Cell Memory in the Mitogen-Activated Protein Kinase Signaling Pathway

Eugenia Lyashenko

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

under the Executive Committee

of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

Cell Memory in the Mitogen-Activated Protein Kinase Signaling Pathway

Eugenia Lyashenko

Abstract

Cells process information from their environment, such as the stimuli to grow, divide, or die, via cell signaling. Deregulated processing of extracellular stimuli can lead to aberrant cell responses and cause cancer. Given that the *in vivo* cell environment constantly changes, it is important to understand how cells incorporate the context of their environment into their decision making processes.

The idea of responding to relative, not absolute, changes in stimuli was first proposed in studies of human perception and became known as Weber's Law. Although, evidence of Weber's Law at the molecular level has been previously presented in studies of several organisms, to the best of our knowledge, it has never been explored in the case of relative sensing of extracellular stimuli in mammalian signaling cascades.

The Mitogen-Activated Protein Kinase (MAPK) signaling pathway has been implicated in multiple human diseases, including cancers, and therefore cell signaling through this pathway is an important subject of research. Here we present a theoretical framework and an experimental validation of the mechanism of Weber's Law in the ability of cells to sense relative changes in the levels of extracellular stimuli in the MAPK signaling pathway. In particular, in this work we consider relative sensing in levels of Epidermal Growth Factor (EGF) in the MAPK pathway.

We derive an analytical model of steady state behavior of the MAPK signaling pathway stimulated with constant doses of EGF. We demonstrate a mechanism that produces phosphorylation responses proportional to relative changes in ligand concentrations. The mechanism of Weber's Law presented here entails the retention of memory of the dose of the past chronic stimulation with EGF. The molecular mechanisms responsible for Weber's Law in MAPK signaling are likely to contribute to many other receptors signaling systems. Therefore, the mechanism of relative sensing of extracellular

ligand concentrations derived here can be generalized beyond the EGF-activated MAPK signaling pathway to many other cell signaling systems.

This thesis also presents a probabilistic framework to explore the parameter space of a detailed mechanistic ODE model of EGFR signaling cascades. The application of the model simulation allows us to generate probabilistic predictions of EGFR system behavior and to explore structure-to-function relationships between the model's parameter space and EGFR system responses.

Overall, this work suggests an alternative view on the role of cellular endocytosis in the MAPK signaling *in vivo*. Specifically, traditionally viewed as a mechanism to downregulate and terminate cell signaling, endocytosis may enable cells to dynamically adjust their sensitivity to extracellular stimuli, and hence allow cells to integrate information about the past stimulations into the cell responses to the consequent stimulations and thus, cell fate decisions.

Table of Contents

vi

LIST OF GRAPHS, IMAGES, AND ILLUSTRATIONS

LIST OF TABLES

ACKNOWLEDGEMENTS

DEDICATION

This thesis is dedicated to my dear parents, Tetyana Loginova and Borys Lyashenko – I'm very grateful, I love you and I admire everything that you are! Also, this thesis is dedicated to the memory of my grandfather, Krinitskiy Aleksey Fedorovich, a medical surgeon during WWII, a talented biochemist and doctor of infectious diseases in peaceful time – I wish we could have many interesting discussions on subjects at the intersection of biology and medicine.

Chapter One: Introduction to Cell Signaling in Mammalian Cells and Relative Sensing of EGF in EGFR Signaling

Cell signaling is crucial to the proper functioning and survival of organisms, just as language is essential for the proper functioning of human societies. Cell signaling enables cells to sense information in their environment and communicate with other cells. Although manifested differently, cell signaling is present across all types of organisms, ranging from simplest unicellular organisms such as yeast and bacteria, to higher organisms, such as humans. Yeast, for example, signal to each other by secreting and responding to pheromones mating factors (Herskowitz, 1995). Bacteria use cell signaling to sense nutrient gradients and move to their source, thus gaining evolutionary advantage in stringent conditions (Mesibov & Adler, 1972). Higher organisms consist of multiple types of tissues. As such, their cells are expected to communicate among themselves and to respond in a coordinated way to stimuli in order to maintain tissue identity and homeostasis. Hence, one would expect cells in higher organisms to have a much more sophisticated machinery of signaling cascades to allow complex communication than do lower organisms.

It has been suggested that developing elaborate signaling cascades and the ability of cells to respond in a coordinated way allowed evolution from unicellular to multicellular organisms (Alberts B, Johnson A, Lewis J, et al, 2002). Through gene duplication and divergence, higher organisms were able to develop elaborate signaling cascades which then allowed cells to carry out diverse cellular functions and respond efficiently to changing environment facing the organism, thus providing survival advantage.

Multiple studies have elucidated the biochemical details of signaling cascades in mammalian cells. However, much less is known about how quantitative information from cell environment is translated in cell decision through cell signaling in mammals (Albeck, Mills, & Brugge, 2013). In addition, it is still not well understood how cells manage to reliably perform their functions while operating in chaotic environments with diverse stimuli. How does cell signaling accurately relate information from cell exterior into cell decisions despite the molecular noise? What features of the signaling responses are most important to the cell? What mechanisms do cells use to control the

signaling? Even less is known about quantitative aspects of cells signaling and how cells encode those features. The main challenge is that many of the cell mechanisms are intertwined dynamic processes. The overall picture is further complicated by the fact that it is still difficult to measure time-varying abundances or signaling molecules *in vivo* (Fisher & Lakshmanan, 1990; Goentoro et al., 2006).

Cell signaling is mediated by signaling molecules secreted by cells throughout the organism. Signaling molecules, such as various growth factors, are recognized and bound by designated cell-surface receptors. Upon ligand binding, receptors undergo activation and initiate series of reactions in the cell interior, ultimately, leading to changes in metabolism, cell shape, motility, or expression of the genes necessary to carry out the response intended by the signaling molecule (Alberts B, Johnson A, Lewis J, et al, 2002). Interestingly, cell surface receptors serve as an interface between cells and their environment and establish specificity and precision between signaling molecules and elicited responses (Shankaran, Resat, & Wiley, 2007).

In higher organisms there are multiple receptor-enabled sensory systems. G-protein coupled receptors are an example of the largest family of cell-surface receptors. Although they are found in all eukaryotes, in higher organisms the GPCR mediate a myriad of diverse functions such as vision and olfactory sensing in human (Alberts B, Johnson A, Lewis J, et al, 2002). Many different signaling molecules can bind to these receptors; in addition, the same ligand can activate multiple types of receptors (Collins, Caron, & Lefkowitz, 1992). Moreover, signaling cascades initiated by these receptors can overlap on multiple intermediate components as well as on downstream targets. Overall, the reuse of the signaling components in many different signaling cascades could allow cells to implement more efficiently various functions. In addition, the differential dynamics of target's responses might distinguish between different stimulants in cell environment, and thus encode different cell decisions. For example, in PC12 cells, pERK can be activated by different ligands (EGF and NGF) (Sasagawa, Ozaki, Fujita, & Kuroda, 2005a), and the produced responses may differ in their dynamics (transient vs sustained), leading to different cellular phenotypes (growth vs differentiation). Hence, since cell signaling serves as a mediator between cell environment and cell

2

fate, regulating signaling response is important for cell survival in a time- and composition-varying environment. What are some underlying mechanisms which allow cells to function effectively?

Cells have developed various mechanisms to cope with changing extracellular stimuli (Lodish H, Berk A, Zipursky SL, et al, 2000) and to retain the memory of previous stimulations. At the signaling molecule level, cells can buffer effective duration and doses of the stimulants either by converting short-lasting treatments into long-lasting response or by depleting stimuli from cell environment attenuating treatment. Ligand depletion can be mediated through ligand-induced internalization and degradation of the cell surface ligand-receptor complexes. For example, Erythropoietin receptors, which maintain production of red blood cells, effectively internalize Epo ligand from the cell surface and this mechanism is partially what enables EpoR to sense a very wide range of Epo concentrations (Becker et al., 2010).

Amplification of levels of intermediate signaling components helps to extend the effect of signaling molecules long after the stimulus has been cleared from the cell surface. Such amplification can be realized through action of positive feedback loops activated in response to ligand treatment. Ca2+ regulated kinase cascades serve as an example of cell retention of information about the stimulus (Alberts B, Johnson A, Lewis J, et al, 2002).

Another way to realize cell memory is via turning on gene transcription in response to ligand stimulation. As a downstream target of signaling cascades, gene transcription thus extends the effects of signaling molecule to the longer time scale necessary to produce gene response.

Alternatively, to deal with extracellular stimuli, cells can regulate their own signaling components. Specifically, if the signaling molecules persist in cell environment, the cells can attenuate and adapt its responses to the constant level of stimulation. Adaptation or desensitization is defined as a reversible process in which the prolonged stimulation with constant levels of input over time decreases cell's response to that level of stimuli (Alberts B, Johnson A, Lewis J, et al, 2002). Adaptation would allow cells to respond to relative, rather than absolute, changes in stimulus and would allow cells to maintain their sensitivity over wide range of background stimuli doses.

One cell mechanism to implement adaptation is ligand-induced temporarily inactivation of receptors or downstream signaling components via inhibitory post-translational modification (Segel,

Goldbeter, Devreotes, & Knox, 1986) (Figure 1). For example, cell surface receptors in bacterial chemotaxis in response to sensing nutrient become methylated, which reduces their activity and ultimately, decreases the rate of bacterial tumbling in response to stimuli (Alon, Surette, Barkai, & Leibler, 1999). Thus, through adaptation, bacteria cells are able to slow down at a higher concentration of nutrient, independently of the absolute levels at the source.

