) Activation, deactivation, and adaptation in vertebrate photoreceptor cells. *Annu Rev Neurosci* 24:779-805.

Cannon, W.B. (1933) The wisdom of the body.

Carpenter, G. (2000) EGF receptor transactivation mediated by the proteolytic production of EGF-like agonists. *Science STRK* (15), PE1.

Dard N & Peter M (2006) Scaffold proteins in MAP kinase signaling: more than simple passive activating platforms. *BioEssays* 28:146-156.

Dong, C. et al. (2002) MAP kinases in the immune response. *Annu Rev Immunol* 20:55.

Downward J (2004) PI 3-kinase, Akt and cell survival. *Semin Cell Dev Biol* 15:177-182.

Ferrell JE. Jr. (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 14:140-148.

Hill AV (1910) The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *Proc Of The Phys Society*.

Holbro, T. and Hynes, N.E. (2004) ErbB receptors: directing key signaling networks throughout life. *Annu Rev Pharmacol Toxicol* 44: 195–217.

Hood, L. et al. (2004) Systems Biology and New Technologies Enable Predictive and Preventative Medicine**.** *Science* Vol. 306. No. 5696, pp. 640 – 643.

Hudmon A & Schulman H (2002) Structure-function of the multifunctional Ca^{2+}/calmodulin-dependent protein kinase II. *Biochem J* 364:593-611.

Kitano, H. et al. (2000) Fundations of systems biology.

Kolch, W. et al. (2005) When kinases meet mathematics: the systems biology of MAPK signalling. *FEBS Letters* 579: 1891–1895.

Luttrell LM (2006) Transmembrane signaling by Gprotein-coupled receptors. *Methods Mol Bio* 332:3-49.

Mitin N, Rossman KL & Der CJ (2005) Signaling interplay in Ras superfamily function. *Curr Biol* 15:R563-574.

Moghal, N. and Sternberg, P.W. (1999) Multiple positive and negative regulators of signaling by the EGF-receptor. Curr. Opin. Cell Biol. 11, 190-196.

Mullschleger S, Leowith R & Hall MN (2006) OR signaling in growth and metabolism. *Cell* 124:471-484.

Papin JA, Hunter T, Palsson BO & Subramaniam S (2005) Reconstruction of cellular signaling networks and analysis of their properties. *Nature Rev Mol Cell Biol* 6:99-111.

Parker PJ (2004) The ubiquitous phosphoinositides. *Biochem Soc Trans* 32:893-898.

Pawson T (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* 116:191-203.

Pawson T & Scott JD (2005) Protein phosphoryation in signaling – 50 years and counting. *Trends Biochem Sci* 30:286-290.

Pierce KL, Premont RT & Lefkowitz RJ (2002) Seven-transmembrane receptors. *Nature Rev Mol Cell Biol* 3:639-650.

Pires-daSilva A & Sommer RJ (2003) The evolution of signaling pathways in animal development. *Nature Rev Genet* 4:39-49.

Reddy, C.C. et al. (1994) Proliferative response of fibroblasts expressing internalization-deficient epidermal growth factor (EGF) receptors is altered via differential EGF depletion effect. *Biotechnol. Prog.* 10, 377-384.

Reiter E & Lefkowitz RJ (2006) GRKs and beta-arrestine roles in receptor silencing, trafficking and signaling. *Trends Endocrino Metab* 17:159-165.

Rhee SG (2001) Regulation of phosphoinositide-specific phospholipase C. *Annu Rev Biochem* 70:281-312.

Robishaw JD & Berlot CH (2004) Translating G protein subunit diversiy into functional specificity. *Curr Opin Cell Biol* 16:206-209.

Roskoski R Jr (2004) Src protein-tyrosine kinase structure and regulation. *Biochem Biophys Res Commun* 324:1155-1164.

Qi M & Elion EA (2005) MAP kinase pathways. *J Cell Sci* 118:3569-3572.

Sahin M, Greer PL, Lin MZ et al (2005) Eph-dependent tyrosine phosphorylation of ephexin1 modulates growth cone collapse. *Neuron* 46:191-204.

Schlessinger J (2000) Cell signaling by receptor tyrosine kinases. *Cell* 103:211-225.

Schwartz MA & Madhani HD (2004) Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. *Annu Rev Genet* 38:725-748.

Science's Signal Transduction Knowledge Environment (Stke): www.stke.org

Seet BT, Kikic I, Zhou MM & Pawson T (2006) Reading protein modifications with interaction domains. *Nature Rev Mol Cell Biol* 7:473-483.

Shaw RJ & Cantley IC (2006) Ras, PI(3)K and mTOR signaling controls tumour cell growth. *Nature* 44:424-430.

Shaywitz AJ & Greenberg ME (1999) CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annu Rev Biochem* 68:821-861.

Singla V & Reiter JF (2006) The primary cilium as the cell's antenna signaling at a sensory organelle. *Science* 313:629-633.

van der Geer, P. et al. (1994) Receptor proteintyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell Biol.* 10, 251–337.

Wassarman DA, Therrien M & Rubin GM (1995) The ras signaling pathway in Drosophila. *Curr Opin Genet Dev* 5:44-50.

Wells, A. (1999) EGF receptor. *Int. J. Biochem. Chell Biol.* 31, 637-643.

Wiley, H.S. et al. (2003) Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol* 13: 43–50.

Chapter Two **Computational Modeling Methods**

INTRODUCTION

Computational models can be very useful tools when they provide experimentally testable predictions. Modeling complex biological networks presents challenges. Models can generate hypothesis, predict and explain outcomes of novel experiments and guide experimental design. There are different types of modeling that can be theoretically exploration and data driven. Choosing the right tool is very important in modeling. The complexity of the biological problem dictates the scale and detail of the model to be considered. There is a variety of computational modeling methods for understanding signaling networks at the systems level. When modeling cellular signaling networks, a balance needs to be considered between details, ease of simulation and interpretation. The level of details is constrained by availability of experimental data. Models based on kinetics of chemical reactions most likely contain a large number of parameters, and the rate constants for many parameters are unknown because of the difficulty of experimentally determining individual rate constants [Brown et al. 2004]. It would be even more difficult when the signaling network involves a large number of protein-protein interactions. A network of N proteins it will consist of N protein concentrations, on average about 5N interactions per protein [Grigoriev, 2003], and the reaction-rate parameters for each connection of binding (two parameter) and enzyme-catalytic (three parameter) reactions. As recently a large scale signaling and proteomic data available, computational modeling can be useful for analyzing quantitative data especially for systems biology.

When considering the spectrum of computational modeling methods, from less detail to more detail, they may be specified as Boolean modeling method, coarse-grained modeling method, ordinary differential equation modeling method, and statistical mechanics method. Each method has its advantages and limitations. Here we introduce several computational modeling methods with examples, and compare their advantages and limitations. The choice of method or combination of methods for computational modeling depends on the biological questions to be addressed at the forefront and the measurement, definition and function of the networks [Janes et al. 2006]. The detail contained in the model is only that needed to capture and predict the biological questions.

BOOLEAN METHOD

Boolean method is a qualitative approach using Boolean (logic) function to simplify model with discrete time. Input and output are binary (1 or 0) values. A "1" means ON or present or active and "0" signifies OFF or absent or inactive.

As early as 1969, Kauffman used Boolean networks to explore the impact of both number of genes, *N*, and connectivity of the network, *K*, on the cell cycle time and number of possible cell types [Kauffman, 1969]. There are some applicable cases of modeling biology using Boolean method. For example, many biological phenomena are ultimately Boolean, especially for cell cycle and development studies [de Jong, 2002]. A cell is switched from mass growth (G1 phase) to DNA replication (S phase) during cell division cycle, a cell is either alive or dead, and cells often have to make binary decisions of committing to apoptosis or not.

One advantage of computational modeling using Boolean method might be to understand the function and behavior of the systems at a much larger scale instead of details of rate constants. With the availability of high-throughput genomic measurements, we may be able to apply reverse engineering to map genetic networks inside cells. Investigation of the generic properties of network models is the utmost interest. Boolean networks have received much attention for several decades in these contexts, such as yeast transcriptional network [Kauffman et al. 2003]. These networks consist of nodes, representing genes and proteins, connected by directed edges, representing gene regulation. Kauffman and others modeled intercellular signaling networks using Boolean method to explore how communications among cells influence the genetic network dynamics in tissue simulations [Kauffman et al. 2004].

The limitation of Boolean method obviously is that information and accuracy are lost if the details are the biological questions to be addressed. If a phenomenon crucially depends on the continuous nature of the system, Boolean method is not appropriate.

COARSE-GRAINED METHOD

Coarse-grained method is a semi-quantitative method. It is currently mechanistically the most advanced method of modeling [Vayttaden, S.J. et al. 2004]. The coarse-grained models are at the level of biochemical representation of cellular signaling and functions. A large amount of data typically reduces to three model quantities: reaction schemes, rate constants, and concentrations of molecules. In more advanced modeling, the localization of molecules and their transport rates may be represented. The coarse-grained models can

identify the structure of the signaling pathway qualitatively as well as the dynamic behavior of the network quantitatively under many different conditions.

The advantage of the coarse-grained method is that predictions from the modeling often help in refining theory and improving our understanding of a system. MAPK oscillatory model by Kholodenko is an example [Kholodenko, 2000]. This model made predictions about the behavior of the MAPK biochemical pathway and showed that the negative feedback loop in the MAPK cascade allows the system to undergo sustained oscillations. Some recent examples of oscillations in the MAPK system [Akhthar et al. 2002; Duffield et al. 2002] have shown that some genes encoding components of the MAPK signaling pathway do show the oscillatory peaks. Although results from the examples do not match the model predictions in the details of time-course and amplitude, the model still provides specific and testable predictions as the basis for better understanding of signaling events.

The limitation of coarse-grained modeling is that the predictions from the models are for a range of parameter values instead of precise parameter values. Therefore, a coarse-grained method may not be suitable for detailed modeling if precise quantitative results are expected.

ORDINARY DIFFERENTIAL EQUATION METHOD

Ordinary differential equation (ODE) is an equation containing derivatives with respect to a single experimental variable, such as time. ODE modeling method is highly specified and dependent on prior knowledge of rate constants derived from experimental

data. The focus is on the detailed kinetics of the molecular interactions formalized on the basis of mass action kinetics.

The advantage of ODE modeling method is that the models capture temporal and/or spatial dynamics at the level of individual reactions [Aldridge et al. 2006; Hlavacek et al. 2006; Levchenko et al. 2000; Chakraborty et al. 2003; Wiley et al. 2003]. For example, the ODE method was used to model the ErbB (ErbB receptors are epidermal growth factor receptor family) signaling network with all four ErbB receptors and two ligands, epidermal growth factor (EGF) and heregulin (HRG). The model simulation analysis led to new insights as to how the ErbB signaling network functions in MCF-7 cancer cells [Birtwistle et al. 2007]. There is a strong point of model validation from the ODE modeling when there is an agreement between the model prediction and an experimental data set.

The limitation or difficulty of ODE modeling method is the necessity of obtaining rate constants from experimental measurements for individual steps in the network, which is especially difficult for highly interacted dynamic signaling networks. For the same example, the ErbB signaling network that has the ligand-binding domain, the dimerization site, the kinase domain, and 10 phosphorylation sites requires more than 10^6 differential equations. This phenomenon, referred as ‘combinatorial complexity’, is a fundamental problem in developing mechanistic, differential equation models of signal transduction networks [Godstein et al. 2004; Blinov et al. 2006]. With their simplified schematic representation of the model structure, the ODE model still consists of 117 species, 235 parameters, and 96 (net) reactions [Birtwistle et al. 2007]. It would take 3 processor-years to simulate 135 candidate sets from this ODE model, which is prohibitively long simulation time.

OTHER RELATED METHODS

Another modeling method, Bayesian statistics, is modeling of causal relationships among random variables using conditional probabilities to associate correlations and influences between network components. The advantage is that the mechanistic information may be incomplete but the interacting biochemical species is informative in the network [Aldridge et al. 2009]. The limitation of Bayesian method is that the model cannot construct cyclical networks.

CONCLUSION

Computational modeling is a tool to improve our understanding of complex biological systems. Sometimes a combination of the methods may be a better choice. For example, dividing a network into functional modules, a detailed method may be applied within the functional module, and a Boolean method may be applied between modules to switch on or off from one module to a neighboring module.

REFERENCES

Akhtar R.A. et al. (2002) Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. *Curr Biol* 12: 540–550.

Aldridge, BB et al. (2006) Physicochemical modeling of cell signaling pathways. *Nat Cell Biol* 8: 1195–1203.

Aldridge, BB et al. (2009) Fuzzy Logic Analysis of Kinase Pathway Crosstalk in TNF/EGF/Insulin-Induced Signaling. *PLoS Comput Biol* 5(4): e1000340.

Birtwistle, MR et al. (2007) Lignad-dependent responses of the ErbB signaling network: experimental and modeling analysis. *Mol. Sys. Biol.* 3:144.

Blinov, ML et al. (2006) A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems* 83:136–151

Brown, K.S. et al. (2004) The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys. Biol.* 184-195.

Chakraborty, AK et al (2003) In silico models for cellular and molecular immunology: successes, promises and challenges. *Nat Immunol* 4: 933–936.

de Jong, H. (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 9(1): 67–103.

Duffield, G. E. et al. (2002) Circadian programs of transcriptional activation, signaling, and protein turnover revealed by microarray analysis of mammalian cells. *Curr. Biol*. 2002, 12, 551 –557.

Goldstein B, et al. (2004) Mathematical and computational models of immune-receptor signalling. *Nat Rev Immunol* 4:445–456.

Grigoriev, A. (2003) On the number of protein-protein interactions in the yeast proteome. *Nucleic Acids Res.* 31:4157-4161.

Ideker, T. and Lauffenburger, DA (2003) Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol*, 21:255-262.

Janes, KA and Lauffenburger DA (2006) A biological approach to computational models of proteomic networks. *Curr. Opin. in Chem Biol.* 10:73–80.

Kauffman, SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, 22:437-467.

Kauffman, S.A. (2004) *Proc. Natl.Acad. Sci. USA* 101, 17102–17107.
Kholodenko, B.N. (2000) *Eur. J. Biochem.* 267, 1583 -1588.

Levchenko A, Bruck J, Sternberg PW (2000) Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc Natl Acad Sci USA* 97: 5818–5823.

Resat, H et al. (2003) An Intergrated Model of Epidermal Growth Factor Receptor Trafficking and Signal Transduction. *Biophysical Journal* Vol. 85 730-743.

Vayttaden, S.J. et al. (2004) A spectrum of models of signaling pathways. *ChemBioChem*, 5:1365-1374.
Hlavacek, WS et al. (2006) Rules for modeling signal-transduction systems. *Sci STKE*: re6.

Wiley, HS et al (2003) Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol* 13: 43–50.

Chapter Three **Parameter Selection Methods**

INTRODUCTION

In cell signaling modeling, there are relatively limited data of stoichiometry and kinetics of the biochemical reactions about species in the signaling network available. Analysis of dynamic properties is limited to a few well characterized pathways. The classical method is to gather rate constants from experimental data. There are several concerns when gathering information in this fashion. First, it is difficult to measure the rate constants for a large network which has a large parameter set. Secondly, these data are collected from different laboratories, and, therefore, they may not be consistent. When the parameter information is not available, as the case is with our model, we must generate parameter values for our model data fitting. There are several methods for parameter selection in mathematical modeling. Two parameter selection methods are Least Squares Method, and Markov Chain Monte Carlo methods.

LEAST SQUARES METHOD

The least squares method minimizes the sum of squared differences between the values of dependent variable and model predicted value. This method was first described by Karl Friedrich Gauss around 1794 [Bretscher, 2005].

Chen et al. [Chen et al. 1991] used Orthogonal Least Squares method for parameter selection of radial basis function centers one by one in a rational way until an adequate network had been constructed. The radial basis function is an alternative to the two-layer neural network signal processing. This approach provided a simple and efficient method for fitting the radial basis function networks. The researcher concluded that Orthogonal Least

Squares method is far superior to a random selection of centers in the radial basis function network.

MARKOV CHAIN MONTE CARLO METHODS

Markov Chain Monte Carlo (MCMC) methods simulate direct draws from some complex, nonstandard multivariate distributions of interest [Chib et al. 1995].

Markov Chain is a sequence of random variables generated by the Markov process which is defined by its transition probabilities. These transition probabilities between different values in the sample space depend only on the random variable's current state. Thus the only information about the past that predicts the future is the current state of the random variable. Knowledge of the earlier state of the random variable does not change the transition probability.

For example [Walsh 2004], assume the state space is rainy, sunny, and cloudy. Then weather follows a Markov process. The probability of tomorrow's weather depends only on today's weather and not the weather on any previous days. The observation that it has rained for three days does not alter the probability that tomorrow's weather will be sunny. Suppose the transition probability given that today is rainy:

P (rain tomorrow | rain today) = 0.5, read as the probability of rain tomorrow is 0.5 given that today it is rainy.

P (sunny tomorrow | rain today) = 0.25, read as the probability of that tomorrow will be sunny is 0.25 given that today it is rainy.

P (cloudy tomorrow | rain today) = 0.25, read as the probability of that tomorrow will be is 0.25 given that today it is rainy.

Monte Carlo sampling uses Bayesian inference that is based on random sampling. It was introduced by Metropolis [Metropolis et al. 1953] at Los Alamos National Laboratory. Some systems cannot be computed exactly, and, therefore they are predicted on the basis of approximation. Examples are the van der Vaals equations for dense gases, and Boltzmann equations for dilute gases, among many others.

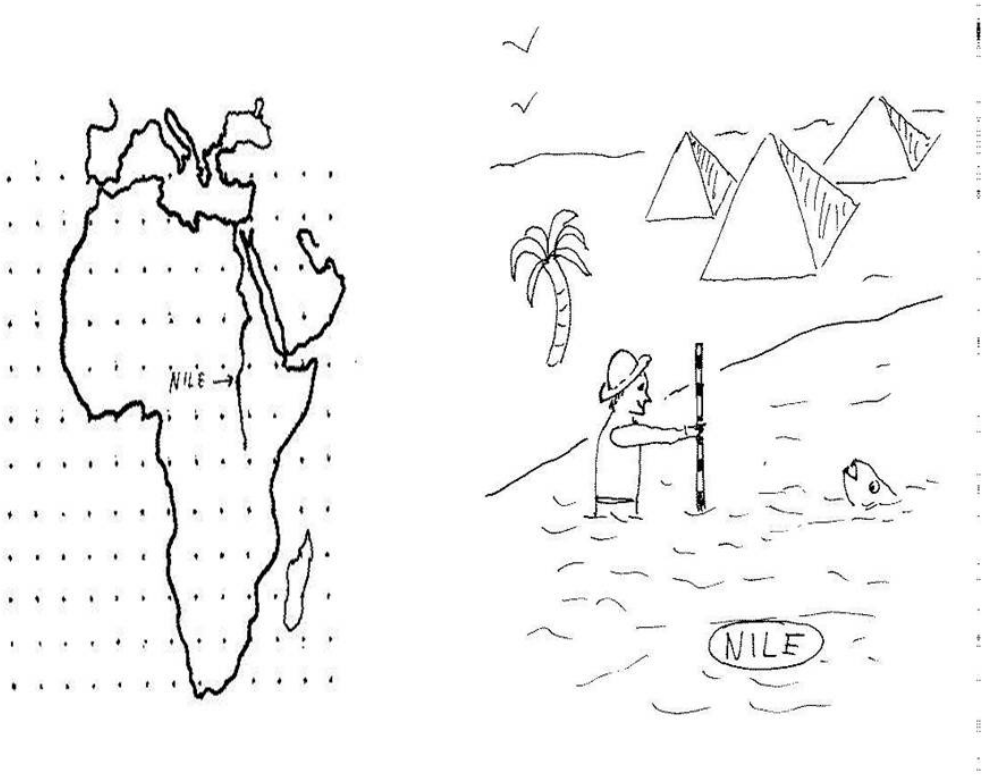

Figure 1. Measuring the depth of the Nile: a comparison of conventional quadrature (left), with the Metropolis scheme (right) [Daan Frenkel, 2004].

Figure 1 illustrates a comparison of two ways of measuring the depth of the river Nile. On the left of side of the figure, depth is measured by conventional quadrature by which

a predetermined set of points is measured, understanding that some points maybe located outside of the water. The right side of the figure, illustrates measuring the depth by Metropolis sampling. This method involves the construction of an importance-weighted random walk. In this random walk, a trial is rejected if the measurement is not in the water and it is accepted if the measurement is in the water. The average of all these measurements yields an estimate of the average depth of the Nile [Daan Frenkel, 2004].

The Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm was introduced by Metropolis and extended by Hastings. Hastings (Hastings, 1970) generalized the Metropolis algorithm by using an arbitrary transition probability function and setting the acceptance probability for a candidate point [Walsh, 2004]. Physicists attempted to compute complex integrals by expressing them as expectations for some distribution and estimating the expectation by drawing samples from that distribution.

The transition probability is a conditional distribution function that represents the probability of moving from the current point to a future point for the same parameter.

The Metropolis-Hastings algorithm has the following steps [Walsh, 2004]:

1. Start with any initial value x_0 satisfying $f(x_0) > 0$.
2. Use current value x_0, sample a random number z from some distribution (uniform, normal), and a standard deviation α for fine tuning, the candidate distribution $y = x_0 + \alpha * z$.
3. Calculate the ratio of the density of candidate y and current x_0 points.
4. According to predefined criteria, accept and set $x_0 = y$ or reject candidate y, and return to step 2.

Markov chain theory suggests that such a chain will eventually converge to a stationary or equilibrium distribution which is the target or posterior distribution. If we sample long enough, we will eventually be sampling from the posterior distribution itself after we discard some sets of samples from burn-in period. The burn-in period is the number of runs before the chain approaches stationarity or equilibrium. Determining the number of runs is a key issue in the successful implementation of Metropolis-Hastings or other MCMC methods. Typically the first 1000 to 5000 elements are thrown out, some applications may have more elements to be thrown out. The various convergence tests are used to assess whether stationary or equilibrium has been reached. The choice of starting values can greatly increase or decrease the required burn-in time. One suggestion is to start the chain as close to the center of the distribution as possible such as using an approximate Maximum Likelihood Estimator as the starting value [Walsh, B. 2004].

Choosing a proposed distribution is another issue in the implementation of the Metropolis-Hastings algorithm. There are two general approaches: random walk and independent chain sampling. The proposed distribution based on a *random walk chain* requires that the new value y equals the current value x plus a random variable z, a standard deviation α for fine tuning.

$$y = x + \alpha * z$$

Density of random variable z is assumed to be symmetric with mean zero. Variance of the proposal distribution is a *tuning parameter* that can be adjusted to get better mixing.

The proposal distribution based on *independent chain* assumes that candidate y is drawn from a distribution of interest and that it is independent of the current value. This proposal distribution is generally not symmetric.

Convergence Diagnostics

Time series trace plot is one type of the tests for convergence. The plot is the number of iterations versus the number of sets of random variables being selected. The trace plot can show the evidence of well mixing or poor mixing [Walsh, 2004]. In a well mixing chain, the chain explores the parameter space and the time series looks like well spaced and has no long flat space (figure 2). In a poor mixing chain, it seems the chain stays in small regions of the parameter space for a long period of time or trapped and there are long flat periods in the time series corresponding to the random variables being rejected.