Another mechanism of desensitization and adaptation acts through production of inhibitory proteins downstream in the signaling cascade. Ligand-induced gene expression of Dual Specificity Phosphatases (Avraham & Yarden, 2011) which dephosphorylate components of the activated MAPK kinase cascade, thus can lead to attenuation of signaling through the cascade.

In addition, cells can adapt and adjust their sensitivity to subsequent stimulations with the same ligand by altering the number of cell surface receptors specific to the ligand (Alberts B, Johnson A, Lewis J, et al, 2002). This adjustment presents one common and effective way to control cell adaptation to extracellular stimuli on multiple downstream targets by regulating the signaling cascade at the receptor level, rather than at the level of multiple individual downstream targets (although the two mechanisms are not mutually exclusive).
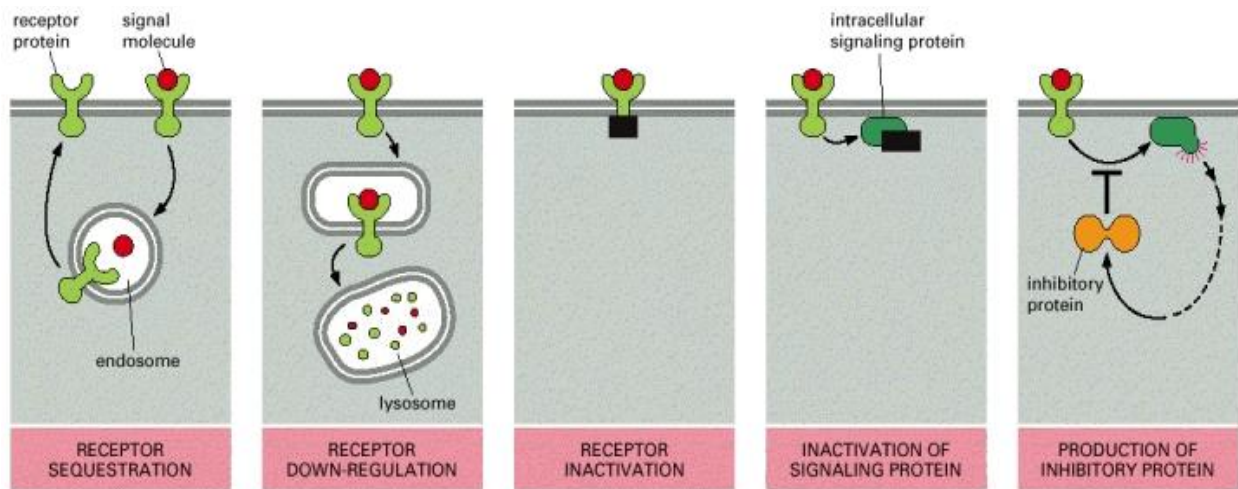
Figure 1.1: Five ways in which target cells can become desensitized to a signal molecule. Adapted from Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002. "General Principles of Cell Communication".

Receptor internalization, recycling and degradation are molecular processes which regulate the number of cell surface receptors. Ligand binding can induce rapid internalization of the ligand-

4

receptor complex, followed by either recycling or degradation of receptors. For some receptor sensory systems stimulations with ligand can lead to receptors mobilization to cell surface, as in the case of EpoR signaling (Becker et al., 2010), whereas in other cases, it can lead to removal of receptors from the cell surface, as in EGF-induced ErbB signaling (Citri & Yarden, 2006). The outcomes of such regulation depend on multiple entangled dynamic processes such as receptor activation, internalization, recycling, degradation, kinetics of which, in turn, is determined by the biochemical properties of ligand-receptor interactions and other reactions.

As was previously mentioned, desensitization is a reversible phenomenon, thus, upon removal of the stimulus from the cell environment, cells can recover their sensitivity to subsequent stimulations. Time to recover the sensitivity can depend on the properties of the sensory system and may affect the precision with which cells respond to subsequent or overlapping signals. Hence, regulating the amount of receptors on the cell surface can help cells be attuned to their environment over time. Since levels of extracellular stimuli can change rapidly, whereas cell response can take longer time scale to the time necessary for gene expression, it might be important to have machinery to synchronize and coordinate cell decisions to the available stimuli on cell surface. Moreover, understanding the quantitative relationship between the responses of downstream targets and cell fate outcomes can help elucidate mechanisms of cell functioning *in vivo* when cells are exposed to repeated stimuli over time.

In this thesis we study quantitative aspects of cell signaling in mammalian cells, focusing on the case of signaling through Epidermal Growth Factor (EGF) Receptors, also referred to as ErbB receptors. ErbB receptors, among other ligands on the cell exterior, can bind epidermal growth factor (EGF), a molecule stimulating cells to grow and divide. Since the ErbB receptors regulate cell growth and division, they have been implicated in multiple cancers and are, thus, an important therapeutic target. It is essential to understand how cells process consecutive signals to grow and divide, as over-responding to them might lead to aberrant cell growth and sometimes to cancer (Yarden, 2001).
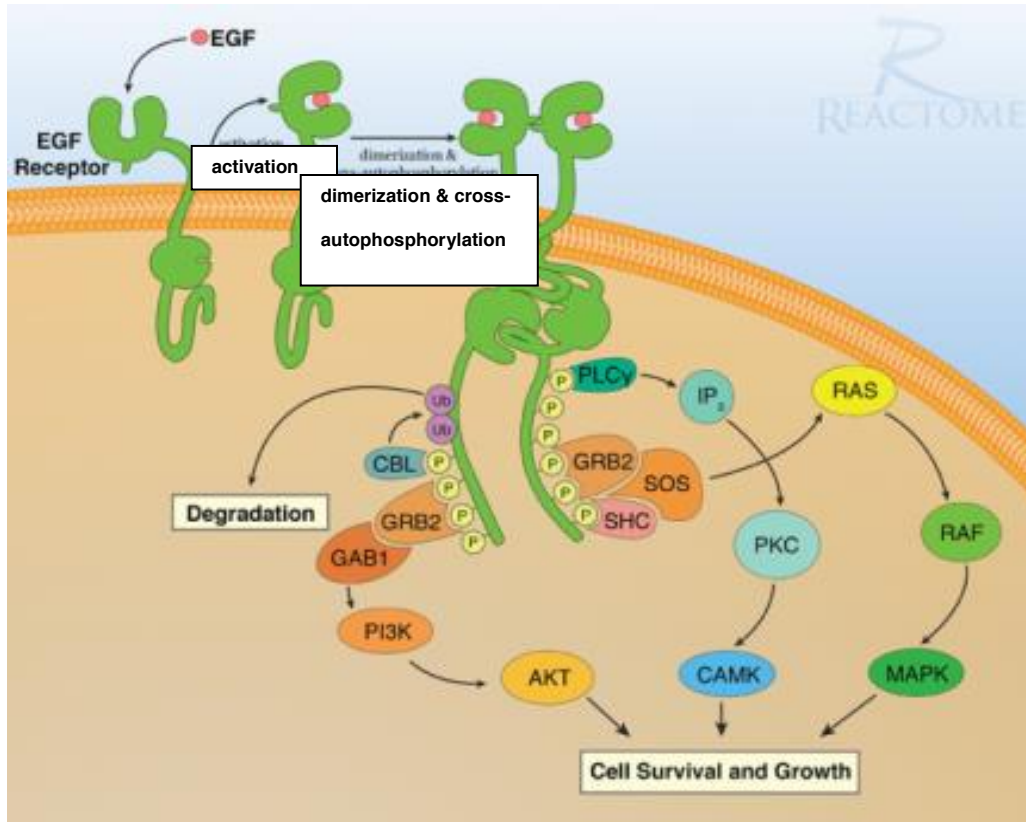
Figure 1.2: Schematics of signaling through EGFR receptors

(source: http://www.ebi.ac.uk/training/online/course/reactome-quick-tour/what-reactome)

Epidermal Growth Factor (EGF) belongs to the family of growth factors – the signaling proteins cells use for communication. They mostly operate as local intermediaries at low concentrations (about 10^{-9}-10^{-11} M). The responses to growth factors take time on the order of hours and generally require intracellular signaling which leads to changes in gene expression (Alberts B, Johnson A, Lewis J, et al, 2002).

The ErbB family receptors are receptor tyrosine kinases. Each receptor consists of an extracellular ligand-binding domain, a transmembrane domain, and a tyrosine kinase domain on the intracellular side (Citri & Yarden, 2006). Upon ligand binding, ErbB receptors form dimers with other ErbB receptors. This leads to autophosphorylation of their tyrosine kinase domains and receptor activation. Active receptor dimers then recruit scaffold and adaptor proteins, thus inducing a series of downstream phosphorylation reactions. ErbB signaling ultimately results in the activation of AKT and MAPK/ERK – important regulators of cell fate. In various cellular contexts, EGF stimulation of ErbB

6

can produce responses with various dynamic profiles and lead to distinct cell decisions. However, it is still not well understood how exactly ErbB signaling dynamics encode cell fate decisions and what are the fundamental quantitative features of its signaling.

ErbB receptor internalization and dephosphorylation by phosphatases are the main known mechanisms controlling ErbB signaling dynamics at the receptor level. A key characteristic feature of the EGFR internalization is that the rate of EGF-induced internalization of receptors is several times faster than the rate of basal ligand-free internalization of receptors. Hence, fast ligand-induced internalization leads to the rapid attenuation of signaling. Depending on the EGF ligand dose and the rate of receptors production, removal of significant number of receptors can render cells temporarily insensitive to consecutive EGF stimuli. This insensitivity can be advantageous if cell is in the middle of responding to previous stimuli to grow, in which case processing an additional activation signal might be redundant. Traditionally, rapid endocytosis of active receptors has been viewed as a mechanism to completely desensitize the cells to consequent ligand stimulations. Such view was derived based on cell signaling studies in which cells were treated with very high, non-physiological levels of stimuli, leading to activation of most receptors and their removal from cell surface (Becker et al., 2010). However, *in vivo*, cells are more likely to be exposed to lower doses of stimuli and thus, it is more plausible that the biological role of endocytosis is not to completely desensitize cells to re-stimulation, but to adjust cell sensitivity to consequent ligand stimulations taking into account previous ligand stimulations.

Since internalization and degradation of receptors requires an assembly of a scaffold and vesicles to transport receptors, this mechanism of receptor downregulation operates on the timescale of minutes to an hour (Sigismund et al., 2012). Dephosphorylation by phosphatases, on the other hand, is a rapid process taking seconds (Kleiman, Maiwald, Conzelmann, Lauffenburger, & Sorger, 2011). Various phosphatases are known to act on phosphor-EGFR throughout different cellular compartments. Overall, phosphatases are considered to be important in inhibiting receptor activation at low levels of stimuli, thus preventing the system from reacting to extracellular noise (Reynolds, Tischer, Verveer, Rocks, & Bastiaens, 2003). Interestingly, phosphatases can decrease the amount of the internalized receptors, as inactive receptors are internalized more slowly. It is still not clear how

exactly internalization and dephosphorylation cooperate together in downregulating ErbB signaling receptors and what are the possible advantages of using both mechanisms to affect cell functioning in its environment.

Given the complexity of the regulation of EGFR signaling through various dynamic mechanisms, as well as the intricacies in assessing ErbB system experimentally, it is, thus, not surprising that until now there were not many quantitative studies focusing on physiological properties of EGFR signaling, going beyond short-lived transient phosphorylation responses and gene expression.

Our main result is a discovery that cells can retain memory of EGF doses during signaling via the number of remaining EGFR receptors on the cell surface, the quantitative description of this feature of cell behavior and a property of relative sensing in input levels emergent from the presented cell memory phenomenon. Through a combination of experimental and modeling approaches, we show how the abundance of EGFR is able to retain information about the EGF doses in chronic stimulations with EGF. This non-trivial quantitative relationship led us to hypothesize that EGFR signaling might be attuned to sense relative, not absolute, increases in EGF levels. By regulating number of EGF receptors on cell surface, cells are able to adjust their sensitivity to extracellular EGF signals. Hence, in this way cells are able to eventually desensitize, or decrease their responses to constitutively present doses of ligand, and respond only to fold-changes in the background ligand doses.

We develop an analytical model of steady state EGFR signaling and with it we derive a mechanism of relative sensing in EGF levels, arising from the property of EGF memory retention. In addition, we develop a detailed mechanistic ODE model and use probabilistic methods to address the problem of model parameter space unidentifiability and generate probabilistic predictions of system behaviors. Finally, we validate predictions of the two models experimentally.

Ability to respond to relative, and not absolute, changes in levels of stimuli is described by Weber's Law and was originally postulated in studies of human perception (Weber, 1905). Evidence of Weber's Law in biological systems other than human perception has been previously reported.

8

In the late 19th century, German physician Ernst Heinrich Weber conducted quantitative experiments on human perception (Weber, 1905). Weber postulated that our ability to perceive absolute changes in background stimuli diminishes as background magnitude of the stimuli increases. Weber's Law in sensory systems states that the just noticeable change in stimulus, required to elicit system response, is proportional to the background of the stimulus. Weber's Law can be formulated as follows. Let ΔR be amount of stimulation that needs to be added to produce detectable difference in response, and let R be the amount of existing stimulation, then

$$\Delta R/R = k$$

where k is a constant specific to the considered sensory system.

Weber's scholar, Theodor Gustav Fechner, generalized the initial law and derived that the maximal amplitude of the response to a change in input stimulus is inversely proportional to the background input level. In other words, the Weber-Fechner's Law suggests that our perception of stimulus intensity is proportional to logarithm of the stimuli. Mathematically, the law can be represented as:

$$y_{max} = c * \Delta R/R$$

where y_{max} stands for the maximum response to a change in signal ΔR and c is a constant.

Weber's and Fechner's work indicates that the relative sensing applies to most human senses of perception, and each particular sensory system will have its own specific proportionality constant, called *Weber's Fraction*. For example, in weight sensing this constant is 0.03, meaning that one is unable to distinguish between distinct weights unless they differ by at least 3% from each other. In the case of vision this constant is 0.01, meaning that one is unable to distinguish between lengths of two lines unless they differ by at least 1% (Nachev, Stich, & Winter, 2013).

In this thesis, we will use mostly the second interpretation of the Weber's Law, concerning the maximal amplitude of the responses. Note that this interpretation concerns only maximal amplitude of the responses, and does not require that the times when these maximums are achieved are the same.

We quantify cell responses to EGF by measuring activation of the downstream EGFR targets, pERK and pAKT. Indeed, we find that despite drastic differences in pERK responses to widely varied

doses of the chronic initial EGF stimulation, after the system adapts, pERK responses to same fold-change inputs are very similar within the physiological range of initial EGF treatments. Thus, being able to generate similar pERK responses to same fold-changes in EGF can allow neighboring cells coordination in cell decisions in spite of variations in background EGF levels, which might occur, for example, due to uneven distribution of ligand in tissues.

It is biologically plausible that the machinery behind EGF-induced signaling has evolved to be attuned to operate in physiological levels of EGF in mammalian tissues and implement ratio-sensing in EGF stimuli. By implementing ratio-sensing in the growth hormone levels, cells in the tissue get advantage in the face of time-varying and noisy EGF levels and hence might benefit the entire organism.

In conclusion, this thesis demonstrates how by studying dynamics of EGFR signaling we discovered a non-trivial quantitative feature of this singaling system. Specifically, we revealed how retaining information about past EGF doses in the number of remaining EGFR receptors enables cells to carry out relative sensing of EGF levels. The implications of this thesis will lead to better understanding of the quantitative aspects of mammalian cell signaling to the *in vivo* cells environment of time-varying inputs, and thus help to design better therapeutic strategies optimized to cell functioning in tissues.

Chapter 2 of this thesis gives an overview of biological concepts related to EGFR receptors, EGF in physiological conditions, general background on cell signaling and some previous quantitative studies of receptors signaling systems. Chapter 3 describes construction of the detailed mechanistic ODE model of ErbB signaling and the computational methods used in this work to sample and explore model parameter space, including Markov Chain Monte Carlo simulations. Chapter 4 presents main results of the thesis: first, it describes probabilistic predictions of ErbB system behaviors from simulating the detailed mechanistic ODE model of ErbB signaling; second, it presents analytical model of ErbB system at steady state to chronic stimulation with EGF and derives Weber's Law property in sensing EGF inherent to the simplified analytical model based on several fundamental, but general assumptions regarding signaling through activated EGFR receptors. Chapter 5 discusses the broader implications and significance of the present work to quantitative

studies of cell memory and to the generalizability of the discovered phenomenon of Weber's Law in sensing levels of extracellular stimuli to other receptor sensory systems, such as signaling through GPCR family of receptors.

Chapter Two: Background on EGFR signaling

Section 2.1: The ErbB receptors

The Epidermal Growth Factor family of receptors, also referred to as the ErbB receptors, is a family of receptor tyrosine kinases implicated in many cancers. The ErbB family of receptors is therefore one of the most well studied receptors signaling systems (Herbst, 2004; Citri & Yarden, 2006). Each receptor is a transmembrane protein, consisting of a ligand-binding extracellular domain, a membrane-tethering transmembrane domain and an intracellular tyrosine kinase domain. There are four types of ErbB receptors: ErbB1 (EGFR), ErbB2 (HER2), ErbB3, and ErbB4. The ErbB types are distinguished by their preferred ligands, as well as by their tyrosine kinase activities. This thesis focuses on the signaling through ErbB1, or EGFR, as it is an important regulator of cell fate decisions and is highly expressed in multiple cancers.
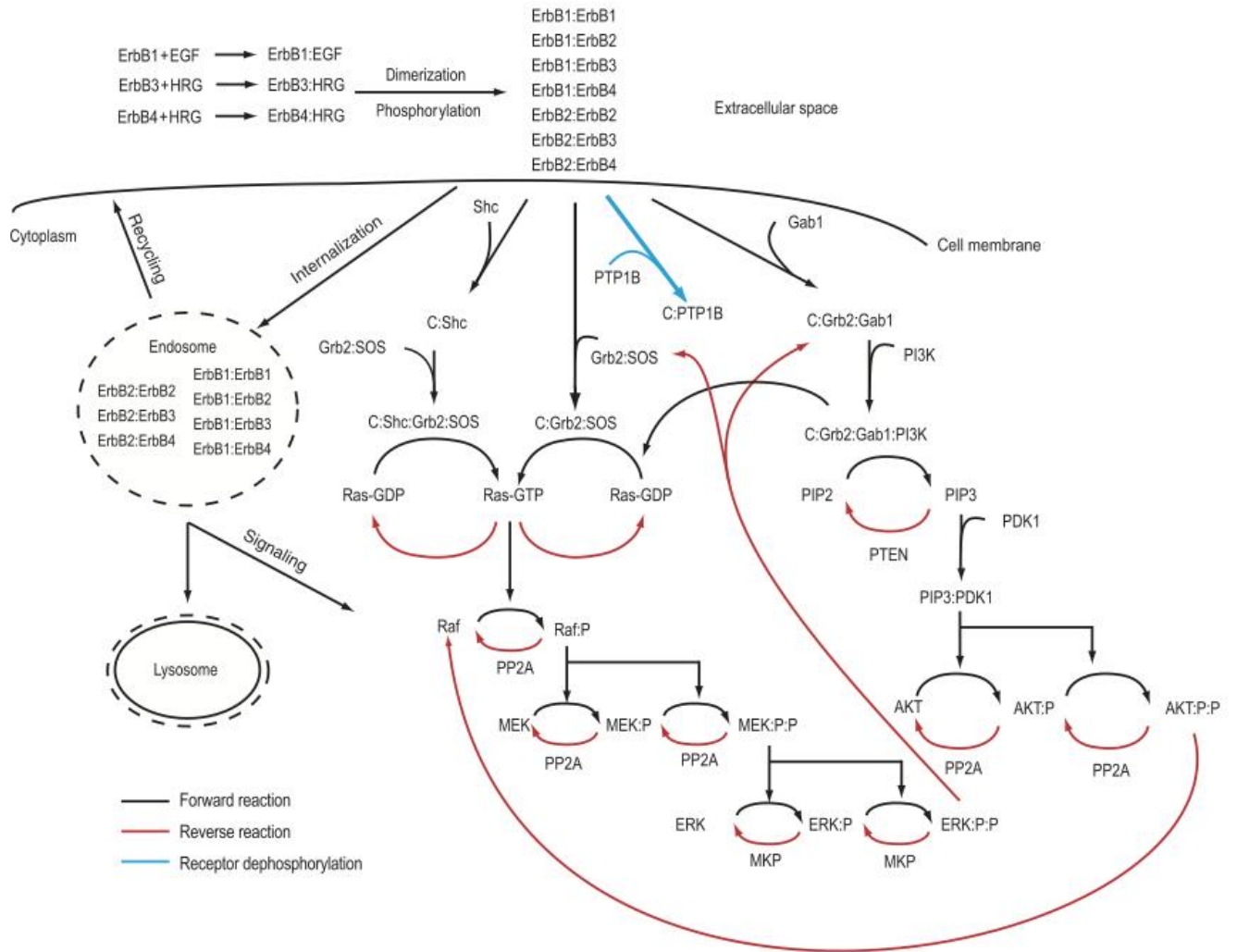
Figure 2.1: Simplified diagram of the ErbB signaling network model. Adapted from (Chen et al., 2009)