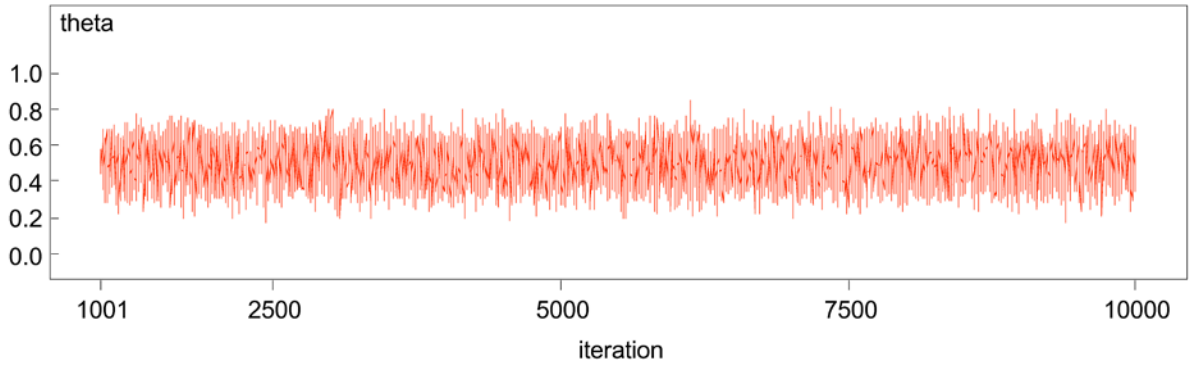

Figure 2. The MCMC chain of 9000 draws from the posterior distribution [Wetzels et al. 2009].

Studies by Roberts (Roberts et al. 1994) showed, that if the target and proposal densities are normal, the scale of the random walk proposal density should be tuned so that

the acceptance rate is approximately 45% in one dimensional problems and approximately 23% as the number of dimensions approaches infinity, with the optimal acceptance rate being around 25% for as few as six dimensions. In statistics, the number of parameters is often referred to as the dimension [Bretscher 2005].

CONCLUSION

Markov Chain Monte Carlo methods are useful methods for parameter selection, and Metropolis-Hastings algorithm is a powerful Markov Chain method to simulate multivariate distributions. Since there is limited amount of information of biochemical reactions available for modeling cell signaling, MCMC increasingly will be used as a parameter selection method.

Issues in simulation of a single long chain [Geyer 1992, Raftery et al. 1992b], or multiple chains each starting from different initial values [Gelman et al. 1992], the initial value, the sample size, and how long the burn-in period should be, that are currently under active research.

REFERENCES

Bretscher, Otto (2005). *Linear Algebra with Applications, 3rd ed.* Upper Saddle River NJ: Prentice Hall.

Chen, S., Cowen, C.F.N, and Grant, P.M. (1991). *IEEE TRANSACTIONS ON NEURAL NETWORKS,* Vol. 2, No. 2.

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings Algorithm. *The American Statistician, Vol.* 49, No. 4, 327-335.

Frenkel, D. (2004) Introduction to Monte Carlo Methods. *NIC Series, Vol. 23*, 29-60.

Gelman, A., and D. B. Rubin. (1992). Inferences from iterative simulation using multiple sequences (with discussion). *Statistical Science* 7: 457 - 511.

Geyer, C. J. 1992. Practical Markov chain Monte Carlo (with discussion). *Stat. Sci.* 7: 473–511.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* 57:97-109.

Metropolis, N. et al (1953) Equations of state calculations by fast computing machines. *J. of Chemical Physics* 21:1087-1091.

Raftery, A. E., and S. Lewis. 1992b. Comment: One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo. *Stat. Sci.* 7: 493–497.

Roberts, G.O. et al (1994) Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *Technical Report,* University of Cambridge.

Walsh, B. (2004) Markov Chain Monte Carlo and Gibbs Sampling. *Lecture Notes for EEB 581*.

Wetzels, R. et al. (2009) Bayesian Inference Using WBDev: A Tutorial for Social Scientists.

Chapter Four **Modeling Cell Signal Transduction Network Using Hill Function and Markov Chain Monte Carlo Methods**

INTRODUCTION

In this study, we developed a new mathematical modeling method to understand the dynamic behavior of Epidermal Growth Factor (EGF) Receptor signal transduction network at the system level. We constructed the network as a signaling regulatory network where there is a lack of biochemical reactions information but the direction of signaling information is known. The readouts of Erk and Akt from experiments are used to fit the model to the experimental data. We used the Hill function to formulate ordinary differential equations (ODE), and the Markov Chain Monte Carlo method for parameter selection. Given the difficulty of experimentally testing the predictions that some proteins may not be experimentally measurable at current technology, and the constraints speed of simulation that combinatorial models may not be feasible to simulate, building a simple model that can explain complex biological processes is an ultimate goal.

Modeling EGFR Signal Transduction Network

Epidermal growth factor (EGF) regulates cell growth, differentiation, proliferation and survival. The EGF receptor (EGFR) is a useful test case for modeling cell signaling networks because it is present in different cell types. Also, antibodies and reagents are available for testing [Wiley et al. 2003]. Modeling complex cell signaling network presents challenges. Models based on the kinetics of chemical reactions most likely contain a large number of parameters and the rate constants for many parameters are unknown due to the difficulty of measuring them experimentally [Brown et al. 2004]. It would be even more

difficult where a large number of protein-protein interactions are present in a signaling network. The most intensively studied signaling pathway is the Mitogen Activated Protein Kinase (MAPK) pathway.

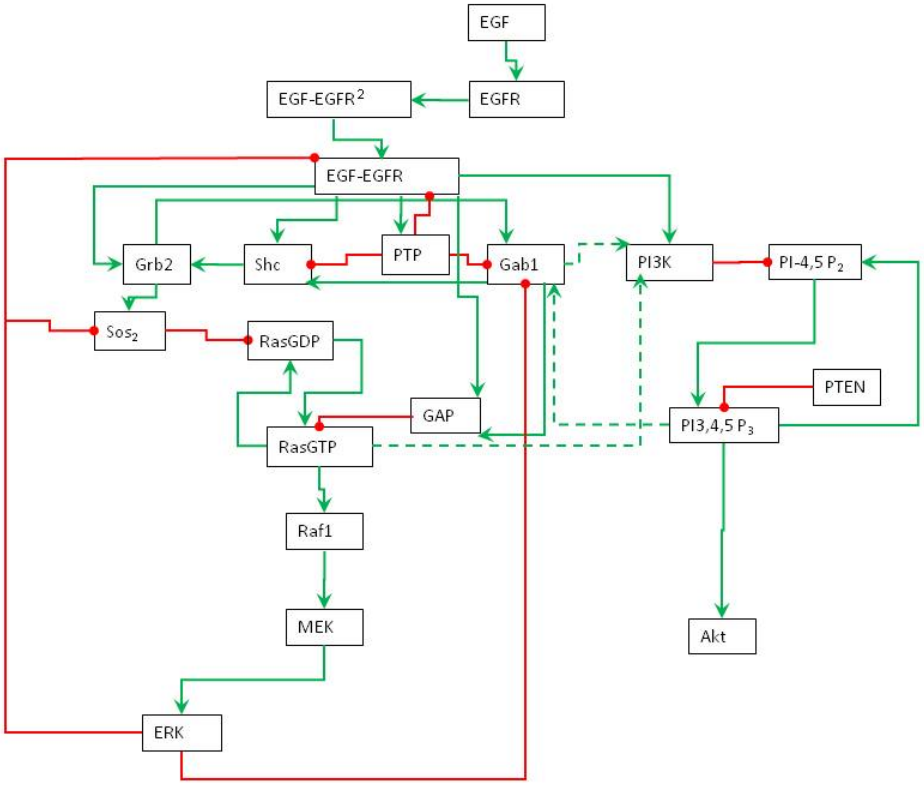

Figure 1: Schematic representation of EGFR signaling transduction network. Green line is activation, red line is inhibition, and dashed line is cross pathways.

The EGFR signal transduction network is constructed as simplified model identified by experimental data in the literature [Birtwistle et al. 2007]. Molecules are species in the EGFR signal transduction network, and weights are interactions among species. The dynamics weights reflect the system behavior of species in the network. Initial values are set

to zero for some species and set to non-zero values for other species. The non-zero initial values are approximated by educated guess (Table 1). The model diagram (Figure 1) describes the regulatory functions and interactions among the species in the signaling network.

Table 1: Protein initial values

	Species	Initial values
1	EGF	0.5
2	EGFR	0.7
3	RasGDP	0.9
4	PTEN	0.5
5	PIP_2	0.1

In EGFR signaling transduction network, EGF is signaling molecule (ligand) and EGFR is plasma membrane receptor. When EGF binds to EGFR, the EGFR is activated and in turn it activates the intracellular signaling pathways. Two activated EGFR form an EGFR dimer. Once a receptor dimer is formed, it gains tyrosine kinase activity and can auto-phosphorylate on several tyrosine residues. The homodimer undergoes rapid EGF-induced internalization and degradation through a multistep process.

There are two major signaling pathways in the network, MAPK pathway and PI3K pathway. The MAPK pathway is involved in cell proliferation. In this pathway activated EGFR activates the adaptor layer consisting of Grb_2, Shc, PTP and Gab1. The function of the adaptor layer is to connect the receptor and the downstream pathway. PTP is an inhibitor which inhibits EGFR, Shc, and Gab_1. Grb_2 activates Sos_2 which stimulates the inactive Ras protein to replace its bound GDP by GTP. Then GAP inhibits GTP to inactive Ras protein by replacing its bound GTP with GDP. RasGTP phosphorylates downstream Raf_1 (MapKKK),

and Raf_1 activates Mek (MapKK), in turn Mek activates Erk (MAPK). RasGTP also phosphorylates PI3K in the PI3K pathway. Erk has negative feedback loop to inhibit Sos_2 and EGFR. Since Ras is an oncogene, the MAPK pathway is an important research subject in cancer research.

The PI3K pathway is involved in cell survival. In this pathway activated EGFR activates PI3K, in turn PI3K produces PIP_3 by inhibition on (consume) $PIP_{2.}$ PTEN inhibits PIP_3 and PIP_3 has a positive feedback loop to activate $PIP_{2.}$ PIP_3 activates downstream Akt. The PI3K pathway is important in cancer research because PI3K and Akt are oncogenes and PTEN is a tumor suppressor. We chose to fit the data response from Akt in this pathway because an Akt readout can be obtained experimentally.

RESULTS AND DISCUSSION

Model Formulation

In this study, we modeled EGFR signaling network as a regulatory model when biochemical kinetics are not available. The model can provide insight on network signaling qualitatively. There are 20 species, 68 parameters and 30 biochemical reactions in the model. Hill function was used to formulate ODE equations with time scaling coefficient and a decay term.

The circuit of the EGFR signaling network (Table 2) was translated from the EGFR signaling network diagram. The downstream species listed in the second column of Table 2 are the species that receive the activation and/or inhibition from regulators which are upsteam species in the pathway listed in the third column. The plus sign corresponds to the green line in the diagram and indicates activation, and the minus sign corresponds to the red line in the diagram and indicates inhibition. Since EGF and PTEN do not receive any activation or inhibition from other species, they are treated as constants.

Table 2: Circuits of the EGFR signaling network, plus sign indicates activation received, minus sign indicates inhibition received. EGF and PTEN receive neither activation nor inhibition from other species, they are treated as constants.

	Downstream species	Regulators
1	EGF	
2	EGFR	+ EGF
3	Dimer	+ 2*EGFR
4	Phos. dimer	+ Dimer – Erk
5	Grb_2	+ Phos. dimer + Shc
6	Shc	+ Phos. dimer + Gab_1 – PTP
7	PTP	+ Phos. dimer
8	Gab_1	+ Grb_2 – PTP – Erk + PIP_3
9	Sos_2	+ Grb_2 – Erk
10	RasGDP	+ RasGTP – Sos_2
11	RasGTP	+ RasGDP – RasGAP
12	RasGAP	+ Phos. dimer + Gab_1
13	Raf_1	+ RasGTP
14	Mek	+ Raf_1
15	Erk	+ Mek
16	PI3K	+ Phos. dimer + Gab_1 +RasGTP
17	PIP_2	+ PI3K + PIP_3
18	PIP_3	+ PIP_2 – PTEN
19	PTEN	
20	Akt	+ PIP_3

Model Equations

There are 20 species in the EGFR signaling network. However, there are only 18 equations translated from the circuit of the EGFR signaling network, since EGF and PTEN do not have any inputs. Therefore EGF and PTEN are treated as constants. In the following equations, r is time scaling coefficient, α is a constant, n is the Hill coefficient, $W_{activation}$ is the sum of the activation inputs and $W_{inhibition}$ is the sum of the inhibition inputs the species received from other species in the signaling network respectively. Each equation has a decay term. For

example, in equation 2 the negative EGFR term is a decay term and accounts for dissociation, degradation, and internalization.

EGF: dy(1) = 0.

EGFR: dy(2) = r2*(((α*$W_{activation}$) ^{n2})/(1+ (α*$W_{activation}$) ^{n2}) – EGFR).

Dimer: dy(3) = r3*(((α * $W_{activation}$) ^{n3})/(1 + (α * $W_{activation}$) ^{n3}) – dimer).

Phos.dimer: dy(4)=r4*(((α*$W_{activation}$) ^{n4})/(1+ (α*$W_{activation}$) ^{n4} +(α*$W_{inhibition}$) ^{n4}) – Pd).

Grb: dy(5) = r5*(((α * $W_{activation}$) ^{n5})/ (1 + (α*$W_{activation}$) ^{n5}) – Grb).

Shc: dy(6) = r6*(((α *$W_{activation}$) ^{n6})/ (1+(α*$W_{activation}$) ^{n6} +(α *$W_{inhibition}$) ^{n6}) – Shc).

PTP: dy(7) = r7*(((α * $W_{activation}$) ^{n7}) / (1 + (α*$W_{activation}$) ^{n7}) – PTP).

Gab: dy(8) = r8*(((α*$W_{activation}$) ^{n8})/(1 + (α*$W_{activation}$) ^{n8} + (α*$W_{inhibition}$) ^{n8}) – Gab).

Sos: dy(9) = r9*(((α *$W_{activation}$) ^{n9})/(1+(α *$W_{activation}$) ^{n9} + (α *$W_{inhibition}$) ^{n9}) – Sos).

RasGDP: dy(10)=r10*(((α*$W_{activation}$) ^{n10})/(1+(α*$W_{activation}$) ^{n10}+(α*$W_{inhibition}$) ^{n10}) –GDP).

RasGTP: dy(11) = r11(*((α*$W_{activation}$) ^{n11})/(1+(α*$W_{activation}$) ^{n11}+(α*$W_{inhibitionn}$) ^{n11}) – GTP).

RasGAP: dy(12) = r12*(((α* $W_{activation}$) ^{n12}) / (1 + (α* $W_{activation}$) ^{n12}) – GAP).

Raf: dy(13) = r13*(((α* $W_{activation}$) ^{n13}) / (1 + (α* $W_{activation}$) ^{n13}) – Raf).

Mek: dy(14) = r14*(((α* $W_{activation}$) ^{n14}) / (1 + (α* $W_{activation}$) ^{n14}) – Mek).

Erk: dy(15) = r15*(((α * $W_{activation}$) ^{n15}) / (1 + (α*$W_{activation}$) ^{n15}) – Erk).

PI3K: dy(16) = r16*(((α * $W_{activation}$) ^{n16}) / (1 + (α*$W_{activation}$) ^{n16}) – PI3K).

PIP_2: dy(17)=r17*(((α*$W_{activation}$) ^{n17})/(1+(α*$W_{activation}$) ^{n17}+(α*$W_{inhibitionn}$) ^{n17}) – PIP_2).

PIP_3: dy(18)=r18*(((α*$W_{activation}$) ^{n18})/(1+(α*$W_{activation}$) ^{n18}+(α*$W_{inhibitionn}$) ^{n18}) – PIP_3).

PTEN: dy(19) = 0.

Akt: dy(20) = r20*(((α* $W_{activation}$) ^{n20}) / (1 + (α* $W_{activation}$) ^{n20}) – Akt).

Model Fitting

As in many cell signaling models, there are limited data pertaining to biochemical reactions. That is the case for our model. Consequently, we used the Markov Chain Monte Carlo method to select parameter values, as described in the Methods section (Markov Chain Monte Carlo Method). We used data points from experimental data to fit Erk and Akt response in the model. After bringing the model response to the target region, 10,000 runs were simulated. Among the 10,000 sets, around 25% of sets met the criteria of experimental data points and were selected. The final result is an average of these selected parameter values.

Figure 2 shows the average EGF-stimulated response for all the parameter sets. The star is data points from experiment (from Beak). The top plot is response from Erk and the bottom plot is from Akt. The stars are data points from experiment, the plot is from simulation. The Y-axis is the normalized activity of concentration and the X-axis is the simulation time of 0 to 1800 seconds.

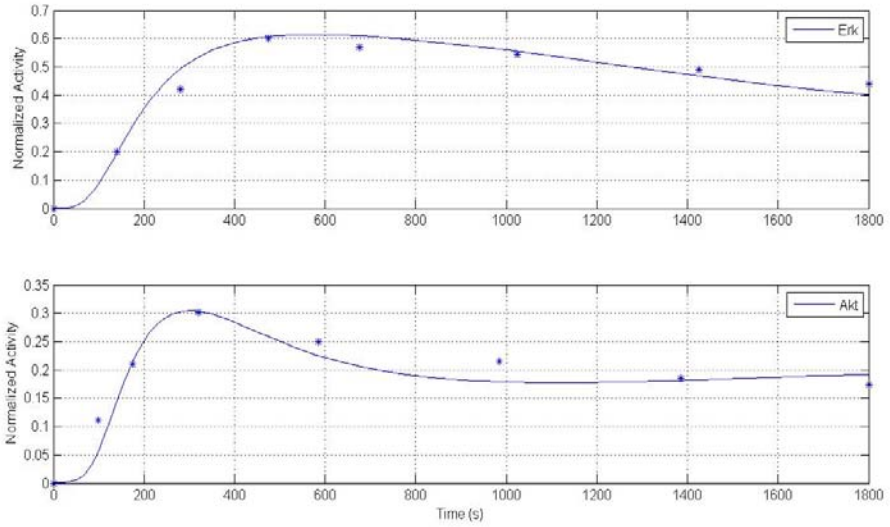

Figure 2: EGF-stimulated response averaged from selected parameter sets. Stars are data points from experiment. Stars are data points from experiment.

Perturbation on Erk and Akt Activation

In dose response to EGF, Erk activation is robust at high ligand dose and Akt activation is not. For Erk, set EGF to 0.9, 0.5 and 0.01, Erk activation does not show great impact in response to EGF perturbation. For Akt, using the same dose, Akt activation shows significant change in response to EGF perturbation. This difference in response to perturbation indicates that Erk is robust to parameter variation. This robustness in Erk most likely is due to saturation. Another possibility of this robustness in Erk may be due to the multiple feedback loops from Erk to EGFR and Sos_2, and Akt has no such feedback loop. The same evidence was shown by [Birtwistle, M.R. et al. 2007] in their ErbB signaling network model. As the researcher concluded that from control theory, it is well known that negative feedback loops provide a system with robustness to disturbances [Ogunnaike, B.A. et al. 1994, Freeman, 2000]. Bistability investigation is necessary to confirm this possibility.

Statistical Data Analysis of Parameter Convergence

To see if the posterior or target distribution has converged, the posterior distribution density and time series trace plots are analyzed by convergence tests. Figure 3 shows that the posterior distribution has converged. The Y-axis is the number of values from different parameter sets, and the X-axis is the weight. Notice that some parameters converged well and others have long tails. These parameter convergence data may provide insights to the signaling network as discussed below. The posterior distribution is not symmetric because negative concentration has no meaning in cell signaling.

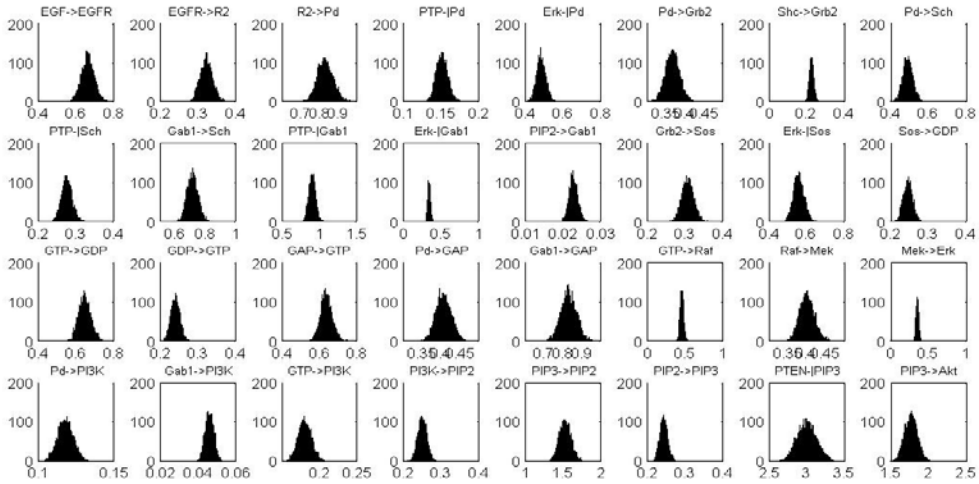

Figure 3: Histograms of posterior distribution showing parameter selection convergence.

Figure 4 is a time series trace to show that the posterior distribution is converged with a well mixing and has no long flat periods. The Y-axis is the weight, and the X-axis is the number of iterations. Black is selected parameter sets, and white is rejected parameter sets. All parameters were tested with time series trace for convergence. Here only show three parameters for demonstration.

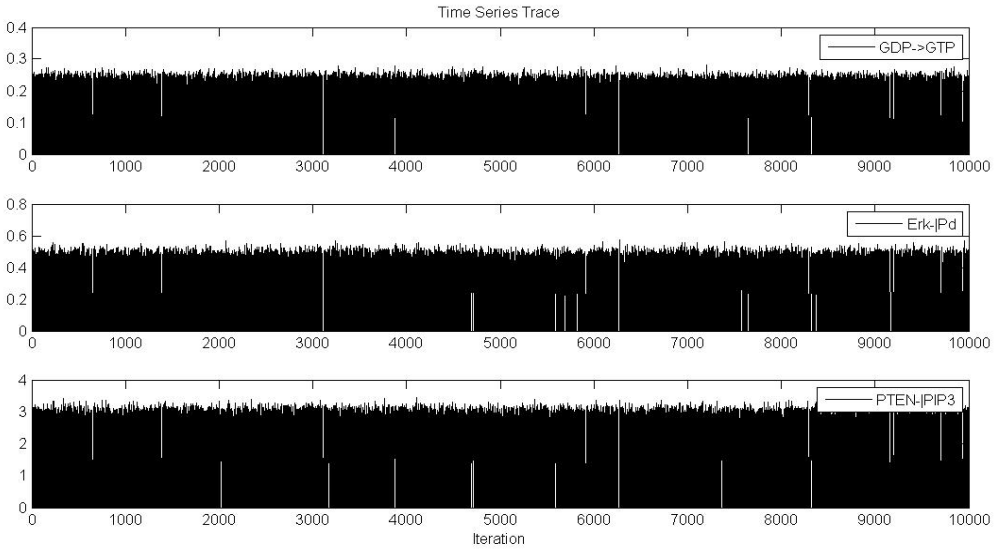

Figure 4: Time series trace of posterior distribution showing that the selected parameter values are well mixed.

Table 3 shows the maxima, minima, means and standard deviations of weights in the EGFR signaling network. The simulation results are not unique, and are averaged from parameter sets selected from simulations. The standard deviation measures the variability of the data. The smaller standard deviation indicates less variability and stable, the larger standard deviation indicates more variability and has large range of values. From statistical data analysis, the standard deviation of weights may provide insights about the regulatory species in the signaling network. Parameters with larger standard deviations in comparison to other parameters in the network may suggest that those species might have the regulatory characteristics that these species are robust regulatory species in the network. Because these species can have wider range of weights, they are considered robust in cell signaling network.

Table 3: Weight parameter sets selected from simulations. Maxima, minima, means and standard deviations of parameter sets showing the degree of convergence. -> sign signifies activation, and -| sign signifies inhibition.