Section 2.2: The ErbB pathways in cancer

Aberrant signaling through ErbB receptors can cause uncontrolled cell growth and proliferation, and thus lead to cancer (Arteaga & Engelman, 2014). The over-activation of the ErbB receptors can arise at various levels in the cellular pathways and through multiple mechanisms. At the receptors level, abnormal signaling may be caused by receptors mutations, their overexpression, or their activation by an autocrine or a paracrine ligand secretion (Kleiman, 2010). In addition, the over-activating mutations in the downstream signaling targets of ErbB, as well as the loss-of-function mutations in the negative regulators of the signaling can also lead to the uncontrolled ErbB signaling and cause cancer.

Given the myriad ways that the abnormal ErbB signaling can lead to oncogenesis, it is not surprising that the ErbB and their pathways are an important target for the anti-cancer therapies. First, the receptors themselves can be targeted by the antibodies that prevent receptors dimerization and activation, or induce their downregulation through yet to be understood mechanisms. An example of such drug is Herceptin, a monoclonal antibody inhibiting ErbB2 receptors. Herceptin has revolutionized breast cancer treatment in the past decades (Cobleigh et al., 1999). Next, to attenuate ErbB phosphorylation various classes of small molecule inhibitors of ErbB kinase domains have been developed (Arteaga & Engelman, 2014). The downstream targets of the ErbB receptors signaling have been targeted as well. For example, small molecule kinase inhibitors of B-Raf are successful in treating melanoma carrying specific B-Raf mutations (Karasarides et al., 2004). Ras has been actively pursued as an anti-cancer drug target, for its role in the phosphorylation signaling cascades and its ability to induce cell death through mechanisms bypassing caspases activation (Skouta, Hayano, Shimada, & Stockwell, 2012).

Interestingly, unexpected outcomes of low dose inhibitor treatments have been reported from the studies of EGFR signaling dynamics: low dose EGFR inhibitor pre-treatment might prevent receptors internalization and lead to their sustained signaling (Citri & Yarden, 2006; Kleiman et al., 2011). Hence, a better understanding of how the inhibitors of ErbB pathway affect the ErbB signaling dynamics and cell physiology will be helpful in optimizing drug dosage and in developing improved treatment strategies.

Section 2.3: EGF in the physiological context

Understanding the effects of EGF, and growth factors in general, on the physiology of living organisms is challenging for several reasons (M. S. Murphy, 1998). First, the biological effects of the growth factors can be context-dependent and specific to each species (Hirai M, Gamou S, Minoshima S, 1988). Second, the regulation of the growth factor production and the expression of their designated receptors *in vivo* are influenced by intertwined dynamic cell processes. Interactions between the growth factors and other molecules in the local tissue environment further obscure the study of growth factors effects *in vivo*. Moreover, the growth factors can act at various distances from their secretion sites throughout the organism: they can signal via endocrine, paracrine, autocrine, or

juxtacrine mechanisms. Lastly, the relationship between the growth factors and their effects on living organisms is highly interlaced: each individual growth factor can have many physiological effects, while multiple distinct growth factors can produce the same effect (M. S. Murphy, 1998).

EGF stimulates cells to grow, divide, or differentiate. EGF is a 6kDA molecule that binds to the EGFR and other receptors on cell surface and initiates cell signaling. EGF is produced and present in multiple human tissues (Kajikawa et al., 1989; Birk et al., 1999). Studies have suggested that the main purpose of EGF in adult mammals is to maintain their epithelial surfaces (Fisher & Lakshmanan, 1990), although EGF also plays a role in mammary tissue development and homeostasis (Yarden & Pines, 2012). On the other hand, a growth inhibition by EGF has also been reported in some cell lines, such as A431 (Hirai M, Gamou S, Minoshima S, 1988). Recent studies explored the effect of growth factors stimulation at a single cell level. It was suggested that the growth factor concentration determines the probability a cell will make a certain cell fate decision (Birtwistle et al., 2012).

Epithelial organs secrete EGF into the lumen (Singh & Harris, 2005). In mammalian tissues, EGF doses can vary across four orders of magnitude (Fisher & Lakshmanan, 1990). In humans, EGF is estimated to range from 0.015-0.036 ng/ml in plasma, and up to 280ng/ml in prostatic fluids. In general, 10ng/ml EGF dose is considered to be a "high-physiological" dose (Chen et al., 2009; Osborne, Hamilton, Titus, & Livingston, 1980). Interestingly, such definition stems from the fact that in many cells EGF near 10ng/ml elicits a saturating response on pERK. The phosphorylation and activation of ERK are considered to be a better proxy for the cell fate decisions, rather than the activation of pEGFR itself. In addition, even low EGF doses are able to produce significant activation of the downstream cascades (Uyemura, Takagi, Yanagida, & Sako, 2005) though it is difficult to detect the pEGFR activation distinguishable from the basal levels in such stimulations. Hence, physiologically relevant doses of EGF are often defined mostly based on the mitogenic effect on activating pERK. However, most recent studies suggest that unlike the nanomolar EGF concentrations used in studies of EGFR, the picomolar ranges of EGF doses might be functioning *in vivo* instead (Rouger, Goaillard, & Marguet, 2014).

Despite the advances in experimental techniques in biology, quantifying levels of extracellular diffusible molecules *in vivo* remains challenging (Goentoro et al., 2006). Moreover, the physiological temporal profiles of the EGF levels can be very diverse. For example, in the organism development, EGF doses can change exponentially with time or in fold changes over the days (Fisher & Lakshmanan, 1990). In some tissues, cells can be exposed to chronic EGF stimulations with more or less constant levels of EGF over time (Alberts B, Johnson A, Lewis J, et al, 2002). In addition, EGF can be secreted in pulses with various time intervals on the order of hours in between the pulses (McEwen, 2001). However, it has been suggested that instead of abrupt pulses in EGF, cells are more likely to experience more gradual changes in EGF levels *in vivo* due to a slow process of EGF diffusion in the blood and tissues (Lazova & Ahmed, 2011).

In vivo EGF can act through autocrine, paracrine, and juxtacrine signaling (Singh & Harris, 2005). In the autocrine signaling cells release signaling molecules that bind to the cell's own receptors or to those of the neighboring cells. In this way, a cell can signal to itself or to its neighboring cells (Alberts B, Johnson A, Lewis J, et al, 2002). The autocrine signaling is most effective when many nearby cells of the same type start producing and responding to the autocrine signals together, thus generating a group response. This signaling mode is important in many contexts, including in development and in maintenance of tissue homeostasis. In general, the autocrine activation of neighboring cells can take several hours, as producing and secreting an signaling molecule can be a slow process and might require certain time to accumulate sufficient levels of the signaling molecule to trigger cell response (Toettcher, Weiner, & Lim, 2013). EGF concentrations in the vicinity of autocrine and paracrine signaling cells can reach nanomolar ranges (Rouger et al., 2014).

Section 2.4: Kinetics of the EGF binding and EGFR receptor dimerization

EGFR can recognize and bind different ligands, including EGF, TNF-alpha, and other molecules (Yarden, 2001). Different affinities of the ligands towards EGFR monomers or dimers can produce different biological effects of the ligand stimulation (Macdonald-Obermann & Pike, 2014). In addition, differences in ligand affinities to receptors in various cell compartments result in different

signaling outcomes. For example, it has been shown that TNF-a, an EGFR ligand, dissociates from the internalized receptors in endosomes, and thus directs the receptors towards their recycling to cell surface; whereas, EGF does not dissociate from the receptors in endosomes and targets them for degradation (Reddy, Wells, & Lauffenburger, 1998).

The EGF ligand binding to receptor monomers induces a change in their conformation, leading to their rapid dimerization and activation (Lemmon & Schlessinger, 2010). It has been suggested that EGF binds preferentially to unoccupied EGFR receptor monomers leading to their rapid dimerization, or to unoccupied receptor dimers, and that the affinity towards the one-EGF occupied dimer is significantly lower (Macdonald & Pike, 2008). The equilibrium constant of EGF unbinding to EGFR has been estimated to be on the order of ~3nM (Chen et al., 2009).

Section 2.5: Mechanisms of ErbB receptor signaling down-regulation

In this section we give an overview of some of the receptor signaling attenuation strategies relevant to the EGFR based on the literature evidence and on the setup of our experimental system. Downregulation of signaling through ErbB can be realized via the following mechanism:

1) Receptor-mediated ligand depletion from the cell surface;
2) Removal of receptors from the cell surface and their consequent degradation in lysosomes;
3) Action of phosphatases;
4) Inactivating phosphorylation;

First, our experimental system setup lets us exclude ligand depletion as a signal decay mechanism. Indeed, in our setup the EGF concentration is maintained almost constant over the time of observation, up to six hours after the initial EGF addition.

Inhibitory post-translational modifications (PTM) can temporarily remove receptors from the signaling pool. Such signal downregulation mechanisms have been reported for various receptors sensory systems: for example, in bacteria chemotaxis the CheY receptors can be reversibly inhibited by methylation (Vladimirov & Sourjik, 2009). Inactivating PTM for ErbB have previously been mentioned in the literature (Countaway, Nairn, & Davis, 1992). However, to the best of our knowledge, these mechanisms are not applicable to the experimental system that we investigate.

Therefore, the ErbB receptors internalization, degradation and action of ErbB phosphatases are the main mechanisms that terminate signaling in our system. The activated EGFR are being constantly dephosphorylated by the RTK phosphatases, as has been shown experimentally for the MCF-10A and other cell lines used in this work (Kleiman et al., 2011). The EGFR can be internalized from the cell surface by a slow basal internalization and degradation process; however, a much faster ligand-induced EGFR internalization is a characteristic feature of the ErbB receptors signaling (Becker et al., 2010). In addition, several types of ErbB internalization have been reported, based on their dependence on clathrin (Sigismund et al., 2012).

Why would cells need to simultaneously use two different mechanisms, namely, receptors sequestration and dephoshorylation, to downregulate EGFR signaling? One possible explanation is that receptors internalization and degradation is a slow process on the order of minutes to hours, as it requires assembly of protein scaffolds and vesicles to shuttle the receptors into the cell interior; whereas, action of phosphatases is a rapid process on the order of seconds, and thus can remove receptors phosphorylation much faster.

Section 2.5.1: Internalization, degradation, and recycling of EGFR

ErbB receptors internalization is a complex dynamical process. Receptor internalization can depend on various factors, such as a nature of the ligand (Roepstorff et al., 2009), ligand concentration (Sigismund et al., 2013), types of the receptors forming the dimers (Hendriks, 2003), intracellular machinery (Goh, Huang, Kim, Gygi, & Sorkin, 2010), and even environmental temperature, which affects plasticity of the cell membrane enabling receptors internalization (Macdonald & Pike, 2008).

There are two major endocytic mechanisms: clathrin-mediated (CME) and non-clathrin endocytosis (NCE) (Sigismund et al., 2005). In CME, receptors are being internalized via assemblies of clathrin coated pit proteins that form buds off the plasma membrane. NCE operates by forming membrane vesicles independent of clathrin. It has been suggested that CME is present at all ligand concentrations, whereas, NCE is triggered at higher EGF doses above a certain threshold (Sigismund et al., 2013). The CME machinery can get saturated at higher ligand doses, leading to the activation

of NCE. The two types of internalization are believed to direct receptors to different fates: CME preferentially guides the receptors to recycling, whereas NCE to degradation (Sigismund et al., 2008)

The analysis of ErbB downregulation is challenging due to the fact that there are many dynamic processes entangled together, namely, receptors internalization, recycling, production and degradation. ErbB internalization and degradation help to maintain cellular homeostasis by controlling the number of receptors available on the cell surface. In the absence of ligand stimulation, the cell surface receptors levels can be kept relatively constant by a low level of basal internalization, degradation, recycling, and production of new EGFR.

Cell context, namely gene expression of other receptors, can influence EGFR internalization. Specifically, the ErbB2 receptors can form dimers with EGFR, however, such heterodimers were shown to internalize slower than EGFR homodimers due to an internalization-defective ErbB2 intracellular domain (Hendriks, 2003).

Internalized receptors can still continue signaling to their downstream targets. Thus, endocytosis has been suggested to have a bidirectional relationship with signaling (Kholodenko, Hancock, & Kolch, 2010) – it helps to terminate signaling and, at the same time, it brings signaling closer to the remote downstream targets.

In addition, endocytosis plays a role in temporary desensitization of cells to EGF. In the course of ligand stimulations, when receptors have been internalized from the cell surface, and prior to their recycling or production of newly synthesized receptors to the cell surface, the ErbB-mediated responses are substantially decreased (Kleiman et al., 2011).

Section 2.5.2: EGFR dephosphorylation by phosphatases

RTK phosphatases have been shown to play an important role in ErbB signaling (Tiganis, 2002). Dimerized receptors are rapidly shuffling between phosphorylated and un-phosphorylated form due to the action of phosphatases, independently of the cellular localization of EGFR receptors (Kleiman et al., 2011; Shankaran et al., 2012). The half-life of the phospho-EGFR signal at inhibition with a small molecule kinase inhibitor is very rapid and was shown to be on the order of 10 - 40 seconds (Kleiman et al., 2011).

ErbB phosphatases are difficult to study experimentally due to their ability to act on multiple targets, as well as due to the fact that multiple different phosphatases, often yet unidentified, are present at different expression levels throughout the cell (Soulsby & Bennett, 2009). Recent experimental approaches focused on applying phosphatase inhibitors to assess the role of phosphatase activity in cell signaling. However, pan-phosphatase inhibitors, such as pervanadate, can inhibit other than the intended phosphatases and thus can generate various unanticipated effects on cell signaling (Huyer et al., 1997).

Phosphatases can play an important role in maintaining cellular homeostasis by suppressing basal phosphorylation on EGFR (Reynolds et al., 2003). In the absence of external stimuli, ErbB can still auto-phosphorylate and transmit signaling downstream. ErbB overexpression can lead to ligand-independent constitutive activation of EGFR observed in multiple cancers (Zhou & Agazie, 2012). Phosphatases are believed to prevent activation of signaling cascades downstream of EGFR by suppressing basal activation of receptors, which can arise due to receptor overexpression or in the context of noisy fluctuations in extracellular stimuli (Östman & Böhmer, 2001).

Interestingly, studying RTK phosphatases can be challenging due to complex regulation of their activity in cell signaling: activated EGFR can phosphorylate and activate their phosphatases or induce gene expression of new phosphatases (Tiganis, 2002)

Section 2.6: Receptor recovery dynamics

Recovery of cell surface receptors is important for cell physiology as it determines cell sensitivity to repeated ligand stimulations. After responding to initial stimuli, cells might enter into a refractory state, duration of which is affected by the speed of receptor recovery. Extended refractory state might help cells complete processing initiated response and prevent new redundant response if stimulus is repeated.

Dynamics of receptor recovery can depend on multiple factors. When receptors are removed via degradation or sequestration, the recovery of functional cell surface receptors can require more time to return existing receptors to the cell surface or to synthesize new receptors. The rate of production of new receptors might depend on the metabolic state of cells, which in turn can be influenced by ligand stimulation. The presence of ligand in the extracellular space might prevent

complete recovery of cell surface receptors, if available ligand leads to receptor activation and downregulation (Reddy et al., 1998). Such characteristics of receptor recovery can be relevant to many different types of receptor sensory systems. For example, it has been shown that to recover cell surface Erythropoietin receptors, Epo ligand has to be depleted from the cell surface (Becker et al., 2010). However, recovering cell surface receptors might not restore cell sensitivity to Epo if at the cell surface receptors undergo a rapid reversible inhibitory post-translational modification. This case is exemplified by inactivating methylation of CheY receptors in bacterial chemotaxis.

To the best of our knowledge, there are no experimental studies of cell sensitivity to restimulation with EGF through EGFR that would encompass measurements of both the responses of downstream targets of EGFR, such as pERK, recovery of EGFR to the cell surface and EGF depletion from the media. For example, some studies assessed recovery of cells sensitivity to repeated ligand stimulation over extended periods of time by measuring phosphorylation on receptors or on pERK (Kuppuswamy & Pike, 1991; Pennock & Wang, 2003), however exact receptors mass as well as EGF depletion were not measured in these studies.

Various estimates of the time of cell surface EGFR recovery are presented in the literature, yet all of them agree that EGFR recovery is a slow process, which might take on the order of seven or more hours (Hirai M, Gamou S, Minoshima S, 1988). The reported time of EGFR recovery would also depend on the cell context and the dose of initial EGF treatment. For example, in response to six hours of incubation with 1ng/ml of EGF, a nearly physiological EGF dose, in HER-14 cells the levels of EGFR on the cell surface decreased by 20%, and were restored to their initial number only 24 hours afterwards (van de Poll, van Rotterdam, Gadellaa, Jacobs-Oomen, & van Zoelen, 2005). Another study reports that four to six hours are required to recover expression of functional cell surface receptors through de novo EGFR synthesis (Vijver, Kumar, & Mendelsohn, 1991).