	Parameter	Max.	Min.	Mean	Std. dev.
1	EGF -> EGFR	0.7871	0.5464	0.6625	0.0335
2	EGFR -> Dimer	0.3804	0.2736	0.3224	0.0159
3	Dimer -> Phos. Dimer	0.9782	0.6981	0.8309	0.0404
4	PTP-| Phos. dimer	0.1835	0.1268	0.1515	0.0075
5	Erk-| Phos. dimer	0.5740	0.4082	0.4850	0.0238
6	Phos. Dimer->Grb_2	0.4630	0.3111	0.3675	0.0183
7	Shc-> Grb_2	0.2651	0.1868	0.2236	0.0110
8	Phos. Dimer->Shc	0.5832	0.4051	0.4885	0.0251
9	PTP-| Shc	0.3267	0.2338	0.2759	0.0137
10	Gab_1 ->Shc	0.8702	0.5989	0.7144	0.0353
11	Phos. Dimer->PTP	0.4426	0.3046	0.3604	0.0180
12	Grb_2 ->Gab_1	0.6015	0.4185	0.5003	0.0250
13	PTP-| Gab_1	1.0753	0.7468	0.9027	0.0450
14	Erk ->Gab_1	0.4039	0.2936	0.3430	0.0172
15	PIP_3 ->Gab_1	0.0273	0.0191	0.0227	0.0011
16	Grb_2 ->Sos_2	0.3643	0.2572	0.3069	0.0153
17	Erk ->Sos_2	0.6732	0.4674	0.5557	0.0281
18	Sos_2 -| RasGDP	0.2876	0.2082	0.2446	0.0120
19	RasGTP->RasGDP	0.7658	0.5349	0.6463	0.0319
20	RasGDP->RasGTP	0.2826	0.2000	0.2409	0.0114
21	RasGAP-| RasGTP	0.7708	0.5411	0.6363	0.0304
22	Phos. Dimer ->RasGAP	0.4913	0.3348	0.4072	0.0201
23	Gab_1-> RasGAP	0.9958	0.6712	0.8293	0.0410
24	RasGTP ->Raf_1	0.5606	0.3766	0.4591	0.0227
25	Raf_1 ->Mek	0.4837	0.3351	0.3999	0.0196
26	Mek ->Erk	0.4040	0.2855	0.3460	0.0171
27	Phos. Dimer ->PI3K	0.1359	0.1016	0.1183	0.0052
28	Gab_1->PI3K	0.0545	0.0361	0.0460	0.0023
29	RasGTP ->PI3K	0.2100	0.1506	0.1806	0.0089
30	PI3K ->PIP_2	0.2944	0.2125	0.2507	0.0120
31	PIP_3 -> PIP_2	1.7879	1.2814	1.5291	0.0759
32	PIP_2 -> PIP_3	0.2821	0.2103	0.2439	0.0103
33	PTEN -| PIP_3	3.4327	2.6055	2.9978	0.1295
34	PIP_3 -> Akt	2.1616	1.5107	1.7639	0.0850

We divided the standard deviations of weights into three categories in Table 3: larger standard deviation with value greater than 0.1 and colored red; smaller standard deviation with value smaller than 0.05 and colored with green; standard deviation with value smaller than 0.1 and greater than 0.05. Parameters with larger standard deviation may suggest stable enzymatic regulation in cell signaling network. In contrast, parameters with smaller standard deviation may suggest robust regulation in cell signaling network. We identified the weight of PTEN -| PIP_3 that had larger standard deviation greater than 0.1 and colored as red. PTEN has inhibition on PIP_3, which negatively regulates downstream species. This property indicates the importance of negative regulation in the cell signaling network. It may also suggest the robustness of these species in the network since these values have a wider range than other weights. We also identified the weight of RasGDP -> RasGTP that had smaller standard deviation less than or equal to 0.01 and colored as green. Ras is converted from GDP, an inactive form to GTP, an active form. This property indicates the importance of stable enzymatic regulation in the cell signaling network.

Model Limitation

This methodology is useful for modeling cell signaling network qualitatively when there is limited experimental data. Like any modeling method, this method has its limitations that it does not provide precise quantitative parameter values.

CONCLUSION

In conclusion, we developed a regulatory model that used EGF as stimulus to activate Erk and Akt in cell signaling transduction network. The model construction is based on the information of the signaling directions and activation or inhibition in the signaling network.

Therefore this methodology provides qualitative insight pertaining to the cell signaling network. We chose Hill function to formulate ODEs because Hill function can determine the cooperativeness of the ligand and enzyme or receptor binding. Thus it provides biological information to the signaling network model. We developed an algorithm to select parameters using the Markov Chain Monte Carlo methods.

MATERIALS AND METHODS

Model Development

Figure 1 depicts the EGFR signal transduction network. Molecules are species and weights are interactions between two species. The dynamics of weights, scaling coefficient and Hill coefficient reflect the system behavior. The circuits of the network are translated into Ordinary Differential Equations (ODE). The initial values are educated guesses from the knowledge we have about the biological process. An example is that we know the concentration of RasGDP should be higher because this inactive form appears in abundance. The concentration is between 0 and 1. The weights are random numbers generated using Matlab [The Mathworks; Natick, MA] and selected by Markov Chain Monte Carlo method. The weights are positive values for both activation and inhibition. There are 20 species and 71 parameters in the EGFR signaling network. The simulations perform a deterministic computation to select all possible combinations of weights, scaling coefficients and Hill coefficients based on the experimental data points from Erk and Akt readout.

Governing Equations

We used Hill function to formulate ODE for each species in the signal transduction network. Since Hill function is used to determine the cooperativity of the ligand binding to the enzyme or receptor, its use has biological meaning in the signaling network modeling.

There are three types of equations: activation, inhibition, and both activation and inhibition.

1. Activation equation:

$$dx/dt = r^* ((\alpha^* W_{activation})^n / (1+ (\alpha^* W_{activation})^n) - x).$$

2. Inhibition equation:

$$dx/dt = r*((\alpha*W_{inhibition})^{n} / (1+ (\alpha*W_{inhibition})^{n}) - x).$$

3. Activation and Inhibition equation:

$$dx/dt = r*((\alpha* W_{activation})^{n} / (1+ (\alpha*W_{activation})^{n} +(\alpha*W_{inhibition})^{n}) - x).$$

x – a species in the signal transduction network. A negative sign in front of x here indicates x is a decay term. The range of concentration is from 0 to 1.

r – time scaling coefficient, ranging from 0 and 1.

α – a constant with a value of eight.

$W_{activation}$ – sum of activation weights is the total activation input a species received from other species in the network. The initial weight is between 0 and 1. It could be greater than 1 as a result of parameter selection.

$W_{inhibition}$ – sum of inhibition weights is the total inhibitory input a species received from other species in the network. The initial weight is between 0 and 1. It could be greater than 1 as a result of parameter selection.

n – Hill function coefficient. The initial Hill coefficient is greater than 1.

Parameter Selection Using Markov Chain Monte Carlo Method

In cell signaling modeling, there are relatively few data of stoichiometry and kinetics of the biochemical reactions. The classical method employed is to gather rate constants from experimental data. There are several concerns when gathering information in this fashion. First, it is difficult to measure the rate constants for a large network which has a large parameter set. Second, these data are collected from different laboratories and, therefore, they

may not be consistent. We used Markov Chain Monte Carlo method and developed an algorithm to efficiently simulate parameter selection in the EGFR signaling network model.

Markov Chain Monte Carlo (MCMC) methods are used to simulate direct draws from some complex, nonstandard multivariate distributions of interest [Chib et al. 1995].

Markov Chain is a sequence of random variables generated by Markov process which is defined by its transition probabilities. These transition probabilities between different values in the sample space depend only on the random variable's current state. Thus the only information about the past used to predict the future is the current state of the random variable. Knowledge about earlier states of the random variable does not change the transition probability.

Monte Carlo sampling uses Bayesian inference that is based on random sampling. It was introduced by Metropolis [Metropolis et al. 1953] at Los Alamos National Laboratory. Some systems cannot be computed exactly and consequently they are predicted on the basis of approximation. Examples are the van der Vaals equations for dense gases, or Boltzmann equations for dilute gases, among many others.

Markov Chain Monte Carlo Simulation

There are different MCMC algorithms depending on the applications. We chose the Metropolis-Hastings algorithm for our simulation. The usual approach is to start with any value and draw the sample space from uniform or normal distribution with specified mean and standard deviation.

Since we do not have any known parameter values for any species in the signaling network, we selected a set of parameters from a uniform distribution to generate Erk and Akt

response as the starting point. Doing so required less burn-in period and more quickly brings the chain to the stationary or equilibrium state.

In computer science, there is an efficient design approach named divide and conquer. This approach breaks a complex problem into sub-problems of similar type and solves sub-problems in an efficient manner. We used this approach in the algorithm design to shorten burn-in period and bring the chain to convergence efficiently. We divided weight, scaling coefficient, and Hill coefficient into blocks with different standard deviations for each block and pathway.

The equation: New = Old *exp (sd*randn).

The algorithm has the following steps:

1. Start with the selected parameter set mentioned above as the starting point.

2. Determine a proposal distribution or a move to the target region. Parameters are divided into different groups according to their characteristics in the cell signal transduction network: pathways, weight, Hill coefficient, and scaling coefficient. Standard deviations are specified for different groups according to their influence on the data fitting.

3. The candidate y is drawn from the process of y = x + sd*z, where x is the current or old value, sd is standard deviation and z is a random increment variable drawn from a normal distribution.

4. The simulation calculates the system of ODEs that are interactions of species in the signaling network.

5. Determine if the response from Erk and Akt is met the criteria of data points. If so, accept the set. Otherwise reject the set and go to step 3.

In our simulation experiments, it may take several trials to bring the model response to the target region. We noticed that by adjusting standard deviation in burn-in period, the burn-in period is shortened. Due to the complexity of this model, the weight, time scaling coefficient, and Hill coefficient have different functions that influence the model fitting. Once the response is in the target region, the parameter sets are selected if they met the criteria of the experimental data points.

ACKNOWLEDGMENTS

The study was supported in part by the North Carolina State University Department of Animal Science Enhancement Fund.

REFERENCES

Alberts, B. et al. (2008) Molecular Biology of THE CELL, Fifth Edition.

Aldridge B.B. (2006) Physicochemical modeling of cell signaling pathways. *Nat Cell Biol* 8: 1195–1203.

Baker MD, Wolanin PM & Stock JB (2006) Signal transduction in bacterial chemotaxis. *BioEssays* 28:9-22.

Ben-Shlomo I, Yu Hsu S, Rauch R et al. (2003) Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Sci STKE* 187:RE9.

Berridge MJ (2005) Unlocking the secrets of cell signaling. *Annu Rev Physiol* 67:1-21.

Berridge MJ, Bootman MD & Roderick HL (2003) Calcium signaling dynamics, homeostasis and remodeling. *Nature Rev Mol Cell Biol* 4:517-529.

Bourne HR (1995) GTPases: a family of molecular switches and clocks. *Philos Trans R Soc. Lond B Biol Sci* 349:283-289.

Bradshaw RA & Dennis EA (eds) (2003) Handbook of Cell Signaling. *Elsevier*: St. Louis.

Brown, K.S. et al. (2004) The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys. Biol.* 184-195.

Birtwistle, MR et al (2007) Lignad-dependent responses of the ErbB signaling network: experimental and modeling analysis. *Mol. Sys. Biol.* 3: 144.

Burns ME & Baylor DA (2001) Activation, deactivation, and adaptation in vertebrate photoreceptor cells. *Annu Rev Neurosci* 24:779-805.

Cannon, W.B. (1933) The wisdom of the body.

Carpenter, G. (2000) EGF receptor transactivation mediated by the proteolytic production of EGF-like agonists. *Science STRK* (15), PE1.

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings Algorithm. *The American Statistician, Vol.* 49, No. 4, 327-335.

Dard N & Peter M (2006) Scaffold proteins in MAP kinase signaling: more than simple passive activating platforms. *BioEssays* 28:146-156.

Dong, C. et al. (2002) MAP kinases in the immune response. *Annu Rev Immunol* 20:55.

Downward J (2004) PI 3-kinase, Akt and cell survival. *Semin Cell Dev Biol* 15:177-182.

Ferrell JE. Jr. (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 14:140-148.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* 57:97-109.

Hill AV (1910) The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *Proc Of The Phys Society*.

Holbro, T. and Hynes, N.E. (2004) ErbB receptors: directing key signaling networks throughout life. *Annu Rev Pharmacol Toxicol* 44: 195–217.

Hood, L. et al. (2004) Systems Biology and New Technologies Enable Predictive and Preventative Medicine**.** *Science* Vol. 306. No. 5696, pp. 640 – 643.

Hudmon A & Schulman H (2002) Structure-function of the multifunctional Ca^{2+}/calmodulin-dependent protein kinase II. *Biochem J* 364:593-611.

Kitano, H. et al. (2000) Fundations of systems biology.

Kolch, W. et al. (2005) When kinases meet mathematics: the systems biology of MAPK signalling. *FEBS Letters* 579: 1891–1895.

Luttrell LM (2006) Transmembrane signaling by Gprotein-coupled receptors. *Methods Mol Bio* 332:3-49.

Metropolis, N. et al (1953) Equations of state calculations by fast computing machines. *J. of Chemical Physics* 21:1087-1091.

Mitin N, Rossman KL & Der CJ (2005) Signaling interplay in Ras superfamily function. *Curr Biol* 15:R563-574.

Moghal, N. and Sternberg, P.W. (1999) Multiple positive and negative regulators of signaling by the EGF-receptor. Curr. Opin. Cell Biol. 11, 190-196.

Mullschleger S, Leowith R & Hall MN (2006) OR signaling in growth and metabolism. *Cell* 124:471-484.

Ogunnaike, B.A., Ray, W.H. (1994) Process Dynamics, Modeling, and Control. New York: Oxford University Press.

Papin JA, Hunter T, Palsson BO & Subramaniam S (2005) Reconstruction of cellular signaling networks and analysis of their properties. *Nature Rev Mol Cell Biol* 6:99-111.

Parker PJ (2004) The ubiquitous phosphoinositides. *Biochem Soc Trans* 32:893-898.

Pawson T (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* 116:191-203.

Pawson T & Scott JD (2005) Protein phosphoryation in signaling – 50 years and counting. *Trends Biochem Sci* 30:286-290.

Pierce KL, Premont RT & Lefkowitz RJ (2002) Seven-transmembrane receptors. *Nature Rev Mol Cell Biol* 3:639-650.

Pires-daSilva A & Sommer RJ (2003) The evolution of signaling pathways in animal development. *Nature Rev Genet* 4:39-49.

Reddy, C.C. et al. (1994) Proliferative response of fibroblasts expressing internalization-deficient epidermal growth factor (EGF) receptors is altered via differential EGF depletion effect. *Biotechnol. Prog.* 10, 377-384.

Reiter E & Lefkowitz RJ (2006) GRKs and beta-arrestine roles in receptor silencing, trafficking and signaling. *Trends Endocrino Metab* 17:159-165.

Rhee SG (2001) Regulation of phosphoinositide-specific phospholipase C. *Annu Rev Biochem* 70:281-312.

Robishaw JD & Berlot CH (2004) Translating G protein subunit diversiy into functional specificity. *Curr Opin Cell Biol* 16:206-209.

Roskoski R Jr (2004) Src protein-tyrosine kinase structure and regulation. *Biochem Biophys Res Commun* 324:1155-1164.

Qi M & Elion EA (2005) MAP kinase pathways. *J Cell Sci* 118:3569-3572.

Sahin M, Greer PL, Lin MZ et al (2005) Eph-dependent tyrosine phosphorylation of ephexin1 modulates growth cone collapse. *Neuron* 46:191-204.

Schlessinger J (2000) Cell signaling by receptor tyrosine kinases. *Cell* 103:211-225.

Schwartz MA & Madhani HD (2004) Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. *Annu Rev Genet* 38:725-748.

Science's Signal Transduction Knowledge Environment (Stke): www.stke.org

Seet BT, Kikic I, Zhou MM & Pawson T (2006) Reading protein modifications with interaction domains. *Nature Rev Mol Cell Biol* 7:473-483.

Shaw RJ & Cantley IC (2006) Ras, PI(3)K and mTOR signaling controls tumour cell growth. *Nature* 44:424-430.

Shaywitz AJ & Greenberg ME (1999) CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annu Rev Biochem* 68:821-861.

Singla V & Reiter JF (2006) The primary cilium as the cell's antenna signaling at a sensory organelle. *Science* 313:629-633.

van der Geer, P. et al. (1994) Receptor proteintyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell Biol.* 10, 251–337.

Walsh, B. (2004) Markov Chain Monte Carlo and Gibbs Sampling. *Lecture Notes for EEB 581*.

Wassarman DA, Therrien M & Rubin GM (1995) The ras signaling pathway in Drosophila. *Curr Opin Genet Dev* 5:44-50.

Wells, A. (1999) EGF receptor. *Int. J. Biochem. Chell Biol.* 31, 637-643.

Wiley, H.S. et al. (2003) Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol* 13: 43–50.

Chapter Five **Modeling PITP Using Hill Function and Markov Chain Monte Carlo Methods**

INTRODUCTION

This study modeled cell signaling network to understand specifically the important role of PITP in the Epidermal Growth Factor (EGF) Receptor signal transduction network at the system level. We constructed the network as a signaling regulatory network. There is a lack of information on biochemical reactions but the directions of signaling information is known. Experimental data of response from Erk and Akt were used to fit the model. The methodology used for this model was previously developed using the Hill function to formulate ordinary differential equations (ODE) and the Markov Chain Monte Carlo method for parameter selection. Concerning the difficulty of experimental testing the predictions that some proteins may not be experimentally measurable at current technology, and the constraints speed of simulation that combinatorial models may not be feasible to simulate, building a simple model that can explain complex biological processes is our ultimate goal.

Modeling PITP in EGFR Signal Transduction Network

Epidermal growth factor (EGF) regulates cell growth, differentiation, proliferation and survival. The EGF receptor (EGFR) is a useful test case for modeling cell signaling networks because it is present in different cell types. Additionally, antibodies and reagents are available for testing [Wiley et al. 2003]. Modeling complex cell signaling network presents challenges. Models based on the kinetics of chemical reactions most likely contain a large number of parameters and the rate constants of many parameters are unknown due to the difficulty of measuring them experimentally [Brown et al. 2004]. It would be even more

difficult where a large number of protein-protein interactions are present in a signaling network.

The EGFR signal transduction network is constructed as a simplified model identified by experimental data in the literature [Birtwistle et al. 2007]. We combined the adaptor layer together with Sos_2 to simplify modeling and to focus on how PITP influences the signaling network behavior. PITP was added in the network. Molecules are species in the EGFR signal transduction network. Weights are interactions among species. The dynamics weights reflect the system behavior of species in the network. The model diagrams describe the regulatory functions and interactions among the species in the signaling network.

Phosphoinositide transfer protein (PITP) is a critical regulator of phosphoinositides in cellular compartments, signal transduction, and membrane traffic. PITP can bind and exchange one molecule of phosphatidylinositol (PI) and facilitate the transfer of these lipids between different membrane compartments. PITPα is expressed ubiquitously in all tissues and is very abundant in the brain. It is detected widely throughout the entire developing central nervous system and in almost all neurons in an adult brain. Dysfunction of PITP may lead to neurodegeneration diseases [Hsuan et al. 2001]. PtdIns 4-OH kinase (PI4K) is an enzyme that converts PI to PIP_2. Suppressor of actin (SAC) is an enzyme that converts PIP_2 back to PI [Liu et al. 2010].

In an EGFR signaling transduction network, EGF is a signaling molecule (ligand) and EGFR is a plasma membrane receptor. When EGF binds to EGFR, the EGFR is activated and in turn it activates the intracellular signaling pathways.

There are three major signaling pathways in the network, PI3K pathway, MAPK pathway, and PLC_r pathway. The MAPK pathway is involved in cell proliferation. In this pathway activated EGFR activates Sos_2 which stimulates the inactive Ras protein to replace its bound GDP by GTP. Then GAP inhibits GTP to inactive Ras protein by replacing its bound GTP with GDP. RasGTP phosphorylates downstream Raf_1 (MapKKK). Raf_1 then activates Mek (MapKK), and Mek activates Erk (MAPK). RasGTP also phosphorylates PI3K in the PI3K pathway. Erk has a negative feedback loop to inhibit Sos_2 and EGFR. Since Ras and Erk are oncogenes, MAPK pathway is important in cancer research.

The PI3K pathway is involved in cell survival. In this pathway activated EGFR activates PI3K. In turn PI3K produces PIP_3 by inhibition (consume) on phosphatidylinositol 4,5-biphosphate (PIP_2). PTEN inhibits PIP_3 and PIP_3 has a positive feedback loop to convert back to PIP_2. PIP_2 is a phosphorylated inositol phospholipid that is present in small amounts in the inner half of the plasma membrane lipid layer. It is produced by phosphorylation of phosphatidylinositol (PI). PIP_3 activates downstream Akt. PITPα is a transfer protein that stimulates PI4K. PI4K inhibits (consumes) PI to produce PIP_2. SAC is an enzyme that inhibits (consumes) PIP_2 and converts it back to PI. The PI3K pathway is important in cancer research because PI3K and Akt are oncogenes and phosphatase and tensin homolog (PTEN) is a tumor suppressor. Experimental data of response from Akt is used to fit the model.

The PLC_r pathway is involved in cell proliferation. In this pathway activated EGFR activates phospholipase C (PLC_r). PITPα stimulates PI4K. PI4K inhibits (consumes) PI to produce PIP_2. SAC is an enzyme that inhibits (consumes) PIP_2 to convert it back to PI. PITPα was identified as an essential component in ensuring substrate supply to PLC [Hsuan et al.

2001]. PLC_r is a plasma-membrane-bound enzyme that cleaves PIP_2. PIP_2 is then broken down in a signaling response and generates two intracellular mediators, IP_3 and DAG. Inositol 1,4,5-trisphosphate (IP_3) diffuses through the cytosol and release Ca^{2+} from the endoplasmic reticulum (ER) by binding to and opening IP_3 - gated Ca^{2+} - releases channels (IP_3 receptors) in the ER membrane. Diacylglycerol (DAG) is embedded in the plasma membrane.

Since the structure of how PITP is connected to other species in the signaling network is not well understood, we constructed four different models designated A, B, C, and D. We present four different model structures here and point out the differences among them in terms of PITP connection to its immediate downstream species. Simulation results presented here are focused on Model B and Model D because our predictions of PITP structure are Model B and Model D. The reason is that PITP activates PI3K pathway and PLC_r pathway in both Model B and Model D. The simulation results indicate an close agreement with PITP experimental data (not published) qualitatively. Results of Model A and Model C are in the supplement.

Model A (Figure 1) PITP is activated by EGFR, and in turn activates PI4Ka. PI4Ka inhibits (consumes) PIa, and PIa produces PIP_2a. SACa inhibits (consumes) PIP_2a, and PIP_2a is feedback to PIa.