In this work, EGFR receptor recovery dynamics plays a role in retaining the memory of the dose of a persistent stimulation with EGF. At repeated stimulation, cell responses depend on the number of functional EGFR present on the cell surface ,as well as, on the dose of EGF in the new stimulus. In addition, changes in various intracellular molecular components downstream of EGFR can affect cell responses to repeated stimuli. However, for the case of pERK signaling, studies

indicate that the downstream kinase cascade signaling module between Ras and ERK itself does not provide any memory of past stimulations (Toettcher et al., 2013). Thus, it is likely that recovery of cell sensitivity to EGF by restoring the number of EGFR on the cell surface is a key process behind cell memory of EGF at the level of downstream intracellular kinases.

Section 2.7: Effects of ErbB expression levels on signaling outcomes

Although the relationship between EGFR receptor activation and mitogenic cell responses is not very clear (van de Poll et al., 2005), it has been suggested that levels of ErbB expression can play an important role in causing aberrant signaling (Hendriks, Opresko, Wiley, & Lauffenburger, 2003). The exact role of EGFR expression in signaling activation and cell fate remains puzzling: cells expressing low EGFR levels can generate much stronger responses to EGF than cells with intermediary levels of receptors; whereas for some cells overexpressing EGFR EGF can be growth inhibitory and induce apoptosis (van de Poll et al., 2005). Among other factors, ErbB gene amplifications as well as mutations in ErbB receptors might control the number of ErbB receptors on the cell surface through overexpression or through, for example, impaired downregulation (Sigismund et al., 2013). These studies suggest that levels of expression of cell surface receptors are important to cell signaling. Moreover, the biochemical properties of various receptor species (e.g., homodimers or heterodimers), such as ligand binding affinities, rates of dimerization, and internalization, can shape signaling dynamics.

Recent studies have explored the relationship between expression of cell surface receptors and signaling outcome. For example, Macdonald & Pike, 2008 suggested that variations in EGFR expression within a physiological range can affect signaling response by determining the distribution of receptors between monomers and dimers before EGF treatment. Given that the two species have differential affinities to EGF ligand, shifting the number of receptors can alter the proportion of monomer to dimer species and thus affect signaling outcomes.

HER2 (ErbB2) overexpression has been detected in many cancers (Baselga & Swain, 2009). Hendricks et al (Hendriks et al., 2003) studied how HER2 amplification affects cell signaling via changing receptors trafficking. Their work suggests that internalization-defective ErbB2 receptors can form dimers with other receptors, such as EGFR, thus preventing EGFR downregulation from the cell

surface and potentially prolonging cell signaling. HER2 overexpression has been reported to convert transient pERK responses to EGF into sustained pERK activation (Birtwistle et al., 2007).

Interestingly, in development, cell surface EGFR can regulate spatial range of diffusible morphogens, by depleting morphogens from the extracellular space (Goentoro et al., 2006). As a result, cells with various levels of EGFR expression when exposed to gradient of the ligand, can produce different signaling responses.

However, other studies suggested that reduced levels of EGFR might be important to avoid EGF-induced growth inhibition in certain cell lines (Hirai M, Gamou S, Minoshima S, 1988). In particular, the authors reported that A431 cells, which express on the order of 10^6 EGFR per cells, were inhibited by periodical high dose EGF (100ng/ml) treatments. The authors proposed that not only EGFR gene amplification loss, but also efficient downregulation of EGFR receptors was important in evasion of EGF-induced cell growth inhibition.

Constitutively high levels of EGFR expression on the cell surface might allow cells to detect high doses of EGF (Becker et al., 2010). However, overexpression of EGFR might not necessarily lead to high cell sensitivity to EGF, if EGF is being efficiently depleted from the cell surroundings, as efficient EGF depletion might reduce EGF mitogenic potential (van de Poll et al., 2005).

Although the exact predictive and diagnostic value of EGFR expression is not well understood yet, a positive correlation between pEGFR and EGFR has been reported (Olsen et al., 2012). Surprisingly, one study found that in some cancer tissues, levels of EGF, pEGFR and EGFR are downregulated compared to the normal control mammary tissue (Olsen et al., 2012). It might be the case that difference in dynamics in pEGFR responses caused by upregulated levels of EGFR, and not the upregulated levels of pEGFR per se, could be of more predictive clinical power. Overall, understanding the abnormalities in dynamics of cell signaling through EGFR might help to consolidate studies with conflicting findings regarding the role of EGFR overexpression in cancer.

Section 2.8: Signaling cascades downstream of the EGFR receptors

Activation of the EGFR receptors ultimately leads to activation of their downstream targets, such as, phospho-ERK and phospho-Akt, which play an important role in determining cell fate. AKT, also referred to as protein kinase B (PKB), is a serine/threonine kinase which regulates cell

metabolism, suppresses apoptosis, and promotes cell size growth (Hemmings & Restuccia, 2012). Mutations in pathways leading to activation of AKT have been found in many cancers (Sulis & Parsons, 2003). ERK stands for extracellular signal-regulated kinase, also referred to as mitogen activated protein kinase (MAPK). The MAPK pathway is among the most well studied mammalian kinase signaling cascades.

The balance between activation and deactivation of pERK and pAKT determines signaling duration and magnitude – features that influence cell fate decisions. Attenuation of signaling on the downstream targets of EGFR can be realized through signal decay on pEGFR itself, action of phosphatases on various components of the signaling cascade, translocation of the targets into a different cell compartment, and activity of negative feedback loops. Degradation of the downstream targets might also contribute to signaling attenuation. However, no significant degradation of the kinases downstream of EGFR has been reported in the course of several hours of EGF stimulation. In addition, upon activation, pERK is known to translocate into the nucleus where it can be a subject to regulation by various other phosphatases (Cohen-Saidon, Cohen, Sigal, Liron, & Alon, 2009). Negative feedback loops from ERK and AKT are cell-context and stimulus dependent features of signaling networks, and thus will not be considered in this work. Experimental studies with the use of kinase inhibitors of upstream activators of pERK, such as phospho-MEK, show that pERK is a subject to rapid dephosphorylation by phosphatases (Hornberg et al., 2005; Kleiman et al., 2011). Hence, in this work we consider signal decay at the receptors level and regulation by phosphatases as primary mechanisms of signal removal on pERK and pAKT.

Section 2.8.1: Phosphatases in the signaling cascade downstream of ErbB

Influencing cell decisions through regulation of phosphorylation responses to growth factors, phosphatases can play an important role in suppressing aberrant signaling and in malignant transformations of cells (Arroyo & Hahn, 2005).

Since complete pERK activation relies on phosphorylation of both serine/tyrosine and threonine residues, dephosphorylation of either of these residues would lead to ERK inactivation (Keyse, 2008). Dual Specificity Phosphatases (DUSP) is an important class of enzymes which

dephosphporylate ERK (Keyse, 2008). pERK can phosphorylate, and thus activate some of its phosphatases; in addition, pERK can induce transcriptional activation of its DUSP (Bermudez et al., 2011). Same phosphatases can regulate multiple kinases in the MAPK cascade across different cellular compartments (Junttila, Li, & Westermarck, 2008).

Phospho-AKT can be regulated by protein phosphatase 2A (PP2A) in a stimulus and cell context dependent manner (Andrabi, Gjoerup, Kean, Roberts, & Schaffhausen, 2007). PP2A can also act on other downstream signaling targets of phospho-EGFR (Sontag, 2001). In signaling cascade upstream of pAKT, a different class of phosphatases can regulate pAKT activation: PTEN – lipid phosphatase of PIP3 is an important tumor suppressor gene upstream of AKT (Sulis & Parsons, 2003).

Section 2.8.2: Signal amplification and ultrasensitivity in the MAPK cascade

A remarkable feature of ERK signaling is its ability to respond to growth factors in a steep as well as in a gradual dose-response manner, depending on stimulus and cell-context. Ultrasensitivity allows to generate detectable mitogenic signals on pERK even at a low number of activated pEGFR (van de Poll et al., 2005). In this case, signal on pEGFR is amplified through the MAPK cascade and thus produces detectable responses on pERK. Steep ERK dose responses have been reported from studies in frog oocytes (Goldbeterl & Koshland, 1984). Ultrasenstive, step-like dose response has been suggested to facilitate all-or-none cell fate decisions, such as cell division. However, in mammalian cells pERK has been reported to have gradual dose responses (Chen et al., 2009).

Multiple molecular mechanisms enabling ultrasensitivity have been suggested, such as saturation of enzymes, cooperativity in enzyme activation, two-step phosphorylation and others (Goldbeterl & Koshland, 1984). In this work, we implement ERK activation as a two-step phosphorylation process. In addition, our model allows ultrasensitivity to arise due to saturation of kinases and phosphatases. As a result, flexibility in parameter assignments allows reproducing both steep and gradual dose responses on pERK.

Section 2.8.3: Cross-talk between ERK and AKT and negative feedback loops

The existence of cross-talk between pERK and pAKT signaling cascades has been previously reported, yet much remains unclear about the details of these interactions (C.-C. Wang, Cirit, & Haugh, 2009). Elucidating cross-talks is challenging as it can be cell-context and stimulus specific. Moreover, observing cross-talk effects can depend on doses of ligand stimulation, doses of inhibitors applied, and timing of the responses (Moelling, Schad, Bosse, Zimmermann, & Schweneker, 2002). Signaling cascades leading to ERK and AKT activation were reported to have cross-talks at various components, and in different cell lines. For example, studies found that AKT can inhibit ERK through Raf (Avraham & Yarden, 2011) in a cell-context dependent manner (Rommel et al., 1999). In addition, components upstream of pERK and pAKT can overlap in a cross-talk. For example, PI3K-dependent MEK activation has been reported (Cirit, Wang, & Haugh, 2010). Moreover, studies suggest that pERK activation is sensitive to PI3K regulation mostly at early stages of signaling, prior to pERK reaching its peak signal (Birtwistle et al., 2007).

pERK can realize a negative feedback loop through activation of its own phosphatases by either phosphorylation or their transcriptional upregulation, as is the case of DUSP activation. In addition, ERK has been reported to negatively regulate various other signaling components upstream in the signaling cascade, such as, for example, Sos (D. Chen, Waters, Holt, & Pessin, 1996), and Raf (Sturm et al., 2010). Furthermore, ERK can regulate various positive and negative feedback loops in the MAPK cascade through inactivation of RKIP, a Raf kinase suppressor (Shin et al., 2009).