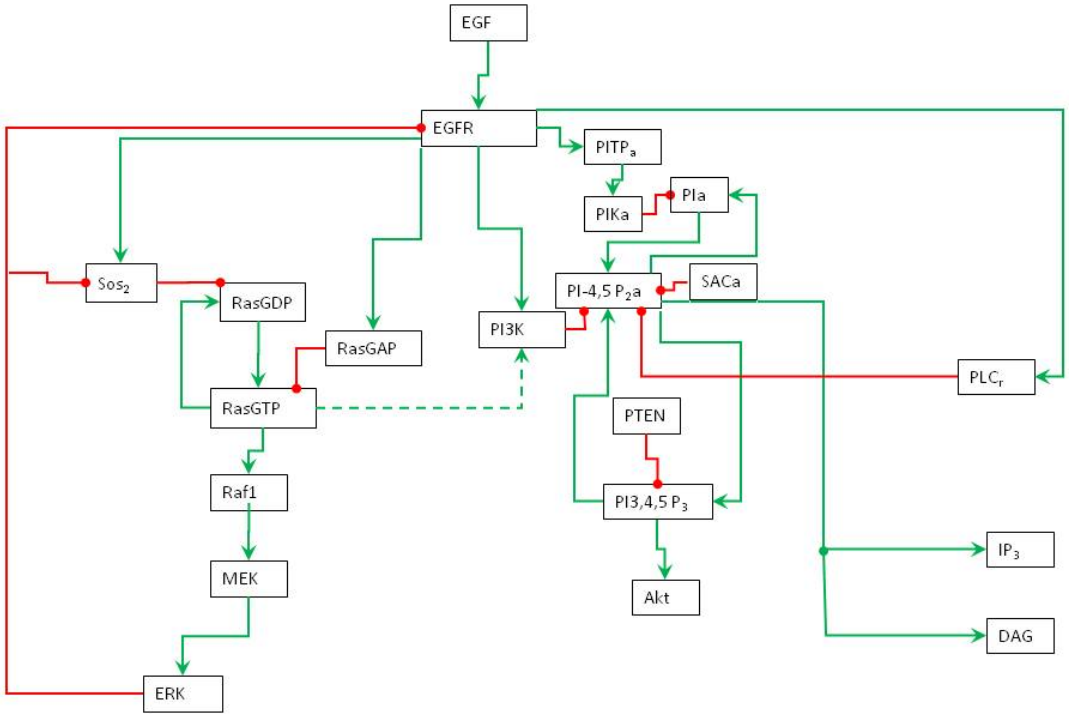

Figure 1: PITP Model A, schematic representation of EGFR signaling transduction network. Green line represents activation, red line is inhibition, and dashed line indicates cross pathways.

Model B (Figure 2), contains a negative feedback loop from PIP_2a to EGFR. This feedback loop distinguishes Model B from Model A.

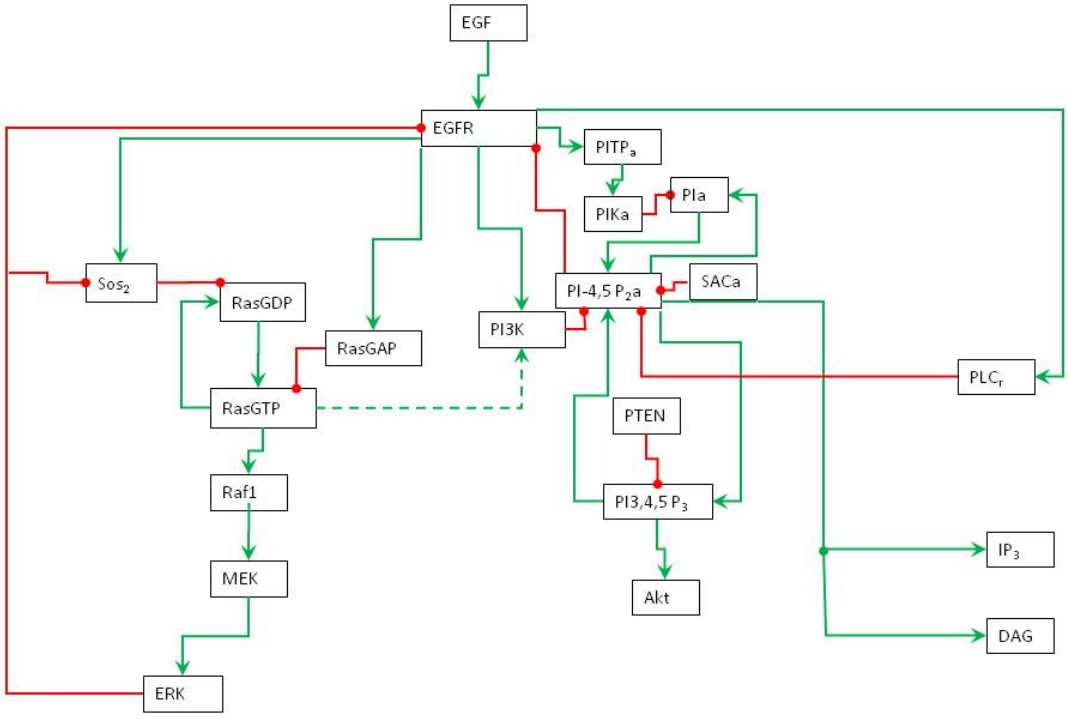

Figure 2: PITP model B, schematic representation of EGFR signaling transduction network. Green line indicates activation, red line is inhibition, and dashed line represents cross pathways.

Model C (Figure 3) includes two PIP_2 pools, PIP_2a and PIP_2b. PIP_2a is downstream of $PITP_a$ and independent of PI3K pathway. There is a negative feedback loop from PIP_2a to EGFR. PIP_2b is downstream of PI3K and PLC_r. PI4Kb inhibits (consumes) PIb, and PIb makes PIP_2b. SACb inhibits (consumes) PIP_2b, and PIP_2b is feedback to PIb. PI3K inhibits (consumes) PIP_2b, and makes PIP_3. PLC_r inhibits (cleaves) PIP_2b and generates two second messengers IP_3 and DAG.

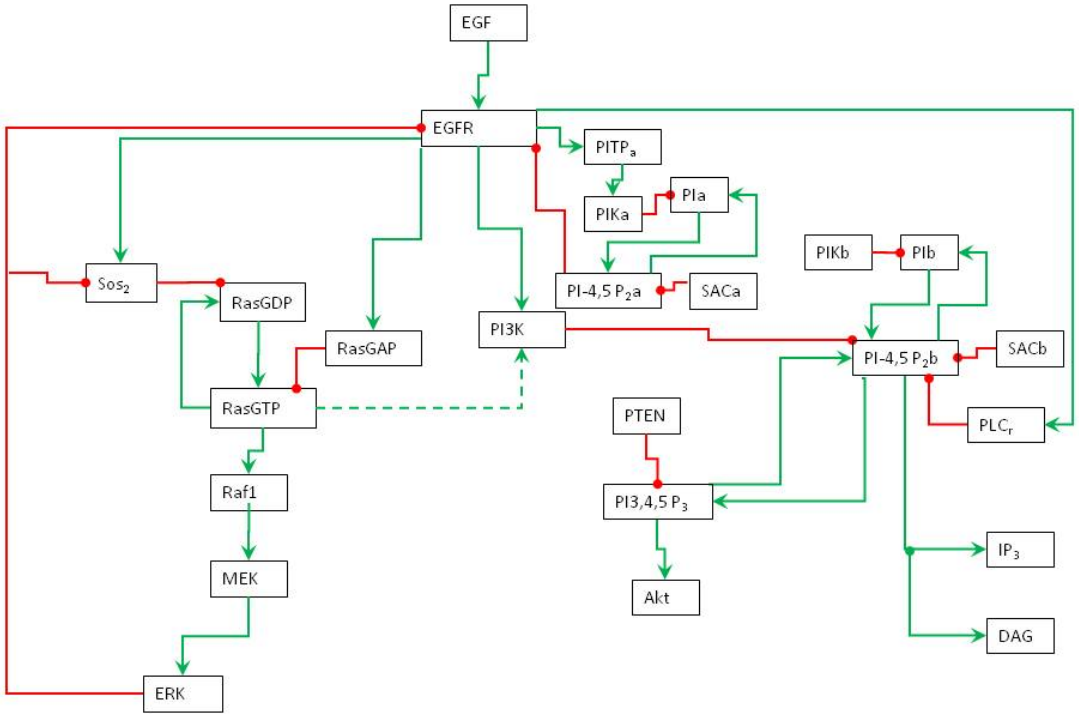

Figure 3: PITP Model C, schematic representation of EGFR signaling transduction network. Green line indicates activation, red line signifies inhibition, and dashed line represents cross pathways. There are two PIP_2 in this model, PIP_2a and PIP_2b.

Model D (Figure 4) in contrast to Mode C, the two PIP_2 are in two different pathways. PIP_2a is downstream of PI3K pathway and PIP_2b is downstream of PLC_r.

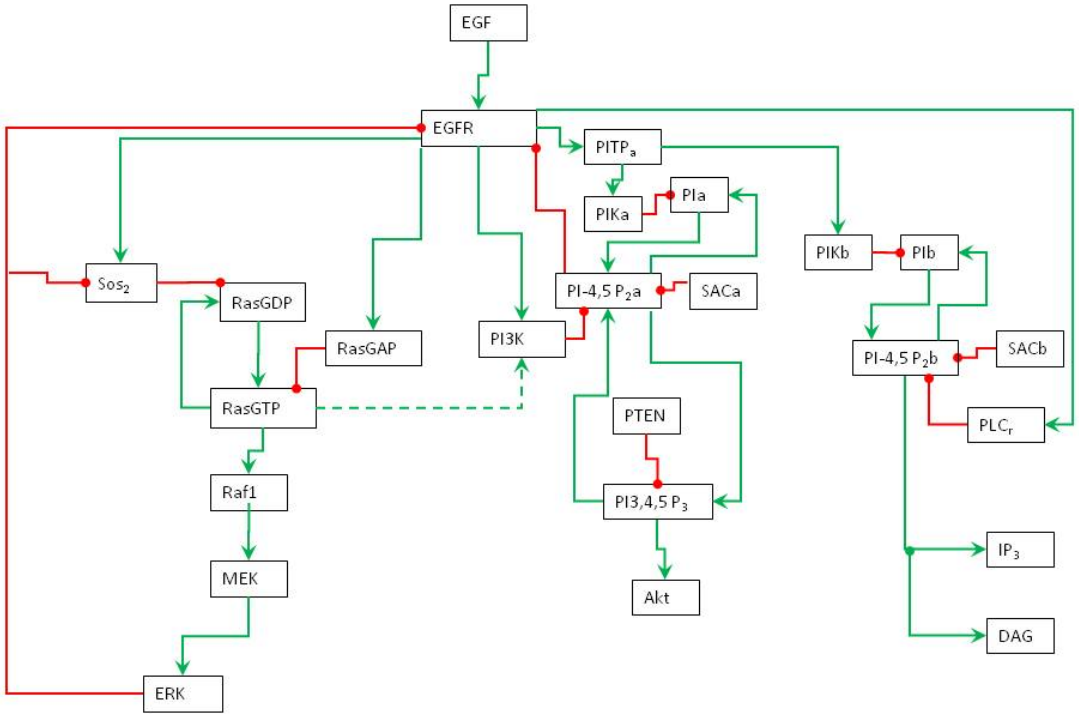

Figure 4: PITP Model D, schematic representation of EGFR signaling transduction network. Green line indicates activation, red line signifies inhibition, and dashed line represents cross pathways. There are two PIP_2 in this model, PIP_2a and PIP_2b. PIP_2a is downstream of PI3K, and PIP_2b is downstream of PLC_r.

RESULTS AND DISCUSSION

Model Formulation

This study modeled EGFR signaling network as a regulatory model when biochemical kinetics are not available. The model can provide qualitative insight on network signaling. The Hill function was used to formulate the ODE equations with time scaling coefficient and a decay term.

In Model A, there are 21 species, 68 parameters and 28 biochemical reactions in the model. The circuit of the EGFR signaling network (see supplement) was translated from the EGFR signaling network diagram Model A (Figure 1). Downstream species listed in the second column are the species that receive activations and/or inhibitions from regulators which are upsteam species listed in the third column. The plus sign corresponds to green line in the diagram and indicates activation. The minus sign corresponds to red line in the diagram and indicates inhibition. Since EGF, PTEN, and SACa do not receive any activation or inhibition from other species in the network, they are treated as constants.

Model B, includes 21 species, 69 parameters and 29 biochemical reactions. The circuit of the EGFR signaling network (Table 1) was translated from the EGFR signaling network diagram Model B (Figure 2). Downstream species listed in the second column are the species that receive activation and/or inhibition from regulators which are upsteam species listed on the third column. The plus sign corresponds to green line in the diagram and indicates activation. The minus sign refers to the red line in the diagram and indicates inhibition. Since EGF, PTEN, and SACa do not receive any activation or inhibition from other

species in the network, they are treated as constants. The only difference between Model B and Model A is that Model B includes a negative feedback loop from PIP_2a to EGFR.

Table 1: Model B**,** Circuits of the EGFR signaling network. Plus sign indicates activation received, and minus sign indicates inhibition received. EGF, PTEN, and SACa receive neither activation nor inhibition from other species, and are treated as constants. The difference between Model B and Model A is that Model B contains a negative feedback loop from PIP_2a to EGFR.

	Downstream species	Regulators
1	EGF	
2	EGFR	+ EGF – PIP_2a – Erk
3	Sos_2	+ EGFR – Erk
4	RasGDP	+ RasGTP – Sos_2
5	RasGTP	+ RasGDP – RasGAP
6	RasGAP	+ EGFR
7	Raf_1	+ RasGTP
8	Mek	+ Raf_1
9	Erk	+ Mek
10	PI3K	+ EGFR + RasGTP
11	PIP_2a	+ PIa + PIP_3 – PI3K – PLC_r – SACa
12	PIP_3	+ PIP_2a – PTEN
13	PTEN	
14	Akt	+ PIP_3
15	PITP	+ EGFR
16	PLC_r	+ EGFR
17	IP_3	+ PIP_2a
18	DAG	+ PIP_2a
19	PIKa	+ PITP
20	PIa	+ PIP_2a - PIKa
21	SACa	

Model C, involves 25 species, 70 parameters and 33 biochemical reactions. The circuit of the EGFR signaling network (see supplement) was translated from the EGFR signaling network diagram for Model C (Figure 3). The downstream species listed in the second column are the species receive activation and/or inhibition from regulators which are

upsteam species listed on the third columns. The plus sign corresponds to green line in the diagram and indicates activation. The minus sign corresponds to red line in the diagram and indicates inhibition. Since EGF, PTEN, PIKb, SACa and SACb do not receive any activation or inhibition from other species in the network, they are treated as constants.

Model D is composed of 25 species, 70 parameters and 33 biochemical reactions. The circuit of the EGFR signaling network (Table 2) was translated from the EGFR signaling network diagram for Model D (Figure 4). The downstream species listed in the second column are the species receive activation and/or inhibition from regulators which are upsteam species listed in the third column. The plus sign corresponds to green line in the diagram and indicates activation. The minus sign corresponds to red line in the diagram and indicates inhibition. Since EGF, PTEN, SACa and SACb do not receive any activation or inhibition from other species in the network, they are treated as constants.

Table 2: Model D, Circuits of the EGFR signaling network. Plus sign indicates activation received, and minus sign indicates inhibition received. EGF, PTEN, SACa and SACb receive neither activation nor inhibition from other species, and they are treated as constants.

	Downstream species	Regulators
1	EGF	
2	EGFR	+ EGF – PIP_2a – Erk
3	Sos_2	+ EGFR – Erk
4	RasGDP	+ RasGTP – Sos_2
5	RasGTP	+ RasGDP – RasGAP
6	RasGAP	+ EGFR
7	Raf_1	+ RasGTP
8	Mek	+ Raf_1
9	Erk	+ Mek
10	PI3K	+ EGFR + RasGTP
11	PIP_2a	+ PIa + PIP_3 – PI3K – SACa
12	PIP_3	+ PIP_2a – PTEN
13	PTEN	
14	Akt	+ PIP_3
15	PITP	+ EGFR
16	PLC_r	+ EGFR
17	IP_3	+ PIP_2b
18	DAG	+ PIP_2b
19	PI4Ka	+ PITP
20	PIa	+ PIP_2a – PI4Ka
21	SACa	
22	PIP_2b	+ PIb – PLC_r – SACb
23	PI4Kb	+ PITP
24	PIb	+ PIP2b – PI4Kb
25	SACb	

Model Equations

The differential equation comprises two parts, activation or inhibition function, plus a term for decay. $W_{activation}$ is the sum of activation inputs and $W_{inhibition}$ is the sum of inhibition inputs the species received from other species in the signaling network, *r* is time scaling coefficient, α is a constant, and *n* is Hill coefficient. Each equation has a decay term, for

example, negative EGFR term in Equation 2 is the decay term and it accounts for dissociation, degradation, and internalization.

Model A (see supplement)

Model B

There are 21 species in the EGFR signaling network but only 18 differential equations are translated from the circuit of the EGFR signaling network. The reason is that EGF, PTEN, and SACa do not have any inputs. Therefore they are treated as constants. The only difference between Model B and Model A is that Model B contains a negative feedback loop from PIP_2 to EGFR and as reflected in the circuit of Model B in Table 2.

EGF: dy(1) = 0.

EGFR: $dy(2)= r2*(((\alpha *W_{activation})^{n2})/(1+ (\alpha*W_{activation})^{n2}+(\alpha*W_{inhibition})^{n2}) – EGFR)$.

Sos_2: $dy(3) = r3*(((\alpha * W_{activation})^{n3})/(1 + (\alpha * W_{activation})^{n3} +(\alpha*W_{inhibition})^{n3}) – Sos_2)$.

RasGDP: $dy(4)=r4*(((\alpha*W_{activation})^{n4})/(1+ (\alpha*W_{activation})^{n4} +(\alpha*W_{inhibition})^{n4}) –GDP)$.

RasGTP: $dy(5)=r5*(((\alpha*W_{activation})^{n5})/(1+(\alpha*W_{activation})^{n5}+(\alpha*W_{inhibition})^{n5})–GTP)$.

RasGAP: $dy(6) = r6*(((\alpha *W_{activation})^{n6})/ (1+(\alpha*W_{activation})^{n6}) – GAP)$.

Raf_1: $dy(7) = r7*(((\alpha * W_{activation})^{n7}) / (1 + (\alpha*W_{activation})^{n7}) – Raf_1)$.

Mek: $dy(8) = r8*(((\alpha*W_{activation})^{n8})/(1 + (\alpha*W_{activation})^{n8}) – Mek)$.

Erk: $dy(9) = r9*(((\alpha *W_{activation})^{n9})/(1+(\alpha *W_{activation})^{n9}) – Erk)$.

PI3K: $dy(10) = r10*(((\alpha*W_{activation})^{n10})/(1+(\alpha*W_{activation})^{n10}) –PI3K)$.

PIP_2: $dy(11)=r11*(((\alpha *W_{activation})^{n11})/(1+(\alpha*W_{activation})^{n11}+(\alpha*W_{inhibition})^{n11}) – PIP_2)$.

PIP_3: $dy(12) = r12*(((\alpha*W_{activation})^{n12})/(1+(\alpha*W_{activation})^{n12}+(\alpha*W_{inhibition})^{n12}) – PIP_3)$.

PTEN: dy(13) = 0.

Akt: $dy(14) = r14*(((\alpha* W_{activation})^{n14}) / (1 + (\alpha* W_{activation})^{n14}) - Akt)$.

PITP: $dy(15) = r15*(((\alpha * W_{activation})^{n15}) / (1 + (\alpha*W_{activation})^{n15}) - Erk)$.

PLC_r: $dy(16) = r16*(((\alpha * W_{activation})^{n16}) / (1 + (\alpha*W_{activation})^{n16}) - PI3K)$.

IP_3: $dy(17) = r17*(((\alpha *W_{activation})^{n17})/(1+(\alpha*W_{activation})^{n17}) - IP_3)$.

DAG: $dy(18) = r18*(((\alpha*W_{activation})^{n18})/(1+(\alpha*W_{activation})^{n18}) - DAG)$.

PI4Ka: $dy(19) = r19*(((\alpha*W_{activation})^{n19})/(1+(\alpha*W_{activation})^{n19}) - PI4Ka)$.

PIa: $dy(20) = r20*(((\alpha*W_{activation})^{n20})/(1+(\alpha*W_{activation})^{n20} +(\alpha*W_{inhibition})^{n20}) - PIa)$.

SACa: $dy(21) = 0$.

Model C (see supplement)

Model D

There are 25 species in the EGFR signaling network but only 20 differential equations are translated from the circuit of the EGFR signaling network. The reason is that EGF, PTEN, PIKb, SACa and SACb do not have any inputs therefore they are treated as constants. In Model D, PITP stimulates both PIP_2a and PIP_2b. PIP_2a is downstream from PI3K, and has its downstream species PIP_3 and Akt. PIP_2b is downstream from PLC_r, and has its downstream species IP_3 and DAG. The difference between Model D and Model C is that PITP stimulates PIP_2a in Model C. PIP_2a is independent of PI3K. PIP_2b is downstream from both PI3K and PLC_r.

EGF: $dy(1) = 0$.

EGFR: $dy(2) = r2*(((\alpha *W_{activation})^{n2})/(1+(\alpha*W_{activation})^{n2} +(\alpha*W_{inhibition})^{n2}) - EGFR)$.

Sos_2: $dy(3) = r3*(((\alpha *W_{activation})^{n3})/(1 + (\alpha * W_{activation})^{n3} +(\alpha*W_{inhibition})^{n3}) - Sos_2)$.

RasGDP: $dy(4)=r4*(((\alpha*W_{activation})^{n4})/(1+ (\alpha*W_{activation})^{n4} +(\alpha*W_{inhibition})^{n4}) - GDP)$.

RasGTP: $dy(5)=r5*(((\alpha*W_{activation})^{n5})/(1+(\alpha*W_{activation})^{n5}+(\alpha*W_{inhibition})^{n5})-GTP)$.

RasGAP: $dy(6) = r6*(((\alpha*W_{activation})^{n6})/(1+(\alpha*W_{activation})^{n6}) - GAP)$.

Raf_1: $dy(7) = r7*(((\alpha * W_{activation})^{n7}) / (1 + (\alpha*W_{activation})^{n7}) - Raf_1)$.

Mek: $dy(8) = r8*(((\alpha*W_{activation})^{n8})/(1 + (\alpha*W_{activation})^{n8}) - Mek)$.

Erk: $dy(9) = r9*(((\alpha*W_{activation})^{n9})/(1+(\alpha*W_{activation})^{n9}) - Erk)$.

PI3K: $dy(10) = r10*(((\alpha*W_{activation})^{n10})/(1+(\alpha*W_{activation})^{n10}) - PI3K)$.

PIP_2a: $dy(11)=r11*(((\alpha*W_{activation})^{n11})/(1+(\alpha*W_{activation})^{n11}+(\alpha*W_{inhibition})^{n11}) - PIP_2a)$.

PIP_3: $dy(12) = r12*(((\alpha*W_{activation})^{n12})/(1+(\alpha*W_{activation})^{n12}+(\alpha*W_{inhibition})^{n12}) - PIP_3)$.

PTEN: $dy(13) = 0$.

Akt: $dy(14) = r14*(((\alpha* W_{activation})^{n14}) / (1 + (\alpha* W_{activation})^{n14}) - Akt)$.

PITP: $dy(15) = r15*(((\alpha * W_{activation})^{n15}) / (1 + (\alpha*W_{activation})^{n15}) - PITP)$.

PLC_r: $dy(16) = r16*(((\alpha * W_{activation})^{n16}) / (1 + (\alpha*W_{activation})^{n16}) - PLC_r)$.

IP_3: $dy(17) = r17*(((\alpha*W_{activation})^{n17})/(1+(\alpha*W_{activation})^{n17}) - IP_3)$.

DAG: $dy(18) = r18*(((\alpha*W_{activation})^{n18})/(1+(\alpha*W_{activation})^{n18}) - DAG)$.

PI4Ka: $dy(19) = r19*(((\alpha*W_{activation})^{n19})/(1+(\alpha*W_{activation})^{n19}) - PI4Ka)$.

PIa: $dy(20) = r20*(((\alpha*W_{activation})^{n20})/(1+(\alpha*W_{activation})^{n20}+(\alpha*W_{inhibition})^{n20}) - PIa)$.

SACa: $dy(21) = 0$.

PIP_2b: $dy(22)=r22*(((\alpha*W_{activation})^{n22})/(1+(\alpha*W_{activation})^{n11}+(\alpha*W_{inhibition})^{n22}) - PIP_2b)$.

PI4Kb: $dy(23) = r23*(((\alpha*W_{activation})^{n23})/(1+(\alpha*W_{activation})^{n23}) - PI4Kb)$.

PIb: $dy(24) = r24*(((\alpha*W_{activation})^{n24})/(1+(\alpha*W_{activation})^{n24}+(\alpha*W_{inhibition})^{n24}) - PIa)$.