We experimentally tested the presence of negative feedback loops between EGF-activated pERK and pAKT in MCF-10A and SKBR3 cells at several different EGF doses and two doses of inhibitors of activation of either pERK or pAKT. The experiments did not reveal the presence of negative feedback loops between pERK and pAKT in SKBR3 cell line. However, they suggested a possibility of a negative feedback loop from pERK and pAKT in MCF-10A cells. With MEK and PI3K inhibitors experiments we used in our system it is not feasible to assess the effect of a negative feedback loop from pERK to itself. One reason for this is that MEK inhibitors prevent pERK phosphorylation and thus would eliminate a feedback loop from pERK if it were present. A more

appropriate approach to detect a negative feedback loop from pERK to itself would be to use inhibitors of kinase activity of pERK. However, it is not clear if such inhibitors would not perturb signaling cascade in various unexpected ways. Since establishing precise regulation of downstream targets by negative feedback loops in the EGFR receptors signaling cascades is not the primary focus of this thesis, we omit implementing any negative feedback loops in our model. Such simplification is acceptable, as our model can still capture reasonably well our experimental data.

Section 2.8.4: Effects of pERK and pAKT signaling responses on cell fate decisions

Activation of ERK and AKT is crucial for cell division, motility, and cell cycle progression (Shaul & Seger, 2007; Kim & Choi, 2010; Hemmings & Restuccia, 2012). ERK can be activated by different stimuli and can elicit a wide range of cell responses depending on the cellular context. Dynamics of signaling responses of pERK and pAKT is crucial as it can encode distinct cell decisions. Yet, it is still not well understood how information about various extracellular stimuli translates into corresponding signaling dynamics on pERK, pAKT, and other targets, and how it encodes specific cell functions. In general, it has been suggested that cells might use two approaches, combinatorial and dynamical, to transduce information through cell signaling. In the combinatorial approach, distinct stimuli that activate the same signaling target can also activate combination of other targets specific to the stimuli. Hence, the combination of these stimulus-specific targets then determines distinct cell responses (Toettcher et al., 2013).

Another way cells can encode signaling information is through the distinct dynamics of signaling targets. Distinct stimuli can elicit activation of same internal targets, yet the target's responses would have different temporal profiles, which would be differently interpreted by the downstream cell machinery into distinct cell functions.

It is likely that cells use both types of information encoding. However, in practice, it is problematic to distinguish between the two mechanisms employed by cells, because network perturbations of specific targets in signaling cascades can often alter dynamics of the signaling.

Even though pERK is an important and well-studied cell signaling target, much remains unknown about how exactly ERK activation encodes information about extracellular stimuli and

translates it into a specific cell response. Many features of ERK signaling responses, such as duration, magnitude and subcellular compartmentalization are believed to be essential for cell decisions (Ebisuya, Kondoh, & Nishida, 2005). Yet, exactly how differences in ERK activities govern cell responses remains a mystery. It was suggested that the main mechanism by which ERK regulates cell fate is through changes in gene expressions. pERK translocates into the nucleus where it activates transcription factors and initiates gene expression (Cohen-Saidon et al., 2009). In addition, ERK can regulate cell function through activation of cytoplasmic targets as well, further propagating the signaling (Ebisuya et al., 2005).

How do various features of ERK activities affect gene expression? In one way, cellular localization of active ERK preferentially to the cytoplasm, and not to the nucleus, can limit ERK-dependent gene transcription. Duration and magnitude of ERK responses are essential in differentiation and cell-cycle regulation (Sasagawa, Ozaki, Fujita, & Kuroda, 2005b). Overall, it has been suggested that distinct ERK activity profiles affect stabilization of immediate-early gene products and other proteins, and this way can lead to different changes in gene expression, ultimately, resulting in different cell fates.

What regulators shape distinct profiles of ERK responses? First of all, the identity and amount of the stimulants affect signaling cascades leading to ERK activation (Chen et al., 2009). Next, phosphatases can control amplitude and duration of ERK responses (Heinrich, Neel, & Rapoport, 2002; Hornberg et al., 2005). ERK phosphatases can be already expressed in cells at the time of signaling; alternatively they can be activated by ERK itself through phosphorylation; finally, ERK activation can induce gene expression of its own phosphatases, as is the case with the Dual Specificity Phosphatases (Keyse, 2008). Amplitude and duration of ERK responses can also be affected by downregulation of the upstream activating machinery, such as internalization and degradation of receptors initiating signaling cascades. In addition, various negative feedback regulators, such as inhibitory post-translational modifications of phospho ERK or the upstream components, such as EGFR or Sos (D. Chen et al., 1996), can lead to signaling attenuation and thus shape dynamics of ERK responses.

What determines the expression and role of the regulators of ERK responses? The abundances, functions and mechanisms of controls of ERK regulators are stimulus, cell-type and cellular context dependent; see Figure (Ebisuya et al., 2005). Such diversity and complexity in regulation can explain, for example, why ERK activation by different ligands in the same cell can lead to different cell response (Sasagawa et al., 2005b), as well as how stimulation with the same ligand can produce drastically different responses in cells of different types (Meloche & Pouysségur, 2007; Ebisuya et al., 2005).

In summary, distinct qualitative and quantitative features in ERK activity, through gene expression, might be translated into specific cell responses. However, quantitative properties of the relationship of ERK activation and cell fate are still poorly understood.

Overall, there is still no consensus as to what profiles of EGF stimulations are required to elicit mitogenic response. First, this relationship is complicated and might be cell-context dependent. Second, different studies vary on the experimental setup they use and cells they work with. It has been suggested that initiation of DNA synthesis in quiescent cells might require at least 8 hours of continuous presence of growth factors (Riddle, Pardee, & Rossow, 1979). Other studies suggested that two short pulses of EGF stimulation are sufficient to transition quiescent cells into G1 phase, followed by entry into S-phase (Pennock & Wang, 2003). Hence, it is possible that to detect subsequent ligand stimulations and activate pERK, cells might need to have functional receptors on their cell surface 8 hours after the initial growth factor stimulation to eventually transition into S-phase (van de Poll et al., 2005).

Recently, effects of ERK activation dynamics onto cell responses were studied through quantitatively relating ERK signaling to the activation of its downstream targets, such as c-fos (Albeck et al., 2013). Transcription factor c-fos controls gene expression relevant to cell differentiation and survival, and has been implicated in many cancers (Milde-Langosch, 2005). c-fos has been shown to mediate entry into the S phase by responding to sustained, and not transient, activation of ERK (L. O. Murphy, Smith, Chen, Fingar, & Blenis, 2002). Such ERK-dynamics dependence of c-fos is likely due to the fact that sustained ERK activity is required for stabilization of c-fos. Although activation of c-fos is believed to be correlated with cell fate, the picture of linking pERK activation to cell fate via c-fos is

still incomplete due to time-separation between processes of activation and the observed response. It might take hours to observe cell responses after pERK stimulation. Frequency of pERK stochastic pulses depends on levels of EGF stimulation and it was suggested that integral of ERK activation over a 2-3 day period results in stabilization and accumulation of downstream transcription factors leading to cell cycle progression (Albeck et al., 2013). Dose responses linking ERK phosphorylation to activities of c-fos and several other ERK's downstream reporters allow to establish quantitative relationship between ERK activity and proxy of cell response. pERK activation at about 10% of its maximal value can have effect on cell proliferation. Hence, to inhibit cell proliferation through ERK one might need to inhibit ERK activity below the 10% threshold (Albeck et al., 2013). Although, this is estimate is quantitative, and not qualitative, it still leaves a wide range of uncertainty – what is an effect on cell proliferation if pERK is inhibited by 50%? It might be useful to have more nuanced quantitative understanding of this relationship between ERK and cell fate, especially, in light of recent tactics in drug development, when partial inhibition of tumor growth or induction of differentiation of tumor cells, and not the complete destruction of tumor, can be a mechanism of action of therapies (Stein et al, 2014).

Emerging experimental optogenetic techniques allow us to directly alter ERK activation dynamics in single cells and observe resulting changes in cell decisions (Toettcher et al., 2013). Such methodology permits to study information processing features of isolated modules in signaling pathways -- for example, Ras/ERK cascade -- and how the dynamics of their responses is deciphered by downstream cell machinery. The advantage of this technique is that it allows direct regulation of targets of interest, bypassing a layer of regulation between extracellular signal and the target of interest. In addition, optogenetic activation allows a better control of timing of reactions and activation (e.g. activation and cell response are not obscured by extended timing between the two, usually present in studies reliant on gene-expression in response to ligand treatment). On the other hand, direct optogenetic activation of the target might disregard the broader context of cellular signaling: for example, stimulus acting via receptor-mediated signal transduction, might lead to activation of not only a specific target of interest, such as ERK, but of its regulators, e.g. its phosphatases. Hence,

although the optogenetic method allows better quantitative understanding of isolated components of signaling cascades, it might omit the broader, systemic regulation features of the signaling cascades.