SACb: $dy(25) = 0$.

Model Fitting

As in many cell signaling models, there is limited data of biochemical reactions available. Since that is true for our model, we used the Markov Chain Monte Carlo (MCMC) method to select parameter values (see methods). The initial values were set to non-zero for some species that do have basal activity (Table 3), others were set to zero. The non-zero initial values are approximation of educated guess. We used experimental data of wild type (from Beak) to fit Erk and Akt response in the model. After bringing the model response to the target region, 10,000 runs were simulated. Among the 10,000 sets, around 25% of the sets met the criteria of the experimental data points and were selected. The final result is an average of these selected parameter values.

Table 3: Protein initial values

	Species	Initial values
1	EGF	0.5
2	EGFR	0.7
3	RasGDP	0.9
4	PIP_2	0.1
5	PI4K	0.1
6	PI	0.5
7	PTEN	0.5
8	SAC	0.5

Figure 5 shows the EGF-stimulated response from Erk and Akt resulting from averages of the parameter sets. The top plot is response from Erk, the bottom plot is from Akt. Transient response from Akt is faster and precedes response from Erk. The stars are data points from experiment, the plot is from simulation. The Y-axis is the normalized activity of concentration; the X-axis is the simulation time of 0 to 1800 seconds.

Figure 6 shows the EGF-stimulated response from PIP_2 and PIP_3 resulting from averages of the parameter sets. The top plot is the response from PIP_2, the bottom plot is from PIP_3. The stars are data points from experiment, the plot is from simulation. The Y-axis is the normalized activity of concentration; the X-axis is the simulation time of 0 to 1800 seconds. The activity from PIP_3 is 10% of activity from PIP_2 at maximum from experimental data. The model shows about 13%.

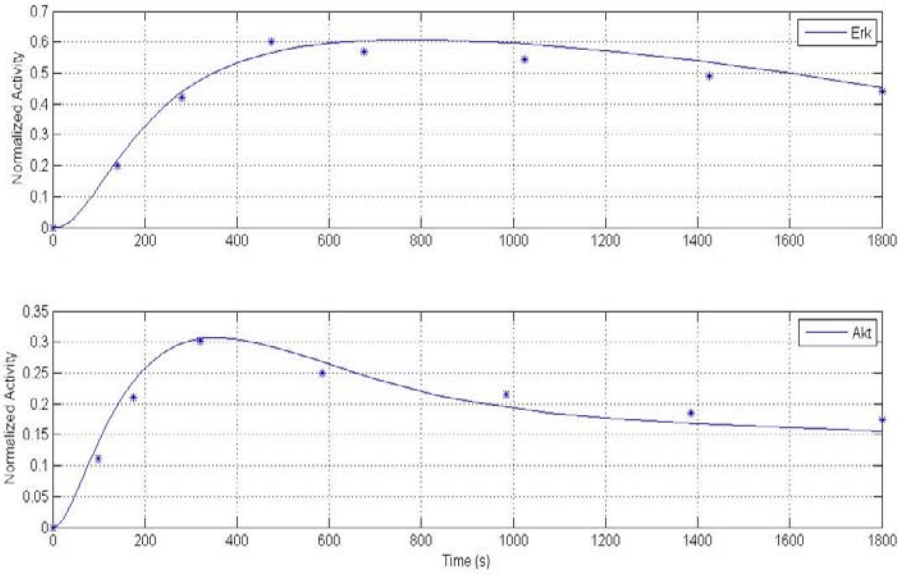

Figure 5: Model B, EGF-stimulated response Erk and Akt averaged from selected parameter sets. Stars are data points from experiment.

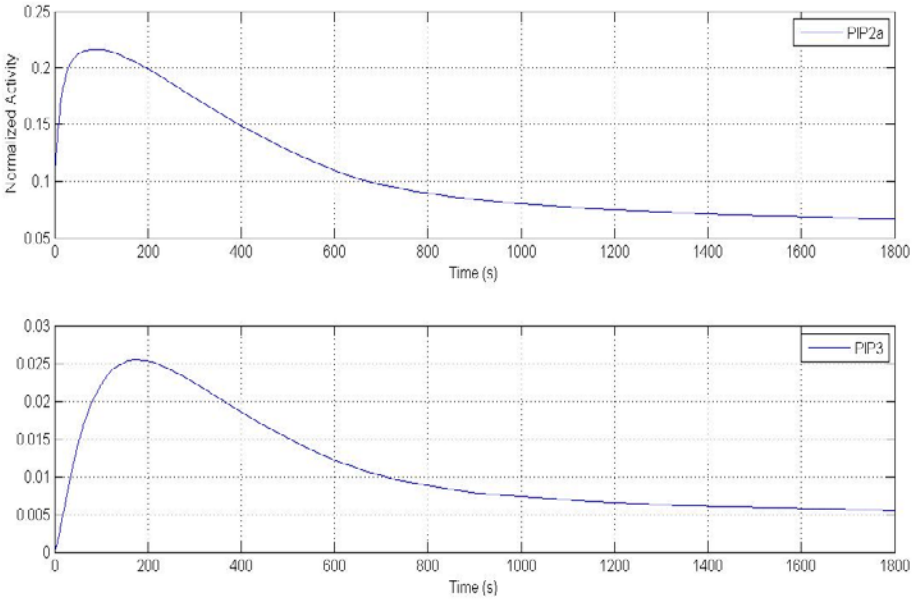

Figure 6: Model B, EGF-stimulated response from PIP_2 and PIP_3 averaged from selected parameter sets.

Figure 7 and Figure 8 show the same type of results from Model D. The activity from PIP_3 is 10% of activity from PIP_2 at maximum from experimental data. The model shows about 13%.

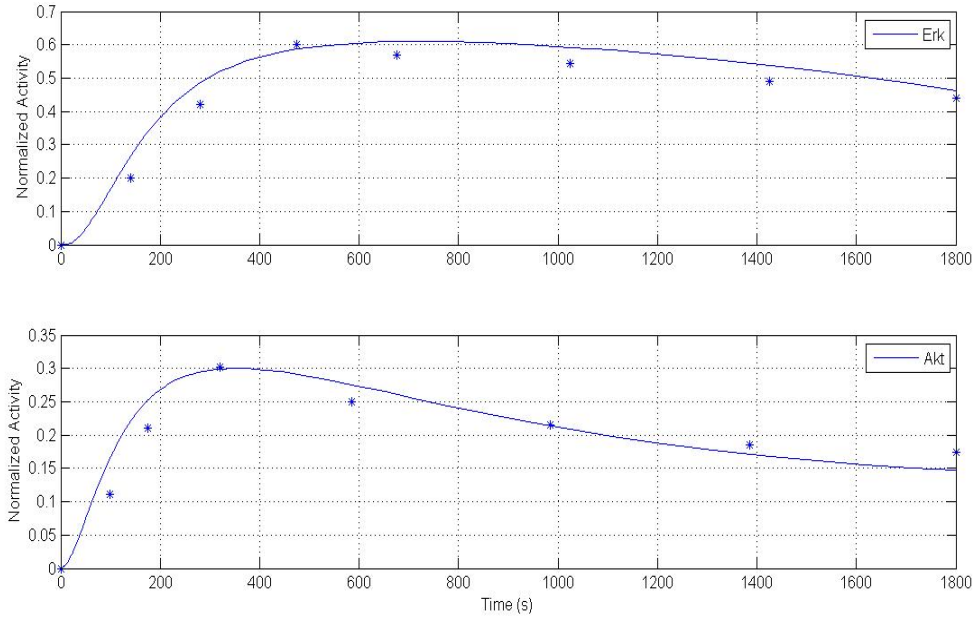

Figure 7: Model D, EGF-stimulated response Erk and Akt averaged from selected parameter sets. Stars are data points from experiment.

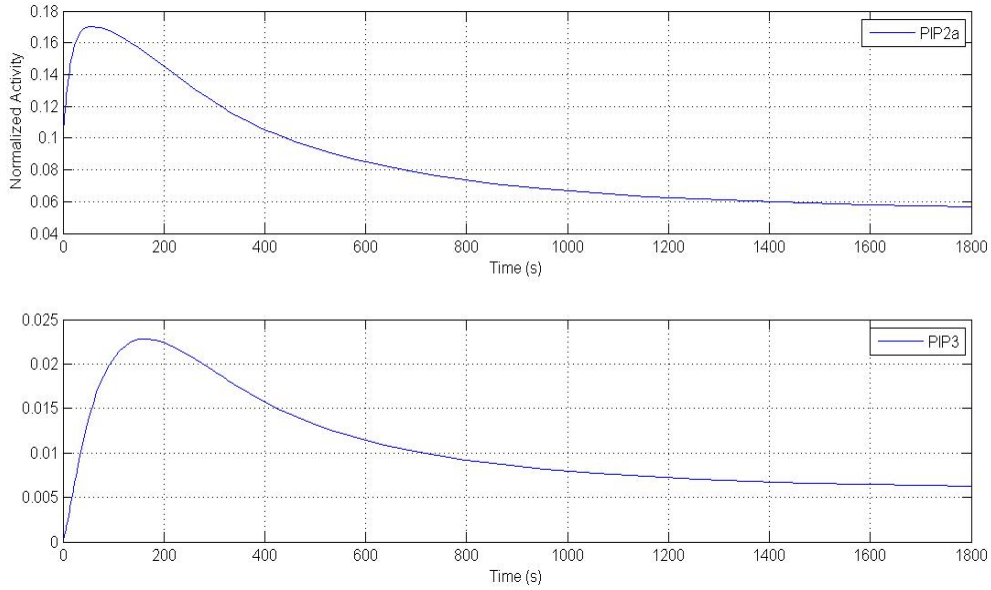

Figure 8: Model D, EGF-stimulated response from PIP_2 and PIP_3 averaged from selected parameter sets.

Statistical Data Analysis of Parameter Convergence

Good methods for testing convergence of target distribution are the histograms of posterior distribution density and time series trace plots. Figure 9 and Figure 10 show that the histograms of the posterior distribution are converged. The Y-axis is the number of values from different parameter sets, and the X-axis is the weight. Notice that some parameters are converged well and some have long tails. This may provide insights to the signaling network (see below). The posterior distribution is not symmetric because negative concentration has no meaning in cell signaling.

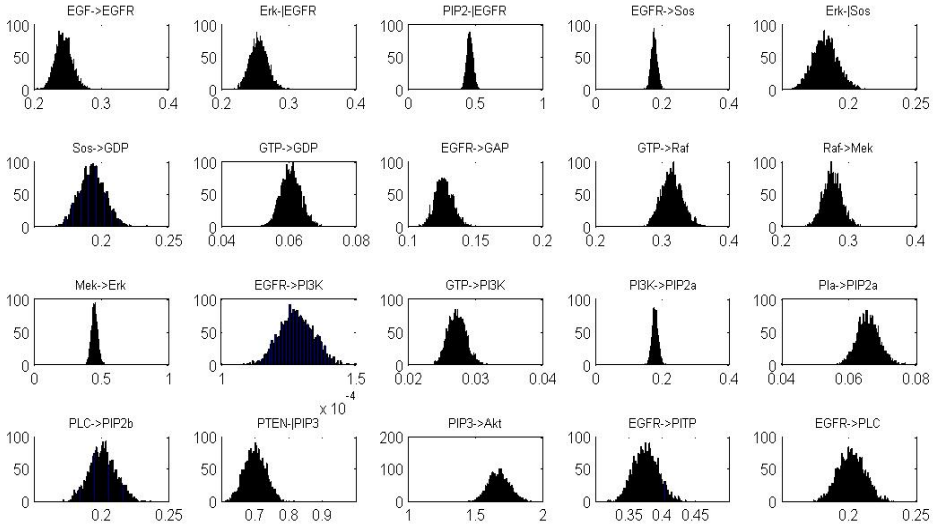

Figure 9: Model B, Histograms of posterior distribution showing the parameter selection convergence.

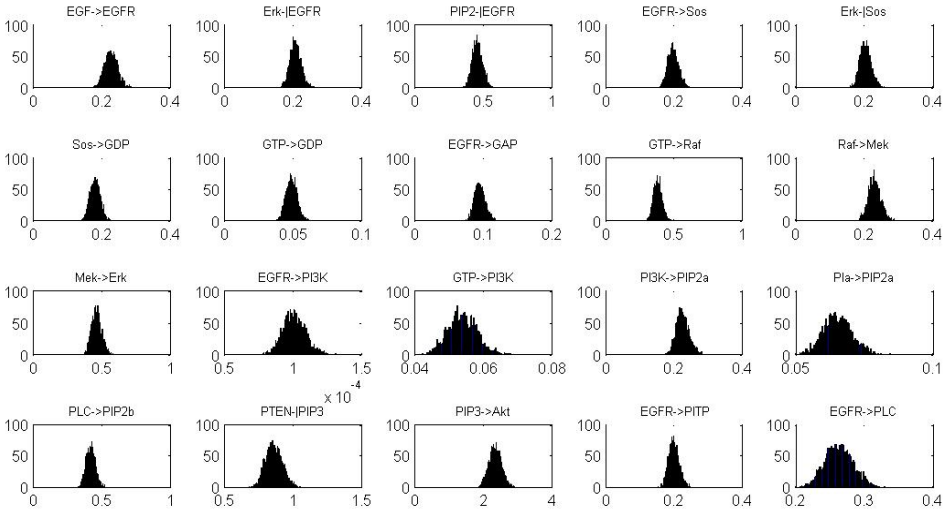

Figure 10: Model D, Histograms of posterior distribution showing the parameter selection convergence.

Figures 11 and 12 show that the time series trace of the posterior distribution is converged. The trace is considered well mixing and has no long flat periods. The Y-axis is the weight, and the X-axis is the number of iterations. Black is selected parameter sets, and white is rejected parameter sets. All parameters were tested for convergence. Here only show three parameters for demonstration.

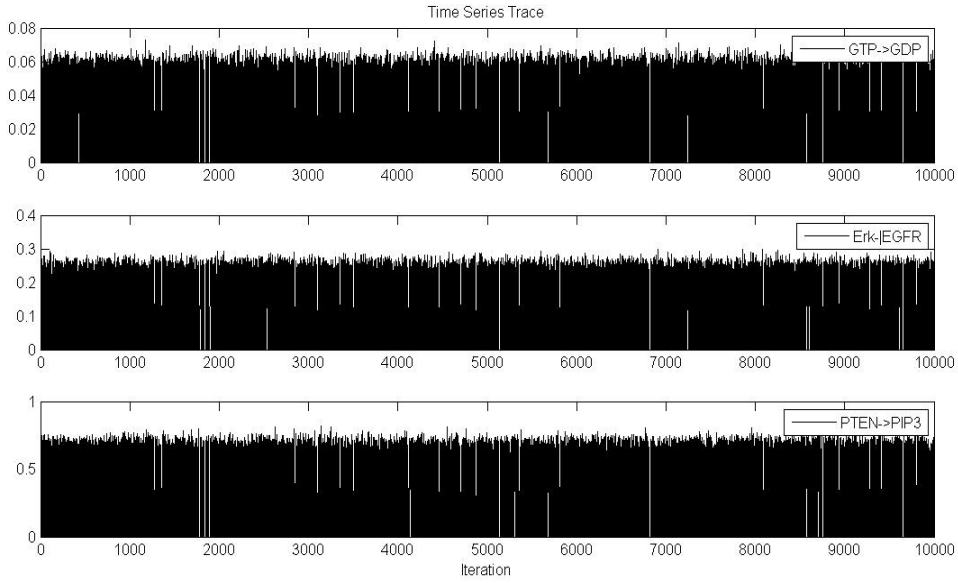

Figure 11: Model B, Time series trace of posterior distribution showing that the selected parameter values are well mixing.

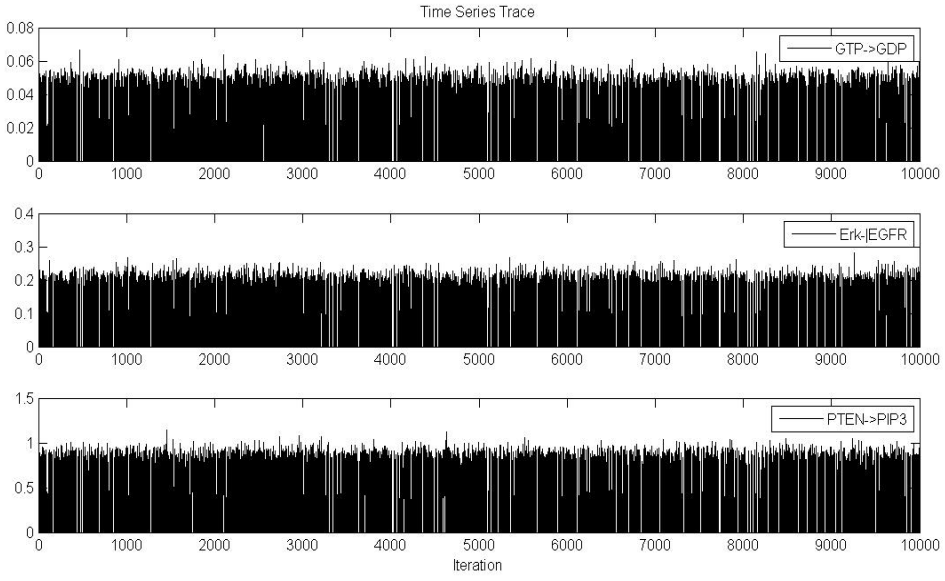

Figure 12: Model D, Time series trace of posterior distribution showing that the selected parameter values are well mixing.

Statistical Data Analysis of Weights in Relation to Signaling Regulation

Tables 3, 4 show the maxima, minima, means and standard deviations of weights in the EGFR signaling network. The simulation results are not unique, and are averaged from parameter sets selected from simulations. The standard deviation measures the variability of the data. The smaller standard deviation indicates less variability and stable, the larger standard deviation indicates more variability and has large range of values. From statistical data analysis, the standard deviations of weights may provide insights about where the strong regulation occurs and where the more balanced part occurs in the signaling network. The parameters with smaller standard deviation in comparison with other parameters in the network may suggest that those species provide stable regulation in the network. The parameters with larger standard deviation in comparison with other parameters in the network may suggest that those species provide robust regulation in the network.

Table 4: Model B, Weight parameter sets selected from simulations. Maxima, minima, means and standard deviations of parameter sets show the convergence. -> sign indicates activation, -| sign represents inhibition.

	Parameter	Max.	Min.	Mean	Std. dev.
1	EGF ->EGFR	0.2866	0.2022	0.2442	0.0123
2	Erk -> EGFR	0.2986	0.2155	0.2551	0.0126
3	PIP_2 -| EGFR	0.5544	0.3893	0.4576	0.0232
4	EGFR -> Sos_2	0.2055	0.1461	0.1755	0.0087
5	Erk -| Sos_2	0.2159	0.1549	0.1822	0.0091
6	Sos_2 -| RasGDP	0.2338	0.1589	0.1931	0.0096
7	RasGTP ->RasGDP	0.0729	0.0507	0.0605	0.0030
8	RasGDP ->RasGTP	0.1503	0.1119	0.1321	0.0062
9	RasGAP -| RasGTP	0.0657	0.0464	0.0556	0.0028
10	EGFR ->RasGAP	0.1473	0.1082	0.1266	0.0063
11	RasGTP ->Raf_1	0.3717	0.2622	0.3139	0.0154
12	Raf_1 ->Mek	0.3256	0.2344	0.2743	0.0136
13	Mek ->Erk	0.5415	0.3770	0.4481	0.0220
14	EGFR ->PI3K	0.0001	0.0001	0.0001	0.0000
15	RasGTP ->PI3K	0.0322	0.0231	0.0272	0.0013
16	PI3K -| PIP_2a	0.2099	0.1481	0.1778	0.0090
17	PIP_3 -> PIP_2a	0.7823	0.5575	0.6608	0.0317
18	PIa -> PIP_2a	0.0765	0.0563	0.0655	**0.0032**
19	PLC_r -| PIP_2a	0.2423	0.1718	0.2009	0.0101
20	PIP_2a -> PIP_3	0.1952	0.1492	0.1709	0.0068
21	PTEN -| PIP_3	0.8224	0.6005	0.6997	0.0314
22	PIP_3 -> Akt	1.9630	1.3304	1.6730	0.0809
23	EGFR -> PITP	0.4541	0.3187	0.3750	0.0186
24	PITP -> PIKa	0.3923	0.2678	0.3196	0.0157
25	PIKa -| PIa	0.0826	0.0590	0.0702	0.0035
26	EGFR -> PLC_r	0.2402	0.1734	0.2016	0.0100
27	PIP_2a -> IP_3	0.2266	0.1616	0.1929	0.0094

Table 5: Model D, Weight parameter sets selected from simulations. Maxima, minima, means and standard deviations of parameter sets show the convergence. -> sign indicates activation, -| sign represents inhibition.

	Parameter	Max.	Min	Mean	Std. dev.
1	EGF ->EGFR	0.2852	0.1745	0.2267	0.0183
2	Erk -> EGFR	0.2805	0.1584	0.2073	0.0163
3	PIP_2 -| EGFR	0.6228	0.3402	0.4571	0.0370
4	EGFR -> Sos_2	0.2520	0.1508	0.1971	0.0156
5	Erk -| Sos_2	0.2667	0.1485	0.2025	0.0163
6	Sos_2 -| RasGDP	0.2351	0.1367	0.1798	0.0145
7	RasGTP ->RasGDP	0.0665	0.0378	0.0489	0.0040
8	RasGDP ->RasGTP	0.2384	0.1380	0.1850	0.0145
9	RasGAP -| RasGTP	0.0619	0.0349	0.0453	0.0037
10	EGFR ->RasGAP	0.1188	0.0733	0.0947	0.0075
11	RasGTP ->Raf_1	0.4968	0.2950	0.3806	0.0304
12	Raf_1 ->Mek	0.2971	0.1768	0.2282	0.0181
13	Mek ->Erk	0.6035	0.3506	0.4611	0.0361
14	EGFR ->PI3K	0.0001	0.0001	0.0001	0.0000
15	RasGTP ->PI3K	0.0692	0.0401	0.0539	0.0043
16	PI3K -| PIP_2a	0.2947	0.1633	0.2243	0.0184
17	PIP_3 -> PIP_2a	1.1981	0.6948	0.9057	0.0668
18	PIa -> PIP_2a	0.0844	0.0503	0.0652	0.0053
19	PLC_r -| PIP_2b	0.5320	0.3115	0.4167	0.0339
20	PIP_2a -> PIP_3	0.1421	0.0800	0.1050	0.0083
21	PIb -> PIP_2b	0.2370	0.1460	0.1906	0.0126
22	PTEN -| PIP_3	1.1430	0.6812	0.8624	0.0630
23	PIP_3 -> Akt	3.1166	1.8432	2.3457	0.1855
24	EGFR -> PITP	0.2687	0.1482	0.1992	0.0156
25	PITP -> PIKa	0.4875	0.2844	0.3694	0.0291
26	PIKa -| PIa	0.0883	0.0513	0.0689	0.0055
27	EGFR -> PLC_r	0.3580	0.2048	0.2629	0.0212
28	PIKa -| PIa	0.1037	0.0603	0.0811	0.0066
29	PIP_2b -> IP_3	0.3244	0.1845	0.2389	0.0199

The simulation results from four models are fitted to the experimental data of response from wild type of Erk and Akt. By comparing the simulation results from these four models, we recognized that there are properties in common among these four models. As

shown in Table 2 (Model B) and Table 3 (Model D), maxima, minima, means and standard deviations of weights in the EGFR signaling network may provide insights on the regulation of the network (Model A and Model C, see supplement). We divided the standard deviation into three categories: larger standard deviation with value greater than 0.03 and colored red; smaller standard deviation with value smaller than 0.01 and colored with green; standard deviation with value greater than 0.01 and smaller than 0.03. Parameters with larger standard deviation may suggest stable enzymatic regulation in cell signaling network. In contrast, parameters with smaller standard deviation may suggest robust regulation in cell signaling network.

In Table 2 (Model B), we identified several weights that resulted in standard deviations greater than 0.1. Two of these weights, PIP_2 -| EGFR and PTEN -| PIP_3, exhibit inhibition which negatively regulate downstream species. This property may indicate the importance of negative regulation in the cell signaling transduction network. The two weights, PIP_2 -> PIP_3, and PIP_3 -> Akt, demonstrate activation, which positively regulates downstream species. The larger standard deviations and the wider range of weights may suggest, in biological system, the indication of robustness of the regulation in the network.

Based on standard deviations of the weights, we divided them into three groups and colored the EGFR signaling network diagrams for better understanding. Red indicates that the parameter has a larger standard deviation greater than 0.03 and a wider range of weights. Green indicates that the parameter has a smaller standard deviation less than 0.01 and a narrower range of weights. Blue indicates that the parameter has a standard deviation

between 0.03 and 0.01. Do these differences in pathways impart significant biological information?

Two common properties are observed from colored diagrams of the four models are: that both RasGDP, RasGTP conversion, and PI, PIP_2 conversion are colored with green. Statistical analysis indicate that these enzymatic reactions are stable and balanced parts in the signaling network, perhaps suggesting a biological characteristic that cell signaling is all about enzymatic regulation in the network. Another common property is that PTEN inhibition on Akt is colored with red, indicating that inhibition has a wider range of weights. From [Li and Sun, 1997, Li et al. 1997; Steck et al. 1997], we know that PTEN is a tumor suppressor and Akt is an oncogene [Staal, 1987; Bellacosa, 1995; Cheng, 1996]. It has been shown that Akt was amplified in a number of human tumors through indirect means such as amplification of PI3K, or, more commonly deletion of PTEN [Vivanco I, 2002]. This characteristic may also be an indication in biology that PTEN has strong local regulation of Akt via PIP_3. Figure 13 (Model B) and Figure 14 (Model D) show the colored diagrams. The colored diagrams for Model A and Model C are shown in the supplement.

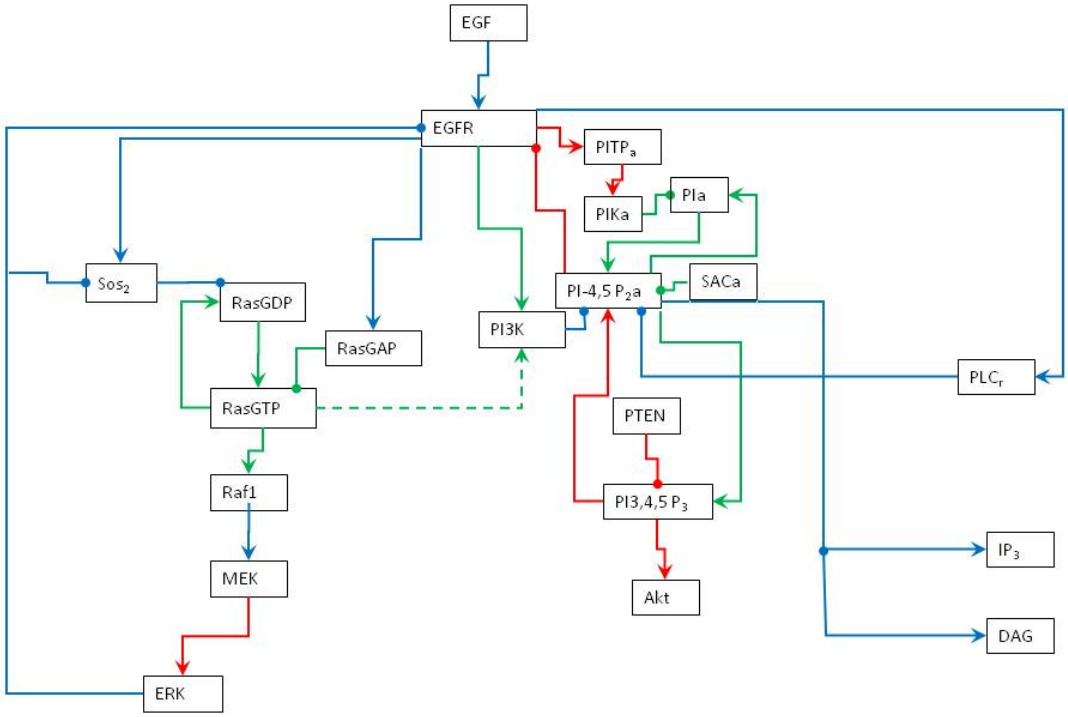

Figure 13: Model B, colored EGFR signaling network indicating the range of the parameter value. Red indicates a standard deviation greater than 0.03; blue signifies a standard deviation between 0.01 and 0.03; and green represents a standard deviation less than 0.01.

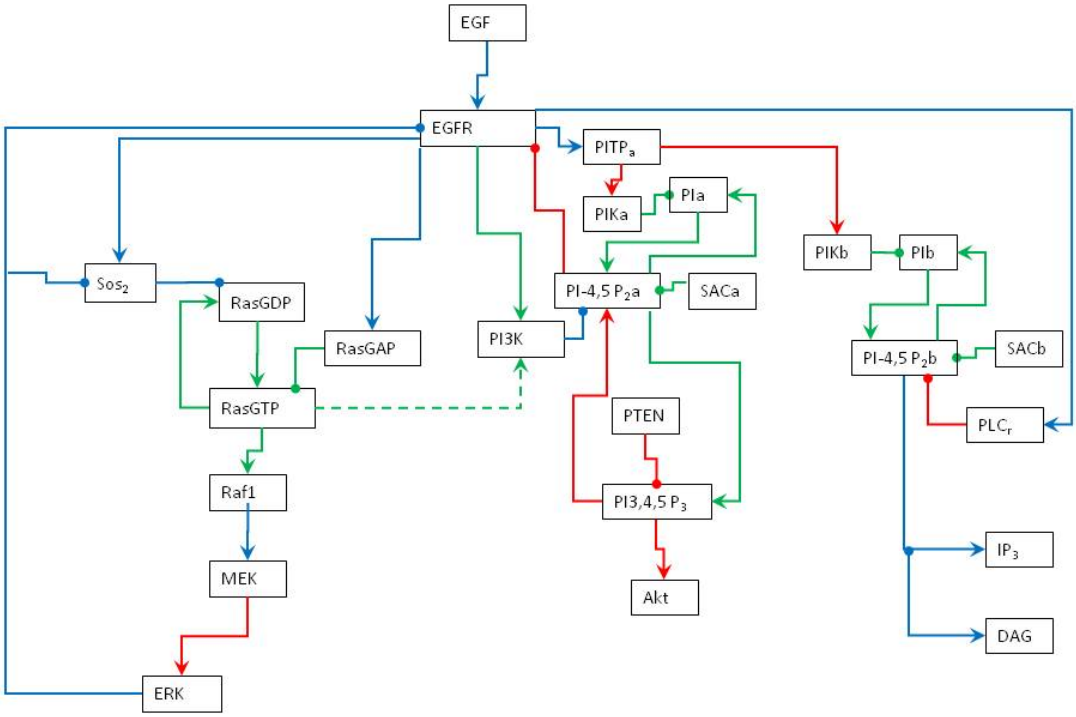

Figure 14: Model D, colored EGFR signaling network indicating the range of the parameter value. Red indicates a standard deviation greater than 0.03; blue signifies a standard deviation between 0.01 and 0.03; and green represents a standard deviation less than 0.01.

Simulation of Mutated Oncogenes and Tumor Suppressors in Cancer Research

Ras [Barbacid, 1987; Bos, 1989] and PI3K [Carpenter, 1990] are oncogenes. PTEN is a tumor suppressor [Li and Sun, 1997, Li et al. 1997; Steck et al. 1997], and cells lacking PTEN function exhibit a two fold increase in PtdIns-3,4,5-P3 levels [Stambolic et al. 1998; Sun et al. 1999]. We used different parameter values to simulate these oncogenes and tumor suppressors, and observed the dynamic behavior of the network at a systems level. When the weight was increased in RasGDP, RasGTP conversion cycle, simulation results show that

response from Erk becomes sustained, which is similar to the experimental data reported in a study by [Birtwistle et al. 2007]. They reported that cancer cells have sustained response from Erk. In modeling, the sustained response most likely is due to saturation. To confirm the possibility of the sustained response caused by oncogene, bistability investigation is necessary. As a consequence of global inhibition from Erk to EGFR and Sos_2, Akt response becomes more transient. Figure 15 (Model B) and Figure 16 (Model D) show Erk and Akt in response to mutated oncogene Ras.

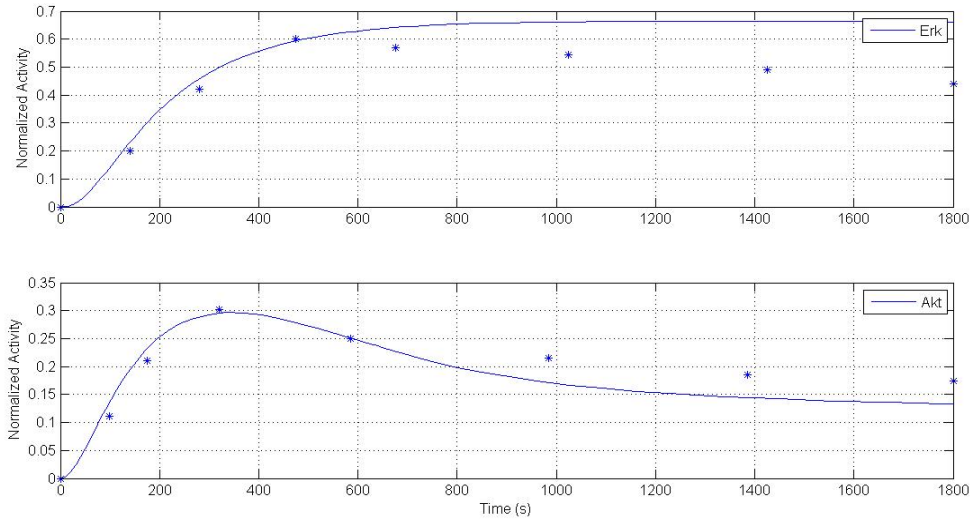

Figure 15: Model B, simulation of mutated oncogene Ras. Erk response becomes sustained, and, as a consequence of global inhibition from Erk, Akt response becomes more transient. Stars are data points from experiment.

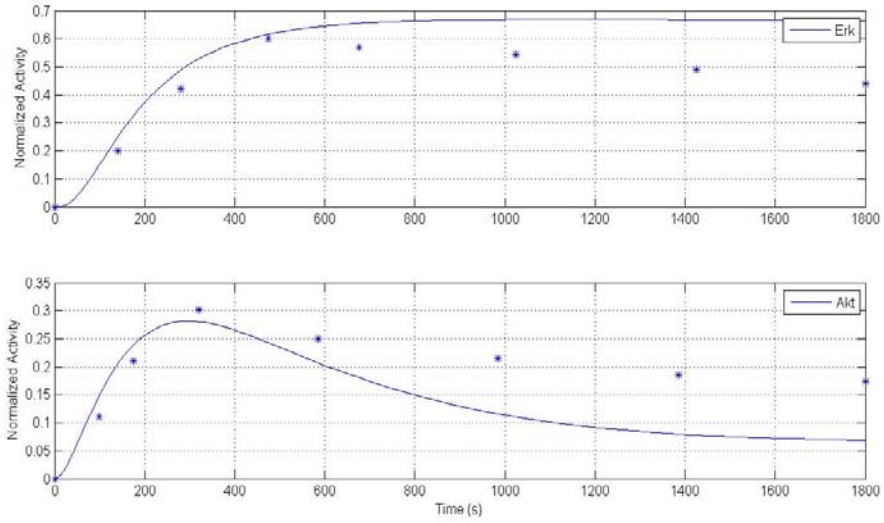

Figure 16: Model D, simulation of mutated oncogene Ras. Erk response becomes sustained, and, as a consequence of global inhibition from Erk, Akt response becomes more transient. Stars are data points from experiment.

Another oncogene PI3K, shows similar dynamic behavior. We increased the weight for the PI3K, PIP_2 conversion cycle, and simulation results showed that response from Akt becomes sustained which is similar to experimental data from studies by [Staal, 1987; Bellacosa, 1995; Cheng, 1996]. Their results indicated that cancer cells have sustained response from Akt. In modeling, the sustained response most likely is due to saturation. To confirm the possibility of the sustained response caused by oncogene, bistability investigation is necessary. No change in Erk response is observed. Figure 17 (Model B) and Figure 18 (Model D) show Erk and Akt in response to mutated oncogene PI3K.

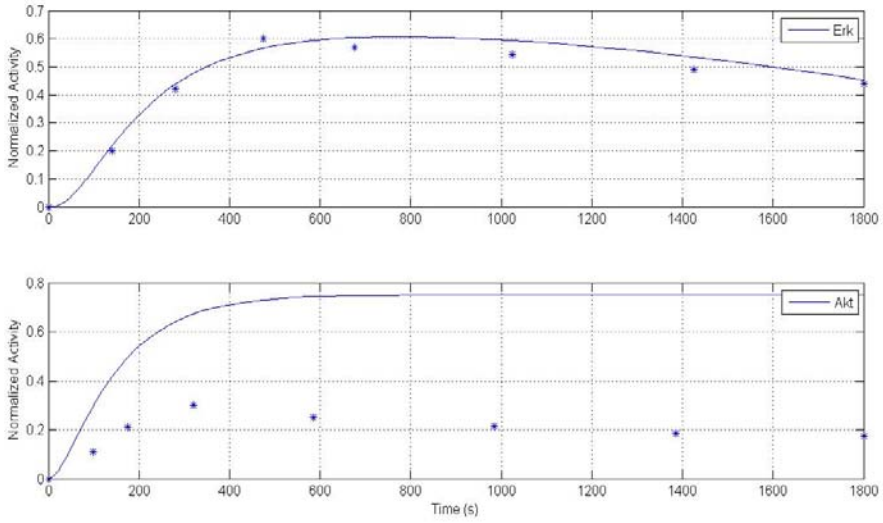

Figure 17: Model B, simulated mutated oncogene PI3K showing that Akt response became sustained with no change from Erk response. Stars are data points from experiment.

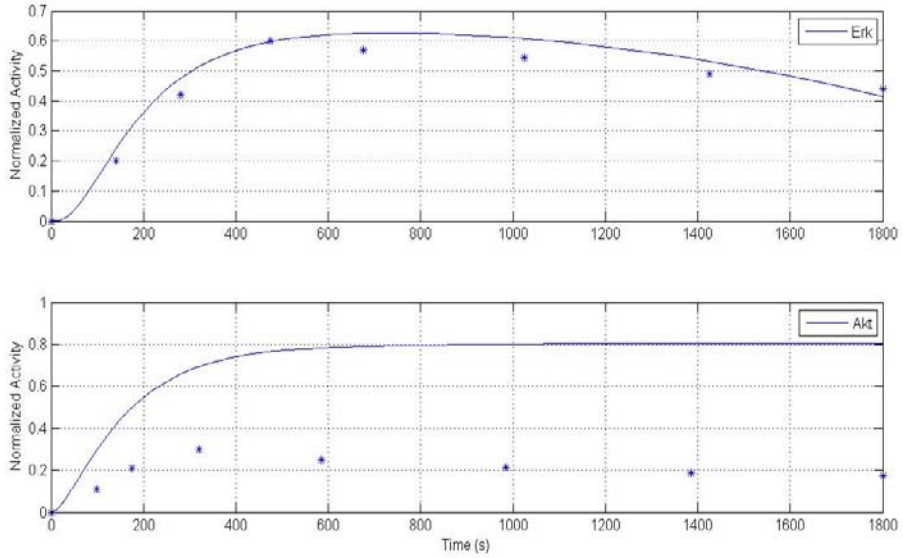

Figure 18: Model D, simulated mutated oncogene PI3K showing that Akt response became sustained with no change from Erk response. Stars are data points from experiment.

For tumor suppressor PTEN, we decreased weight from PTEN to PIP_3 to simulate mutated PTEN. Results showed that the amplitude of Akt response increased and the Akt response became more sustained. This result may suggest that PTEN as a tumor suppressor has stronger negative regulation on Akt. Even though our model does not quantify the level of dysfunction of mutated PTEN, it does provide an indication that PTEN has an important role on regulation of Akt. Figure 19 (Model B) and Figure 20 (Model D) show Erk and Akt in response to mutated tumor suppressor PTEN.

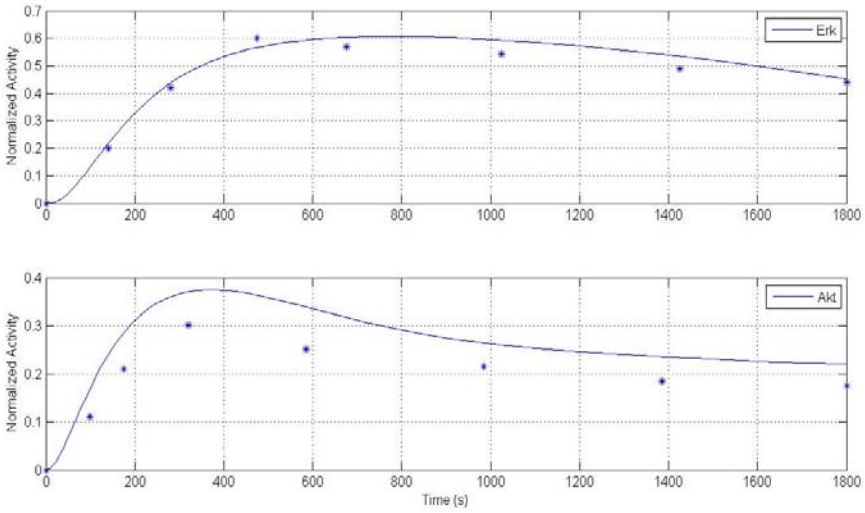

Figure 19: Model B, simulate PTEN as a tumor suppressor. Akt response amplitude increased and became more sustained. Stars are data points from experiment.

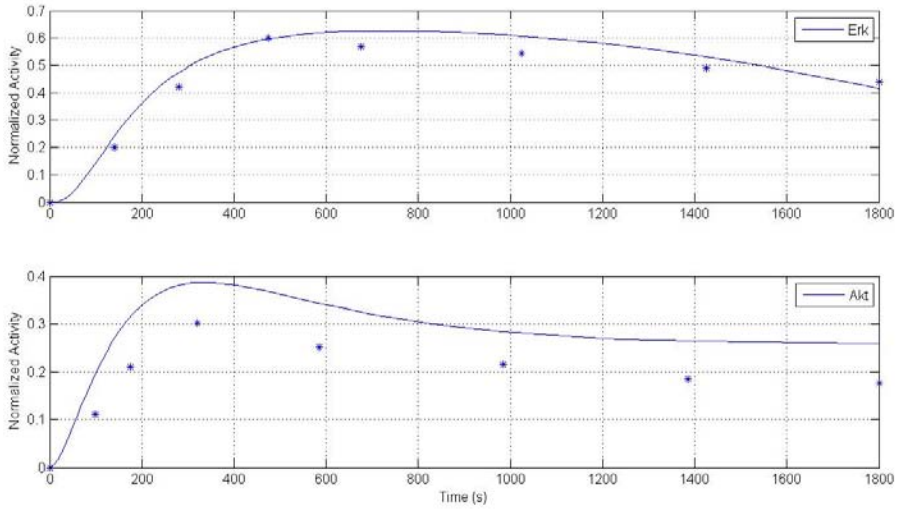

Figure 20: Model D. Simulate PTEN as a tumor suppressor. Akt response amplitude increased and became more sustained. Stars are data points from experiment.

Perturbation of PITP

We perturbed PITP as knock out PITP, by setting the weight from EGFR to PITP as zero. The simulation results showed that the perturbation on PITP had no affect on PIP_2. Therefore, no effect on downstream species PIP_3 and Akt were evident. From standard deviation analysis, we have discussed that PI and PIP_2 conversion is a stable and balanced part of the PI3K pathway. This finding may suggest that PI and PIP_2 conversion is more self - sustained due to basal activity and positive feedback loops from PIP_2 to PI and from PIP_3 to PIP_2 as well. This simulation results indicate an close agreement with PITP experimental data (not published) qualitatively. Another possibility, as pointed out by [Tanaka, S., et al 1994, Cunningham, E., et al. 1996], may be that in functional studies in mammalian cells,

ectopically expressed Sec14p can be used to compensate for the loss of PITPs. Our studies did not model this.

Model Prediction and Limitation

From the simulation results and data analysis of four models, our predictions are that Model B and Model D may represent the PITP structure in the EGFR signaling network and in a closer agreement with the experimental data (not published). The reason for our prediction is that in both Model B and Model D, the PI3K pathway and the PLC_r pathway are stimulated by PITPα. In Model B, PITPα stimulates both PIP_2a, and PIP_2a is accessed by PI3K pathway and PLC_r pathway. In Model D, PITPα stimulates both PIP_2a and PIP_2b, PIP_2a is accessed by the PI3K pathway, and PIP_2b is accessed by PLC_r pathway. But in Model C, PI3K pathway and PLC_r pathway are not stimulated by PITPα. As Hsuan et al. (2001) pointed out, that PITPα was identified as an essential component in ensuring substrate supply to PLC.

The methodology used for our modeling may be good for theoretical exploration, hypothesis generation, and prediction formulation. Like any modeling method, our model has its limitations. It does not provide detailed quantitative information about the signaling network, such as, the level of mutated PTEN necessary to cause PTEN dysfunction.

Conclusion

In conclusion, we were first to develop a regulatory model using EGF as stimulus to activate Erk and Akt in PITP signaling transduction network. The model construction is based on the information of the signaling directions and activation or inhibition in the

signaling network. Therefore this methodology provides qualitative insights about cell signaling network.

We chose the Hill function to formulate ODEs because Hill coefficient can determine the cooperativeness of the ligand and enzyme or receptor binding, thus providing biological information to the signaling network model. We developed an algorithm to select parameters using Markov Chain Monte Carlo methods, and tested parameter convergence. The analysis of standard deviations of weights does provide some insights. Notably, the stable enzymatic regulation of Ras and PI and strong negative regulation of PTEN on Akt in the signaling network. Oncogene simulation results show the relationship to cancer research.

From the simulation results of four models, our prediction of PITP structure in the EGFR signaling network is Model B. The reason is that PITP activates PI3K pathway and PLC_r pathway in Model B which may closely represent experimental data (not published). The statistical data analysis shows that the standard deviations of weights in Model B are more stable. Model B can be the starting point for future modeling on EGF-stimulated PITP signaling network, more realistic kinetics modeling can be used to gain additional insights.

Our model has its limitation in that it does not provide detailed quantitative information about the signaling network.

MATERIALS AND METHODS

Model Development

Figure 1-4 depict PITP Model A through Model D of the EGFR signal transduction networks. Molecules are species, and weights are interactions between two species. The dynamics of weights, scaling coefficients, and the Hill coefficient reflect the system behavior. Circuits of the network are translated into Ordinary Differential Equations (ODE). The initial values are educated guesses from what we know about the biological process. For example we know that the concentration of RasGDP should be higher because this inactive form is present in abundance. The concentration is between 0 and 1. The weights are random numbers generated by Matlab [The Mathworks; Natick, MA] and selected by Markov Chain Monte Carlo method. Weights are positive values for both activation and inhibition. There are 20 species and 70 parameters in the EGFR signaling network. Simulations perform a deterministic computation to select all possible combinations of weights, scaling coefficients and Hill coefficients based on the experimental data points from Erk and Akt readout.

Governing Equations

We used the Hill function to formulate ODEs for each species in the signal transduction network. Since the Hill function is used to determine the cooperativeness of the ligand binding to the enzyme or receptor, there is biological meaning to justify use of the Hill function in the signaling network modeling.

There are three types of equations: activation, inhibition, and both activation and inhibition.

1. Activation equation:

$$dx/dt = r^* ((\alpha^* W_{activation})^n / (1+ (\alpha^* W_{activation})^n) - x).$$

2. Inhibition equation:

$$dx/dt = r^*((\alpha^* W_{inhibition})^n / (1+ (\alpha^* W_{inhibition})^n) - x).$$

3. Activation and Inhibition equation:

$$dx/dt = r^*((\alpha^* W_{activation})^n / (1+ (\alpha^* W_{activation})^n + (\alpha^* W_{inhibition})^n) - x).$$

x – a species in the signal transduction network. A negative sign preceding x here indicates x is a decay term. The range of concentration is between 0 and 1.

r – time is a scaling coefficient in the range of 0 to 1.

α –a is constant with value eight.

$W_{activation}$ – sum of activation weights is the total amount of activation input a species received from other species in the network. The initial weight is between 0 and 1. It could be greater than 1 as a result of parameter selection.

$W_{inhibition}$ – sum of inhibition weights is the total amount of inhibition input a species received from other species in the network. The initial weight is between 0 and 1. It could be greater than 1 as a result of parameter selection.

n – Hill function coefficient. The initial Hill coefficient is greater than 1.

Parameter Selection Using Markov Chain Monte Carlo Methods

In cell signaling modeling, there is relatively limited pool of data on stoichiometry and kinetics of the biochemical reactions. The classical method is to gather rate constants from experimental data. There are several issues when gathering information in this fashion. First, it is difficult to measure the rate constants for a large network which has a large

parameter set. Second, these data are collected from different laboratories which may lead to inherent inconsistencies. We used the Markov Chain Monte Carlo (MCMC) method and developed an efficient algorithm to simulate parameter selection in EGFR signaling network model.

Markov Chain Monte Carlo (MCMC) methods are used to simulate direct draws from some complex, nonstandard multivariate distributions of interest [Chib et al. 1995].

A Markov Chain is a sequence of random variables generated by a Markov process which is defined by its transition probabilities. These transition probabilities between different values in the sample space depend only on the random variable's current state. Thus, the only information about the past available to predict the future is the current state of the random variable. Knowledge about an earlier state of the random variable does not change the transition probability [Walsh, 2004].

Monte Carlo sampling uses Bayesian inference that is based on random sampling. It was introduced by Metropolis [Metropolis et al. 1953] at Los Alamos National Laboratory. Some systems cannot be computed exactly and therefore they are predicted on the basis of approximation. Some examples are the van der Vaals equations for dense gases, and Boltzmann equations for dilute gases, among many others.

Markov Chain Monte Carlo Simulation

There are different MCMC algorithms depending on the applications. We chose Metropolis-Hastings algorithm for our simulation. The usual approach is to start with any value and draw the sample space from a uniform or normal distribution with specified mean and standard deviation.

Since we do not have any known parameter values in the signaling network, we selected a set of parameters from the uniform distribution, which can generate Erk and Akt response as the starting point. Doing so requires a shorter burn-in period and brings the chain to the stationary state.

We used a design approach named divide and conquer. It is well known in computer science field and is the basis for binary search, quick sort, among others. This approach breaks a complex problem into sub-problems of similar type and solves sub-problems in an efficient manner. We applied this approach in the algorithm design to shorten the burn-in period and bring the chain to convergence more quickly. Parameters were divided into blocks: pathways, weight, scaling coefficient, and Hill coefficient. The standard deviation may be different for each block and pathway.

The equation: New = Old * exp (sd*randn). New is the candidate parameter value, Old is the current parameter value, sd is the standard deviation of the posterior distribution, and randn is the command in Matlab [The Mathworks; Natick, MA] to generate a random normalized variable.

The algorithm we developed has the following steps:

1. Start with the selected parameter set mentioned above as the starting point.

2. Determine a proposal distribution or a move to the target region. Parameters are divided into different groups according to their characteristics in the cell signal transduction network: pathways, weight, Hill coefficient, and scaling coefficient. Standard deviations are specified for different groups according to their influence on the data fitting.

3. The candidate y is drawn from the process of y = x + sd*z, where x is the current or old value, sd is standard deviation and z is an increment random variable drawn from normal distribution.

4. The simulation calculates the system of ODEs that are interactions of species in the signaling network.

5. Determine if the response from Erk and Akt meets the criteria of data points from experience. If so, accept the set; otherwise reject the set and go to step 3.

In our simulation experiments, it may take several trials to bring the model response to the target region. We noticed that by adjusting standard deviation in the burn-in period, the burn-in period could be shortened. This result is due to the complexity of this model that the weight, time scaling coefficient, and Hill coefficient have different functions in influence the model fitting. Once the response is in the target region, the parameter sets are selected, if they meet the criteria of the experimental data points.

Supplement

Table 1: Model A**,** Circuits of the EGFR signaling network. Plus sign indicates activation received, minus sign indicates inhibition received. EGF, PTEN, and SACa receive neither activation nor inhibition from other species, they are treated as constants.

	Downstream species	Regulators
1	EGF	
2	EGFR	+ EGF – Erk
3	Sos_2	+ EGFR – Erk
4	RasGDP	+ RasGTP – Sos_2
5	RasGTP	+ RasGDP – RasGAP
6	RasGAP	+ EGFR
7	Raf_1	+ RasGTP
8	Mek	+ Raf_1
9	Erk	+ Mek
10	PI3K	+ EGFR + RasGTP
11	PIP_2a	+ PIa + PIP_3 – PI3K – PLC_r –SACa
12	PIP_3	+ PIP_2a – PTEN
13	PTEN	
14	Akt	+ PIP_3
15	PITP	+ EGFR
16	PLC_r	+ EGFR
17	IP_3	+ PIP_2a
18	DAG	+ PIP_2a
19	PIKa	+ PITP
20	PIa	+ PIP_2a – PIKa
21	SACa	

Table 2: Model C, Circuits of the EGFR signaling network. Plus sign indicates activation received, minus sign indicates inhibition received. EGF, PTEN, PIKb, SACa and SACb receive neither activation nor inhibition from other species, they are treated as constants.

	Downstream species	Regulators
1	EGF	
2	EGFR	+ EGF – Erk – PIP_2a
3	Sos_2	+ EGFR – Erk
4	RasGDP	+ RasGTP – Sos_2
5	RasGTP	+ RasGDP – RasGAP
6	RasGAP	+ EGFR
7	Raf_1	+ RasGTP
8	Mek	+ Raf_1
9	Erk	+ Mek
10	PI3K	+ EGFR + RasGTP
11	PIP_2a	+ PIa – SACa
12	PIP_3	+ PIP_2b – PTEN
13	PTEN	
14	Akt	+ PIP_3
15	PITP	+ EGFR
16	PLC_r	+ EGFR
17	IP_3	+ PIP_2b
18	DAG	+ PIP_2b
19	PIKa	+ PITP
20	PIa	+ PIP_2a – PIKa
21	SACa	
22	PIP_2b	+ PIb + PIP_3 – PI3K – PLC_r – SACb
23	PIKb	
24	PIb	+ PIP2b – PIKb
25	SACb	

Model Equations

Model A

There are 21 species in the EGFR signaling network but only 18 differential equations are translated from the circuit of the EGFR signaling network. The reason is that

EGF, PTEN, and SACa do not have any inputs. Therefore they are treated as constants. For completeness, their equations are equal to zero.

EGF: dy(1) = 0.

EGFR: dy(2) = r2*(((α*$W_{activation}$) ^{n2})/(1+(α*$W_{activation}$) ^{n2} +(α*$W_{inhibition}$) ^{n2}) – EGFR).

Sos_2: dy(3) = r3*(((α * $W_{activation}$) ^{n3})/(1 + (α * $W_{activation}$) ^{n3} +(α*$W_{inhibition}$) ^{n3}) – Sos_2).

RasGDP: dy(4)=r4*(((α*$W_{activation}$) ^{n4})/(1+ (α*$W_{activation}$) ^{n4} +(α*$W_{inhibition}$) ^{n4}) –GDP).

RasGTP: dy(5)=r5*(((α*$W_{activation}$) ^{n5})/(1+(α*$W_{activation}$) ^{n5} +(α*$W_{inhibition}$) ^{n5})–GTP).

RasGAP: dy(6) = r6*(((α *$W_{activation}$) ^{n6})/ (1+(α*$W_{activation}$) ^{n6}) – GAP).

Raf_1: dy(7) = r7*(((α * $W_{activation}$) ^{n7}) / (1 + (α*$W_{activation}$) ^{n7}) – Raf_1).

Mek: dy(8) = r8*(((α*$W_{activation}$) ^{n8})/(1 + (α*$W_{activation}$) ^{n8}) – Mek).

Erk: dy(9) = r9*(((α *$W_{activation}$) ^{n9})/(1+(α *$W_{activation}$) ^{n9}) – Erk).

PI3K: dy(10) = r10*(((α*$W_{activation}$) ^{n10})/(1+(α*$W_{activation}$) ^{n10}) –PI3K).

PIP_2: dy(11)=r11*(((α *$W_{activation}$) ^{n11})/(1+(α*$W_{activation}$) ^{n11}+(α*$W_{inhibition}$) ^{n11}) – PIP_2).

PIP_3: dy(12) = r12*(((α*$W_{activation}$) ^{n12})/(1+(α*$W_{activation}$) ^{n12}+(α*$W_{inhibition}$) ^{n12}) – PIP_3).

PTEN: dy(13) = 0.

Akt: dy(14) = r14*(((α* $W_{activation}$) ^{n14}) / (1 + (α* $W_{activation}$) ^{n14}) – Akt).

PITP: dy(15) = r15*(((α * $W_{activation}$) ^{n15}) / (1 + (α*$W_{activation}$) ^{n15}) – PITP).

PLC_r: dy(16) = r16*(((α * $W_{activation}$) ^{n16}) / (1 + (α*$W_{activation}$) ^{n16}) – P**LC_r**).

IP_3: dy(17) = r17*(((α *$W_{activation}$) ^{n17})/(1+(α*$W_{activation}$) ^{n17}) – IP_3).

DAG: dy(18) = r18*(((α*$W_{activation}$) ^{n18})/(1+(α*$W_{activation}$) ^{n18}) – DAG).

PIKa: dy(19) = r19*(((α*$W_{activation}$) ^{n19})/(1+(α*$W_{activation}$) ^{n19}) – PIKa).

PIa: dy(20) = r20*(((α*$W_{activation}$) ^{n20})/(1+(α*$W_{activation}$) ^{n20} +(α*$W_{inhibition}$) ^{n20}) – PIa).

SACa: dy(21) = 0.

Model C

There are 25 species in the EGFR signaling network but only 20 differential equations are translated from the circuit of the EGFR signaling network. The reason is that EGF, PTEN, PIKb, SACa and SACb do not have any inputs. Therefore they are treated as constants. For completeness, their equations are equal to zero.

EGF: dy(1) = 0.

EGFR: dy(2) = r2*(((α*$W_{activation}$) ^{n2})/(1+(α*$W_{activation}$) ^{n2}+(α*$W_{inhibition}$) ^{n2}) – EGFR).

Sos_2: dy(3) = r3*(((α * $W_{activation}$) ^{n3})/(1 + (α * $W_{activation}$) ^{n3} +(α*$W_{inhibition}$) ^{n3}) – Sos_2).

RasGDP: dy(4)=r4*(((α*$W_{activation}$) ^{n4})/(1+ (α*$W_{activation}$) ^{n4} +(α*$W_{inhibition}$) ^{n4}) –GDP).

RasGTP: dy(5)=r5*(((α*$W_{activation}$) ^{n5})/(1+(α*$W_{activation}$) ^{n5}+(α*$W_{inhibition}$) ^{n5})–GTP).

RasGAP: dy(6) = r6*(((α *$W_{activation}$) ^{n6})/ (1+(α*$W_{activation}$) ^{n6}) – GAP).

Raf_1: dy(7) = r7*(((α * $W_{activation}$) ^{n7}) / (1 + (α*$W_{activation}$) ^{n7}) – Raf_1).

Mek: dy(8) = r8*(((α*$W_{activation}$) ^{n8})/(1 + (α*$W_{activation}$) ^{n8}) – Mek).

Erk: dy(9) = r9*(((α *$W_{activation}$) ^{n9})/(1+(α *$W_{activation}$) ^{n9}) – Erk).

PI3K: dy(10) = r10*(((α*$W_{activation}$) ^{n10})/(1+(α*$W_{activation}$) ^{n10}) –PI3K).

PIP_2a: dy(11)=r11*(((α *$W_{activation}$) ^{n11})/(1+(α*$W_{activation}$) ^{n11}+(α*$W_{inhibition}$) ^{n11}) – PIP_2a).

PIP_3: dy(12) = r12*(((α*$W_{activation}$) ^{n12})/(1+(α*$W_{activation}$) ^{n12}+(α*$W_{inhibition}$) ^{n12}) – PIP_3).

PTEN: dy(13) = 0.

Akt: dy(14) = r14*(((α* $W_{activation}$) ^{n14}) / (1 + (α* $W_{activation}$) ^{n14}) – Akt).

PITP: dy(15) = r15*(((α * $W_{activation}$) ^{n15}) / (1 + (α*$W_{activation}$) ^{n15}) – PITP).

PLC_r: dy(16) = r16*(((α * $W_{activation}$) ^{n16}) / (1 + (α*$W_{activation}$) ^{n16}) – PLC_r).

IP_3: $dy(17) = r17*(((\alpha * W_{activation})^{n17})/(1+(\alpha*W_{activation})^{n17}) - IP_3)$.

DAG: $dy(18) = r18*(((\alpha*W_{activation})^{n18})/(1+(\alpha*W_{activation})^{n18}) - DAG)$.

PI4Ka: $dy(19) = r19*(((\alpha*W_{activation})^{n19})/(1+(\alpha*W_{activation})^{n19}) - PI4Ka)$.

PIa: $dy(20) = r20*(((\alpha*W_{activation})^{n20})/(1+(\alpha*W_{activation})^{n20} +(\alpha*W_{inhibition})^{n20}) - PIa)$.

SACa: $dy(21) = 0$.

PIP_2b: $dy(22)=r11*(((\alpha * W_{activation})^{n22})/(1+(\alpha*W_{activation})^{n11}+(\alpha*W_{inhibition})^{n22}) - PIP_2b)$.

PI4Kb: $dy(23) = 0$.

PIb: $dy(24) = r24*(((\alpha*W_{activation})^{n24})/(1+(\alpha*W_{activation})^{n24} +(\alpha*W_{inhibition})^{n24}) - PIa)$.

SACb: $dy(25) = 0$.

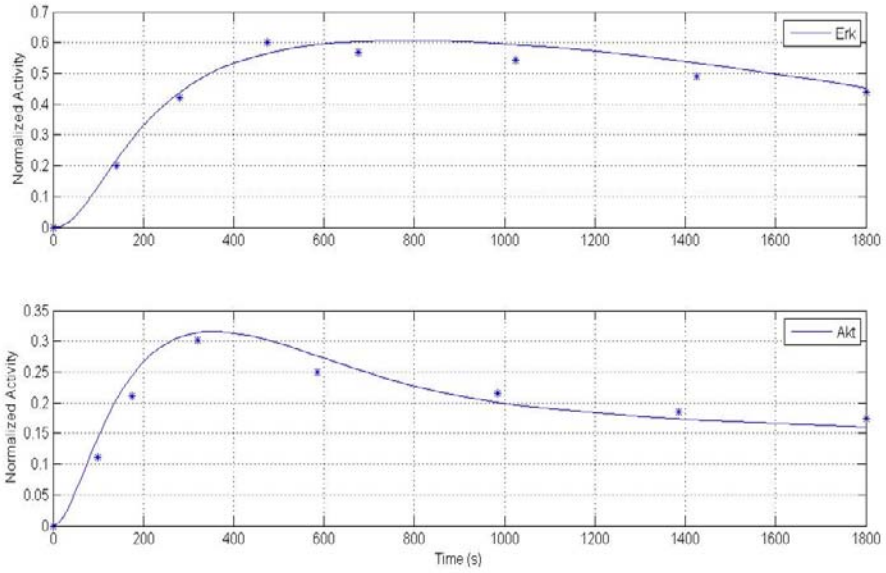

Figure 1: Model A, EGF-stimulated response averaged from selected parameter sets. Stars are data points from experiment.

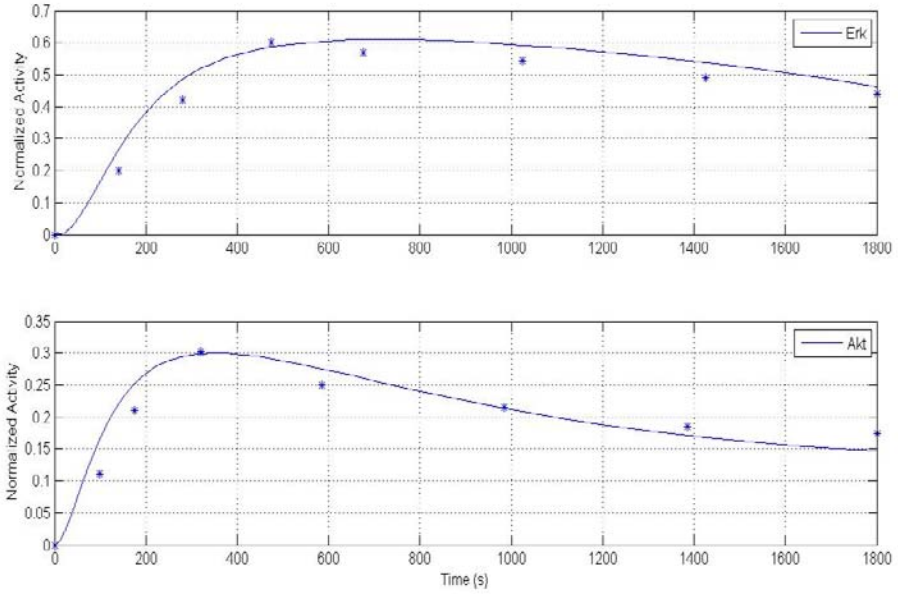

Figure 2: Model C, EGF-stimulated response averaged from selected parameter sets. Stars are data points from experiment.

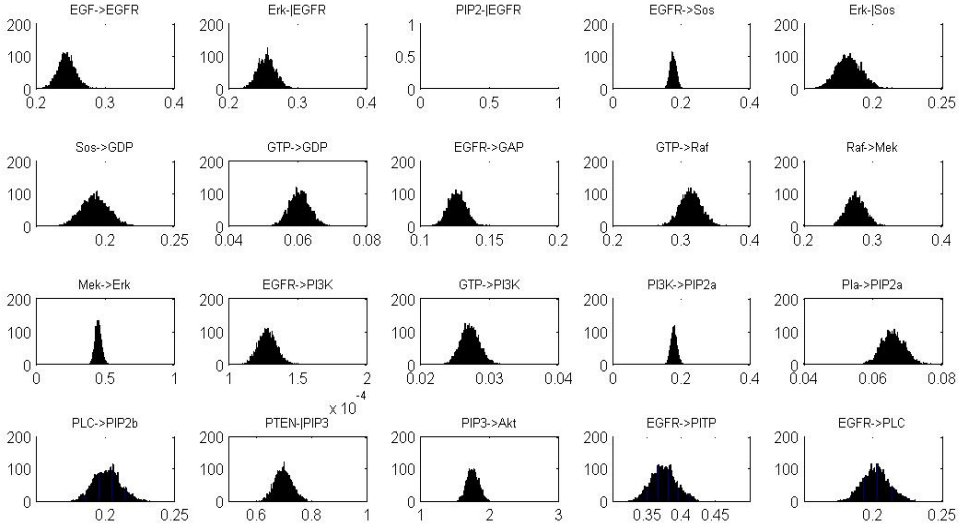

Figure 3: Model A, Histograms of posterior distribution showing the parameter selection convergence. PIP_2-| EGFR is zero.

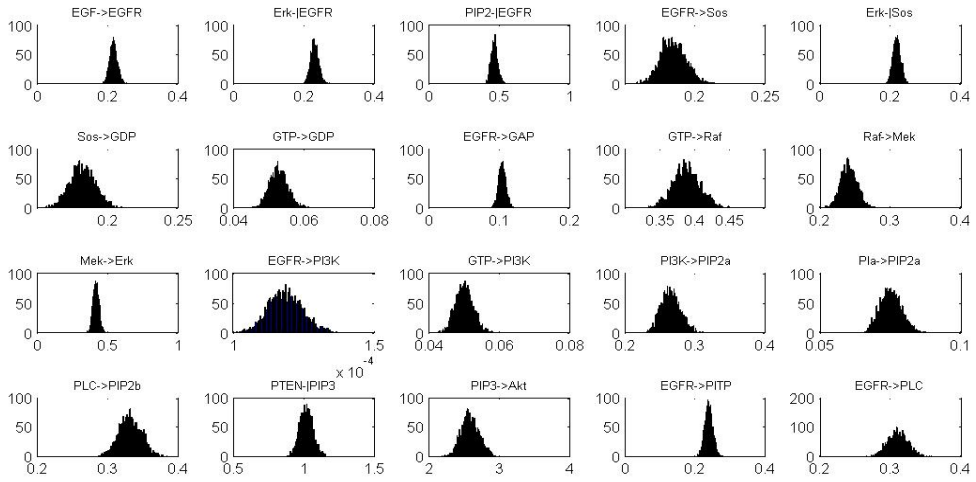

Figure 4: Model C, Histograms of posterior distribution showing the parameter selection convergence.

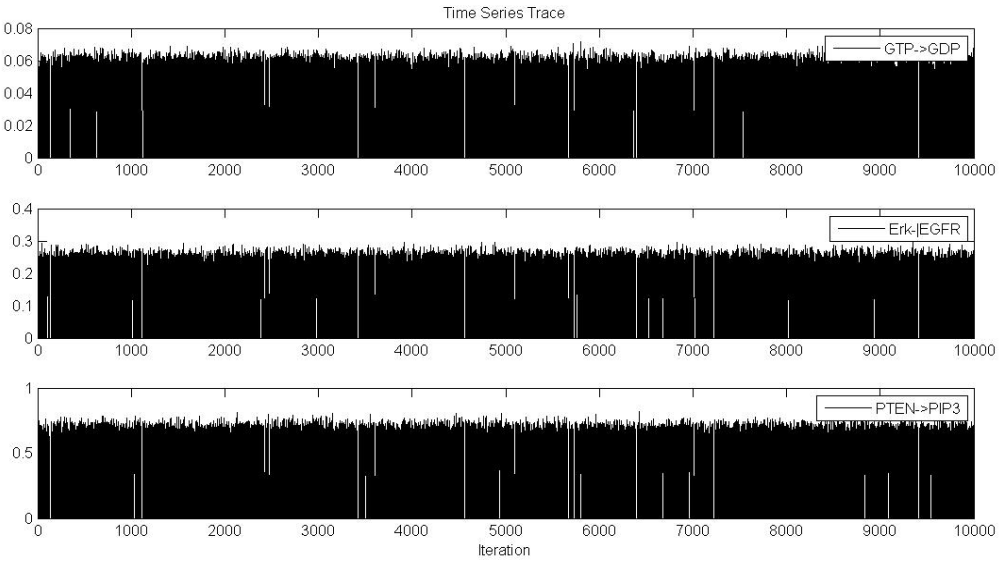

Figure 5: Model A, Time series trace of posterior distribution showing that the selected parameter values are well mixed.

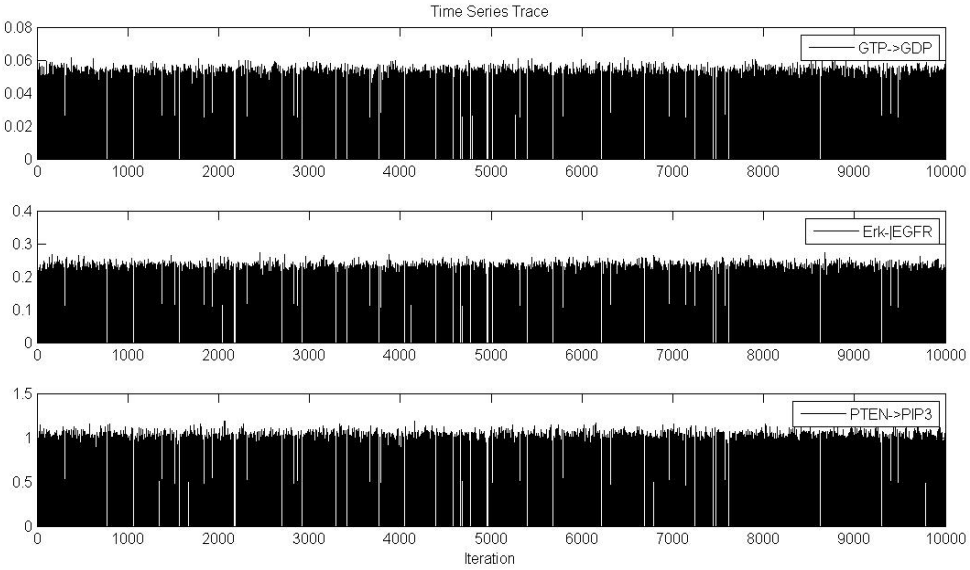

Figure 6: Model C. Time series trace of posterior distribution showing that the selected parameter values are well mixed.

Table 3: Model A, Weight parameter sets selected from simulations. Maxima, minima, means and standard deviations of parameter sets show the convergence. -> sign signifies activation, -| sign signifies inhibition.

	Parameter	Max.	Min.	Mean	Std. dev.
1	EGF ->EGFR	0.2930	0.2031	0.2441	0.0125
2	Erk -> EGFR	0.2987	0.2128	0.2548	0.0127
3	EGFR -> Sos_2	0.2116	0.1480	0.1760	0.0089
4	Erk -| Sos_2	0.2177	0.1541	0.1821	0.0091
5	Sos_2 -| RasGDP	0.2257	0.1637	0.1926	0.0097
6	RasGTP ->RasGDP	0.0719	0.0507	0.0605	0.0029
7	RasGDP ->RasGTP	0.1576	0.1137	0.1328	0.0062
8	RasGAP -| RasGTP	0.0661	0.0474	0.0554	0.0028
9	EGFR ->RasGAP	0.1520	0.1068	0.1265	0.0063
10	RasGTP ->Raf_1	0.3740	0.2589	0.3133	0.0154
11	Raf_1 ->Mek	0.3240	0.2339	0.2744	0.0135
12	Mek ->Erk	0.5799	0.3748	0.4482	0.0224
13	EGFR ->PI3K	0.0002	0.0001	0.0001	0.0000
14	RasGTP ->PI3K	0.0330	0.0219	0.0272	0.0014
15	PI3K -| PIP_2a	0.2141	0.1478	0.1778	0.0089
16	PIP_3 -> PIP_2a	0.7684	0.5478	0.6603	0.0313
17	PIa -> PIP_2a	0.0775	0.0547	0.0656	0.0033
18	PLC_r -| PIP_2a	0.2370	0.1658	0.2012	0.0099
19	PIP_2a -> PIP_3	0.1965	0.1458	0.1712	0.0074
20	PTEN -| PIP_3	0.8196	0.5967	0.6978	0.0324
21	PIP_3 -> Akt	2.0888	1.4766	1.7610	0.0880
22	EGFR -> PITP	0.4647	0.3186	0.3744	0.0190
23	PITP -> PIKa	0.3701	0.2683	0.3200	0.0159
24	PIKa -| PIa	0.0844	0.0593	0.0703	0.0034
25	EGFR -> PLC_r	0.2391	0.1683	0.2017	0.0100
26	PIP_2a -> IP_3	0.2312	0.1617	0.1932	0.0097

Table 4: Model C, Weight parameter sets selected from simulations. Maxima, minima, means and standard deviations of parameter sets show the convergence. -> sign signifies activation, -| sign signifies inhibition.

	Parameter	Max.	Min.	Mean	Std. dev.
1	EGF ->EGFR	0.2525	0.1838	0.2157	0.0109
2	Erk -> EGFR	0.2730	0.1991	0.2298	0.0113
3	PIP_2 -| EGFR	0.5506	0.3916	0.4638	0.0232
4	EGFR -> Sos_2	0.2186	0.1541	0.1846	0.0091
5	Erk -| Sos_2	0.2608	0.1796	0.2164	0.0109
6	Sos_2 -| RasGDP	0.2166	0.1552	0.1821	0.0090
7	RasGTP ->RasGDP	0.0617	0.0455	0.0528	0.0026
8	RasGDP ->RasGTP	0.1918	0.1414	0.1639	0.0075
9	RasGAP -| RasGTP	0.0590	0.0413	0.0494	0.0025
10	EGFR ->RasGAP	0.1237	0.0865	0.1040	0.0052
11	RasGTP ->Raf_1	0.4603	0.3339	0.3879	0.0192
12	Raf_1 ->Mek	0.2820	0.2011	0.2409	0.0118
13	Mek ->Erk	0.5124	0.3451	0.4164	0.0208
14	EGFR ->PI3K	0.0001	0.0001	0.0001	0.0000
15	RasGTP ->PI3K	0.0618	0.0422	0.0501	0.0025
16	PI3K -| PIP_2b	0.3134	0.2182	0.2651	0.0139
17	PIP_3 -> PIP_2b	1.1620	0.8111	0.9737	0.0474
18	PIa -> PIP_2a	0.0895	0.0629	0.0750	0.0038
19	PLC_r -| PIP_2b	0.3851	0.2768	0.3318	0.0163
20	PIP_2b -> PIP_3	0.1549	0.1099	0.1293	0.0064
21	PIb -> PIP_2b	0.2589	0.1871	0.2227	0.0104
22	PTEN -| PIP_3	1.2397	0.8387	1.0160	0.0489
23	PIP_3 -> Akt	3.0673	2.1879	2.5855	0.1275
24	EGFR -> PITP	0.2936	0.1976	0.2389	0.0119
25	PITP -> PIKa	0.5271	0.3722	0.4488	0.0221
26	PIKa -| PIa	0.0922	0.0673	0.0793	0.0038
27	EGFR -> PLC_r	0.3802	0.2570	0.3097	0.0151
28	PIKa -| PIa	0.0995	0.0724	0.0855	0.0042
29	PIP_2b -> IP_3	0.3401	0.2318	0.2879	0.0148

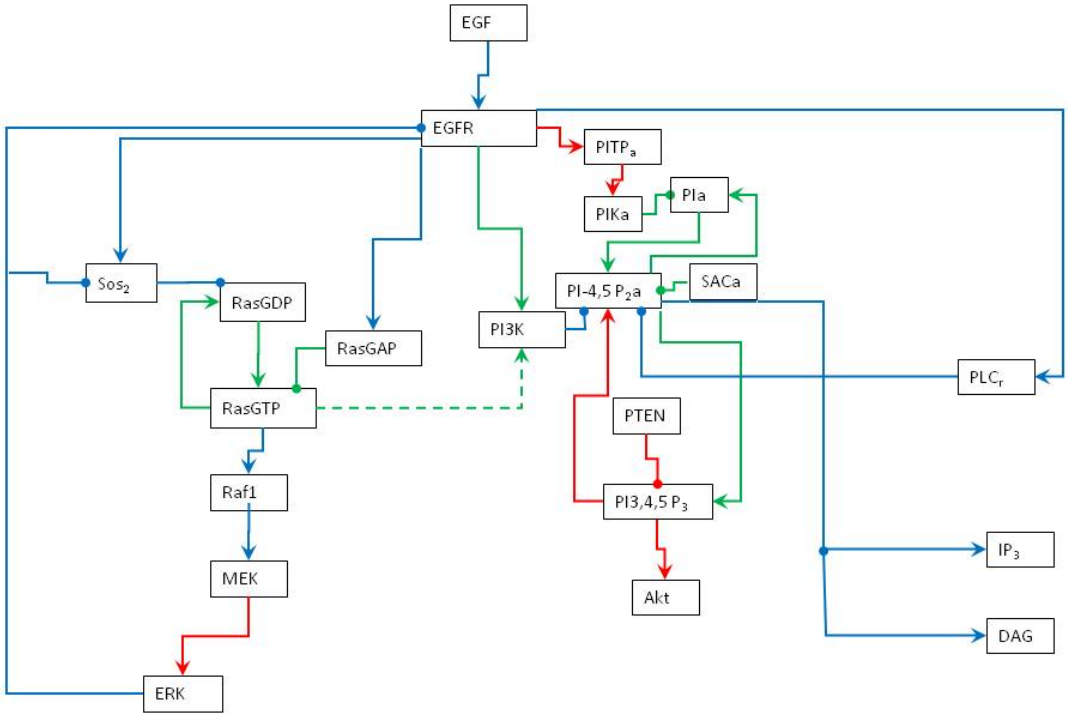

Figure 7: Model A, colored EGFR signaling network indicating the range of the parameter value. Red indicates a standard deviation greater than 0.03; blue signifies a standard deviation between 0.01 and 0.03; green represents a standard deviation less than 0.01.

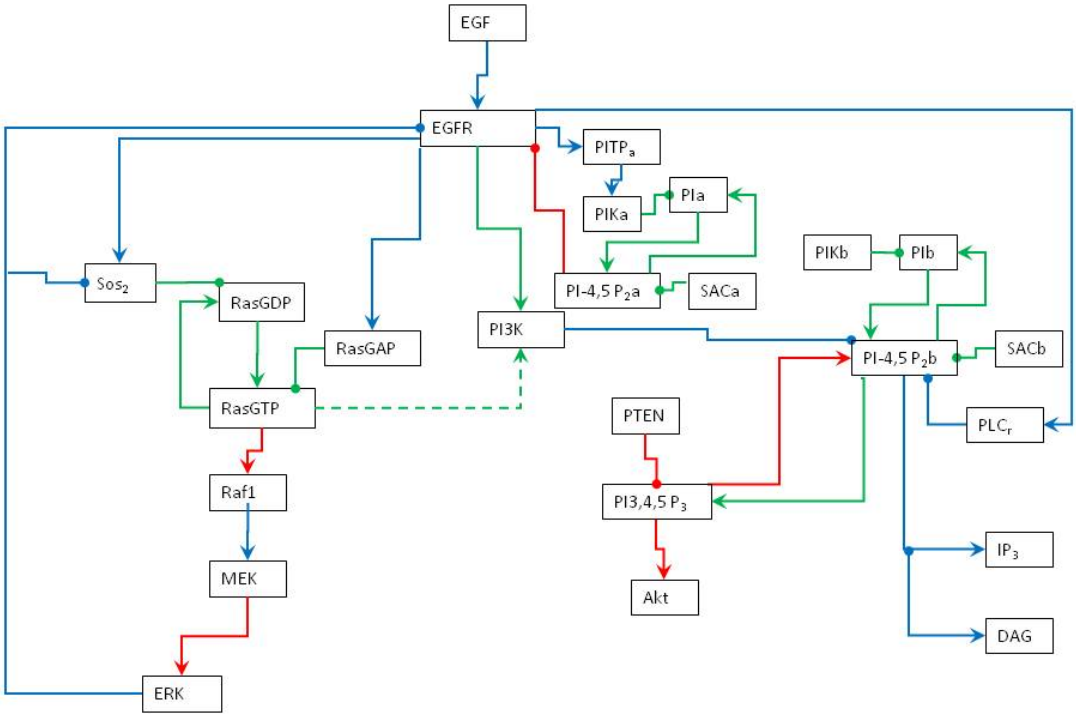

Figure 8: Model C with colored EGFR signaling network indicating the range of the parameter value. Red indicates a standard deviation greater than 0.03; blue signifies a standard deviation between 0.01 and 0.03; green represents a standard deviation less than 0.01.

ACKNOWLEDGMENTS

The study was supported in part by the North Carolina State University Department of Animal Science Enhancement Fund.

REFERENCES

Alberts, B. et al. (2008) Molecular Biology of THE CELL, Fifth Edition.

Aldridge B.B. (2006) Physicochemical modeling of cell signaling pathways. *Nat Cell Biol* 8: 1195–1203.

Baker MD, Wolanin PM & Stock JB (2006) Signal transduction in bacterial chemotaxis. *BioEssays* 28:9-22.

Bellacosa, A., de Feo D, Godwin, A.K., et al. (1995) Molecular alterations of the AKT2 oncogene in ovarian and breast carcinomas. *Int J Cancer*. 64:280–5.

Ben-Shlomo I, Yu Hsu S, Rauch R et al. (2003) Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Sci STKE* 187:RE9.

Berridge MJ (2005) Unlocking the secrets of cell signaling. *Annu Rev Physiol* 67:1-21.

Berridge MJ, Bootman MD & Roderick HL (2003) Calcium signaling dynamics, homeostasis and remodeling. *Nature Rev Mol Cell Biol* 4:517-529.

Bourne HR (1995) GTPases: a family of molecular switches and clocks. *Philos Trans R Soc. Lond B Biol Sci* 349:283-289.

Bradshaw RA & Dennis EA (eds) (2003) Handbook of Cell Signaling. *Elsevier*: St. Louis.

Brown, K.S. et al. (2004) The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys. Biol.* 184-195.

Birtwistle, MR et al (2007) Lignad-dependent responses of the ErbB signaling network: experimental and modeling analysis. *Mol. Sys. Biol.* 3: 144.

Burns ME & Baylor DA (2001) Activation, deactivation, and adaptation in vertebrate photoreceptor cells. *Annu Rev Neurosci* 24:779-805.

Cannon, W.B. (1933) The wisdom of the body.

Barbacid, M. (1987) Ras genes. *Annu. Rev. Biochem.* **56**, 779–827

Bos, J. L. (1989) Ras oncogenes in human cancer: a review. *Cancer Res.* **49**, 4682–4689

Carpenter, C. L., Duckworth, B. C., Auger, K. R., Cohen, B., Schaffhausen, B. S., & Cantley, L. C. (1990). Purification and characterization of phosphoinositide 3-kinase from rat liver. *J Biol Chem,265*(32), 19704-19711.

Carpenter, G. (2000) EGF receptor transactivation mediated by the proteolytic production of EGF-like agonists. *Science STRK* (15), PE1.

Cheng, J.Q., Ruggeri, B., Klein, W.M., et al. (1996) Amplification of AKT2 in human pancreatic cells and inhibition of AKT2 expression and tumorigenicity by antisense RNA. *Proc Natl Acad Sci USA*. 93:3636–41.

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings Algorithm. *The American Statistician, Vol.* 49, No. 4, 327-335.

Cunningham, E., Tan, SK., Swigart, P., Hsuan, J., Bankaitis, V., Cockcroft, S. (1996) The yeast and mammalian isoforms of phosphatidylinositol transfer protein can all restore phospholipase C-mediated inositol lipid signalling in cytosol-depleted RBL-2H3 and HL60 cells. *Proc Natl Acad Sci USA* 1996, **93**:6589-6593.

Dard N & Peter M (2006) Scaffold proteins in MAP kinase signaling: more than simple passive activating platforms. *BioEssays* 28:146-156.

Dong, C. et al. (2002) MAP kinases in the immune response. *Annu Rev Immunol* 20:55.

Downward J (2004) PI 3-kinase, Akt and cell survival. *Semin Cell Dev Biol* 15:177-182.

Ferrell JE. Jr. (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 14:140-148.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* 57:97-109.

Hibshoosh, H., Wigler, M. H., & Parsons, R. (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science, 275*(5308), 1943-1947.

Hill AV (1910) The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *Proc Of The Phys Society*.

Holbro, T. and Hynes, N.E. (2004) ErbB receptors: directing key signaling networks throughout life. *Annu Rev Pharmacol Toxicol* 44: 195–217.

Hood, L. et al. (2004) Systems Biology and New Technologies Enable Predictive and Preventative Medicine**.** *Science* Vol. 306. No. 5696, pp. 640 – 643.

Hsuan, J. and Cockcroft, S. (2001) The PITP family of phosphatidylinositol transfer proteins. *Genome Biology* 2(9): reviews 3011.3-1011.8.

Hudmon A & Schulman H (2002) Structure-function of the multifunctional Ca^{2+}/calmodulin-dependent protein kinase II. *Biochem J* 364:593-611.

Kitano, H. et al. (2000) Fundations of systems biology.

Kolch, W. et al. (2005) When kinases meet mathematics: the systems biology of MAPK signalling. *FEBS Letters* 579: 1891–1895.

Li, D. M., & Sun, H. (1997). TEP1, encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor beta. *Cancer Res, 57*(11), 2124-2129.

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S. I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., Bigner, S. H., Giovanella, B. C., Ittmann, M., Tycko, B.,

Liu, Y. and Bankaitis, V.A. (2010) Phosphoinositide phosphatases in cell biology and disease. *Progress in Lipid Research* **49:**201-217.

Luttrell LM (2006) Transmembrane signaling by Gprotein-coupled receptors. *Methods Mol Bio* 332:3-49.

Metropolis, N. et al (1953) Equations of state calculations by fast computing machines. *J. of Chemical Physics* 21:1087-1091.

Mitin N, Rossman KL & Der CJ (2005) Signaling interplay in Ras superfamily function. *Curr Biol* 15:R563-574.

Moghal, N. and Sternberg, P.W. (1999) Multiple positive and negative regulators of signaling by the EGF-receptor. Curr. Opin. Cell Biol. 11, 190-196.

Mullschleger S, Leowith R & Hall MN (2006) OR signaling in growth and metabolism. *Cell* 124:471-484.

Ogunnaike, B.A., Ray, W.H. (1994) Process Dynamics, Modeling, and Control. New York: Oxford University Press.

Papin JA, Hunter T, Palsson BO & Subramaniam S (2005) Reconstruction of cellular signaling networks and analysis of their properties. *Nature Rev Mol Cell Biol* 6:99-111.

Parker PJ (2004) The ubiquitous phosphoinositides. *Biochem Soc Trans* 32:893-898.

Pawson T (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* 116:191-203.

Pawson T & Scott JD (2005) Protein phosphoryation in signaling – 50 years and counting. *Trends Biochem Sci* 30:286-290.

Pierce KL, Premont RT & Lefkowitz RJ (2002) Seven-transmembrane receptors. *Nature Rev Mol Cell Biol* 3:639-650.

Pires-daSilva A & Sommer RJ (2003) The evolution of signaling pathways in animal development. *Nature Rev Genet* 4:39-49.

Reddy, C.C. et al. (1994) Proliferative response of fibroblasts expressing internalization-deficient epidermal growth factor (EGF) receptors is altered via differential EGF depletion effect. *Biotechnol. Prog.* 10, 377-384.

Reiter E & Lefkowitz RJ (2006) GRKs and beta-arrestine roles in receptor silencing, trafficking and signaling. *Trends Endocrino Metab* 17:159-165.

Rhee SG (2001) Regulation of phosphoinositide-specific phospholipase C. *Annu Rev Biochem* 70:281-312.

Robishaw JD & Berlot CH (2004) Translating G protein subunit diversiy into functional specificity. *Curr Opin Cell Biol* 16:206-209.

Roskoski R Jr (2004) Src protein-tyrosine kinase structure and regulation. *Biochem Biophys Res Commun* 324:1155-1164.

Qi M & Elion EA (2005) MAP kinase pathways. *J Cell Sci* 118:3569-3572.

Sahin M, Greer PL, Lin MZ et al (2005) Eph-dependent tyrosine phosphorylation of ephexin1 modulates growth cone collapse. *Neuron* 46:191-204.

Schlessinger J (2000) Cell signaling by receptor tyrosine kinases. *Cell* 103:211-225.

Schwartz MA & Madhani HD (2004) Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. *Annu Rev Genet* 38:725-748.

Science's Signal Transduction Knowledge Environment (Stke): www.stke.org

Seet BT, Kikic I, Zhou MM & Pawson T (2006) Reading protein modifications with interaction domains. *Nature Rev Mol Cell Biol* 7:473-483.

Shaw RJ & Cantley IC (2006) Ras, PI(3)K and mTOR signaling controls tumour cell growth. *Nature* 44:424-430.

Shaywitz AJ & Greenberg ME (1999) CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annu Rev Biochem* 68:821-861.

Singla V & Reiter JF (2006) The primary cilium as the cell's antenna signaling at a sensory organelle. *Science* 313:629-633.

Staal, S.P. (1987) Molecular cloning of the akt oncogene and its human homologues AKT1 and AKT2: amplification of AKT1 in a primary human gastric adenocarcinoma. *Proc Natl Acad Sci USA.* 84:5034–7.

Stambolic, V., Suzuki, A., de la Pompa, J. L., Brothers, G. M., Mirtsos, C., Sasaki, T., Ruland, J., Penninger, J. M., Siderovski, D. P., & Mak, T. W. (1998). Negative regulation of PKB/Aktdependent cell survival by the tumor suppressor PTEN. *Cell, 95*(1), 29-39.

Steck, P. A., Pershouse, M. A., Jasser, S. A., Yung, W. K., Lin, H., Ligon, A. H., Langford, L. A., Baumgard, M. L., Hattier, T., Davis, T., Frye, C., Hu, R., Swedlund, B., Teng, D. H., & Tavtigian, S. V. (1997). Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nat Genet, 15*(4), 356-362.

Sun, H., Lesche, R., Li, D. M., Liliental, J., Zhang, H., Gao, J., Gavrilova, N., Mueller, B., Liu, X., & Wu, H. (1999). PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and Akt/protein kinase B signaling pathway. *Proc Natl Acad Sci USA. 96*(11), 6199-6204.

Tanaka, S. and Hosaka, K. (1994) Cloning of a cDNA encoding a second phosphatidylinositol transfer protein of rat brain by complementation of the yeast *sec14* mutation. *J Biochem* **115:**981-984.

van der Geer, P. et al. (1994) Receptor proteintyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell Biol.* 10, 251–337.

Vivanco, I., Sawyers, C.L., (2002) The phosphatidylinositol 3-kinase AKT pathway in human cancer. *Nat Rev Cancer*. 2:489–501.

Walsh, B. (2004) Markov Chain Monte Carlo and Gibbs Sampling. *Lecture Notes for EEB 581*.

Wassarman DA, Therrien M & Rubin GM (1995) The ras signaling pathway in Drosophila. *Curr Opin Genet Dev* 5:44-50.

Wells, A. (1999) EGF receptor. *Int. J. Biochem. Chell Biol.* 31, 637-643.

Wiley, H.S. et al. (2003) Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol* 13: 43–50.

Chapter Six **Future Direction**

INTRODUCTION

When there is limited amount of data available in cell signaling network, the Hill function and the Markov Chain Monte Carlo (MCMC) methods can model the network without having biochemical reaction information. The model is constructed on the information of signaling direction, and is able to provide insight on the network and make predictions qualitatively.

In the future, we plan to improve this method by setting some parameters as fixed values, and select other parameter values. This modification may provide more balanced and more quantitative details on the dynamic behavior of the network.

We hope to collaborate with experimental researchers who use our model to guide experiment design and use the experimental data to fine tune our model. This approach will permit model predictions to be validated by experimental data.

We would like to develop more modeling methods for cell signaling, neurobiology, and cancer research. We hope to add more modeling methods to our repertoire.

To gain greater efficiency, we suggest that parallel simulations be performed whenever possible. Parallel simulation can be data parallel, thread parallel, or component parallel. Data parallel involves dividing a big data file (one or several million sets) into several data files and simulate them at the same time. Thread parallel is accomplished by dividing a big simulation into several subprograms that do not have dependency (numerator and denominator) and running the subprograms in parallel. Component parallel involves decomposing a big simulation into different modules which belong to different levels.

Modules are then simulated at the same level in parallel and the results passed back to modules.

We also suggest that a combination of modeling methods can be used in cell signaling network modeling. The cell signaling network can be divided into components or modules depending on how detail we want to model different parts of the network. The different modeling methods can be applied to different components. This approach will allow more details for some parts of the cell signaling network and speed up simulations for other parts of the network.