How do cells interpret quantitative changes in ERK dynamics? Here we have already mentioned how differences in duration in ERK signaling can produce different cell responses via stabilization of c-fos. However, not much is understood about effect of magnitude of ERK responses on cell fate. Does relative or absolute change in pERK in response to stimulation is important for cell decision? It has been hypothesized that effect of fold-changes in pERK on cell decisions might depend on the range of pERK activation; specifically, a two-fold change in pERK might result in a five-fold change in cell response if pERK levels are far below saturating; whereas, a two-fold change in pERK might be unnoticeable at cell response level, if pERK is approaching saturation (Albeck et al., 2013). In addition, fold-changes have been suggested to be important to cell fate in nuclear phospho ERK response to EGF treatment (Cohen et al., 2008). However, mechanisms that cells might use to encode and decipher pERK relative changes are not well established; incoherent feed forward loop at transcription regulation has been proposed as a potential mechanism, yet it was not confirmed experimentally (Goentoro, Shoval, Kirschner, & Alon, 2009). In this work we will focus on cell mechanism that allows cells to generate same pERK responses to step-increases in EGF stimuli independently of the background level of EGF across physiological range.

Figure 2.2: Schematic representation of regulatory mechanisms of ERK activation and cellular responses. Adapted from (Ebisuya et al., 2005)

Section 2.9: Perfect adaptation in cell signaling systems

Previously we discussed how adaptation can allow biological systems to decrease their responses to chronic stimulation, while retaining sensitivity to relative changes in stimuli across a range of background stimuli (see Introduction). Adaptation is ubiquitous across different living

organisms and biological systems, due to survival advantages it provides to organisms and cells in changing environments.

A classic example of adaptation at molecular level is perfect adaptation in bacterial chemotaxis. Perfect adaptation allows bacteria to move towards the source in a nutrient gradient, independently of the absolute concentration of nutrient at the source. Bacteria can either run or tumble. Tumbling allows bacteria search directions with higher nutrient content, whereas runs allow bacteria continue in the preferred, nutrient-rich direction. Sensing nutrients regulates bacterial proteins responsible for motility; perfect adaptation on the responses of these proteins allows bacteria to decrease its motility and stay in high-concentration nutrient region, yet capable to respond to further increases in nutrient levels (Alon et al., 1999).

Recently, perfect adaptation in nuclear phospho ERK has been reported across distribution of single cells in response to EGF stimuli (Cohen-Saidon et al., 2009). However, no mechanism behind the perfect adaptation on nuclear pERK is proposed.

In general, mechanistically, adaptation requires a delayed feedback loop, and in addition it may employ shorter-timescale mechanisms such as covalent protein modifications (Koshland, Goldbeter, & Stock, 1982). For example, adaptation in bacteria chemotaxis is realized through reversible methylation of CheY receptors responsible for bacteria motile function. Nutrient-induced phosphorylation of CheY activates the receptors and leads to bacteria runs. However, over time, CheY is being methylated, which reduces the motility. Demethylation of CheY restores the ability of bacteria to respond to further relative increases in stimuli levels. Such mechanism of realizing perfect adaptation is classified as integral feedback control. Other topologies capable of achieving perfect adaptation include incoherent feed-forward loop and have been recently explored in detail (Ma, Trusina, El-Samad, Lim, & Tang, 2009).

To observe perfect adaptation in sensory systems experimentally, signaling responses would be required to return to a pre-stimulation level independently of the dose of the applied chronic stimulation. However, it might be impracticable to experimentally detect precise adaptation in responses due to measurement detection errors. In addition, it is often infeasible and unnecessary for biological systems to implement exact adaptation, as biology often operates on approximate, not

33

exact values. Hence, it is reasonable to consider responses which return to the pre-stimulus level within a certain threshold, for example, within 5-10% (Goentoro et al., 2009).

Moreover, since maintaining the perfect adaptation state might require energy expenditures from cells, there might be cases when perfect adaptation is not necessary, and approximate adaptation instead is a more suitable option (Koshland et al., 1982). For example, in tissues, hormone levels might decrease over time due to diffusion, washing away from cell surface by blood, or due to depletion through internalization of ligand-bound cell surface receptors. Fading of the extracellular stimulus would aid in attenuating the signaling, and thus help cells to return to pre-stimulus level even if cells do not possess machinery for perfect adaptation. For example, Epo receptors function by depleting Epo from their media through internalization of ligand-bound receptors (Becker et al., 2010).

As of today, there are only few studies on adaptation in signaling systems, especially EGFR related signaling. A recent study explored perfect adaptation on Ras in response to activation of G protein coupled receptors. They find that adaptation in Ras pathway is realized through quantitative changes in upstream components and not through negative feedback loops (Takeda et al., 2012).

It is not clear what are all possible cellular mechanisms behind perfect adaptation, what are the implications of adaptation to cell physiology and function. In this work we consider adaptation of phospho signals in EGFR signaling cascade and will explore differences in adaptation properties in cancer versus normal mammary epithelial cell lines. One hypothesis is that it is not the ability to adapt perfectly, per se, but the time scale of adaptation, which might be important to cell fate decision. Mutations in signaling cascade might delay adaptation, prolonging signaling, which might lead to aberrant cell growth and cancer.

What would prefect adaptation in cell signaling through EGFR mean to cell physiology? Is this property lost in cancer cells? Is it still present in cancer cells the dynamics of pEGFR, pERK, pAKT responses is altered? What role does perfect adaptation in EGFR signaling play in deciphering consecutive EGF extracellular signals? What mechanisms do cells use in perfect adaptation to EGF at the levels of pEGFR and at the level of its targets? We will attempt to address these questions in our work on the example of EGF signaling through EGFR signaling cascades.

34

Section 2.10: Weber's Law in living systems

Beyond ordinary human senses, the first evidence of Weber's Law in biological systems emerged in the end of 19th century from studies of bacterial chemotaxis (Mesibov, Ordal, & Adler, 1973). In this section, we will collect other known examples of Weber's Law at cellular and molecular levels as well as provide a rationale for why relative sensing in signals might be advantageous. It is possible that necessity to adapt to changing environments and to perform numerous complex functions, led living organisms to develop effective information processing strategies at cell, tissue and organism levels. Hence, it is not surprising if the organisms at these various levels converged to have a similar property of relative sensing.

Cells in tissue might benefit from mechanisms (such as relative sensing) aiding effective cell group responses. For example, it has been suggested that collective cell responses can be advantageous from the point of view of information processing (Cheong, Rhee, Wang, Nemenman, & Levchenko, 2011): responses in a number of cells could help mitigate molecular noise. However, such information processing gains might be limited by the number of cells exposed to the same stimulus. Hence, the information processing strategy in cells might be further enhanced, for instance, through the integration of memory of the past signals in cell responses. Cell responses relative to the pre-stimulus state might be more robust to noise originated from cell-to-cell variability in gene expression or from uneven distribution of stimulus among the cells, characteristic to cell populations in tissues.

Second, implementing relative sensing mechanisms allows cells to function in many different contexts. Various tissues in the organism can be exposed to the chronic stimulation with the same ligand, but of different doses. Hence, relative sensing might allow cells to use same mechanisms to perform same functions across tissues, independently of the stimuli doses characteristic to different tissues.

Third, relative sensing can help cells maintain sensitivity to repeated signals. For instance, if a cell, while processing information from incoming signal, is receiving a next signal, of the same nature, but at a higher level, relative sensing can allow cells to adjust their sensitivity and generate detectable response to the repeated stimuli only if it is significantly different from the original stimuli.

35

Relative sensing is present across different simpler species and can help the organisms facilitate various physiological behaviors. A classic example of relative sensing in inputs in biology is bacterial chemotaxis in nutrient gradients. It has been reported that E.coli is able to navigate towards regions with the highest concentration of nutrient independently of the absolute level of nutrient at the source (Lazova & Ahmed, 2011). Furthermore, yeast has been suggested to implement relative sensing in glucose and galactose levels (Savir, Escalante-chong, & Springer, 2005). It is conceivable, that more precise stimuli sensing can help yeast optimize its growth strategies depending on the abundance of its preferred growth nutrient. Drosophila olfactory searches were suggested to be scale-invariant in odorant levels (Martelli, Carlson, & Emonet, 2013), and thus resemble bacteria chemotaxis nutrient searchers. The relative sensing in drosophila olfactory searches in mediated through olfactory receptors.

Recently, several examples of relative signaling in mammalian cells were reported. Although these mechanisms might be essential for cell functioning in tissues, their effects on broader cell physiology remain to be elucidated. TNF-induced signaling has been suggested to determine downstream transcription in single cells via fold-changes, not absolute levels, in the activity of the transcription factor NF-kB (Lee, Walker, Savery, Frank, & Gaudet, 2014). Geontoro and colleagues (Goentoro & Kirschner, 2009) observed apparent correlation between Wnt initiated fold changes in b-catenin levels and transcription of response genes, as well as the phenotypes they produce in drosophila development. As has been previously mentioned, nuclear phosphor-ERK levels might respond to EGF in fold-changes (Cohen-Saidon et al., 2009).

Interestingly, the systems in the above examples most likely implement relative sensing through different underlying mechanisms. In the case of bacterial chemotaxis, relative sensing is known to result from integral feedback control via inhibiting reversible methylation of cell surface receptors; whereas for NF-kB, b-catenin and nuclear phosphor ERK examples, the fold-change detection is thought to be realized by transcription factors machinery. This outcome suggests that distinct biological mechanisms, such as methylation, a post-translational modification, as well as transcription regulation, can be responsible for the relative sensing of inputs. This again reinforces the idea of converging evolution to the same relative sensing function across different levels in biological

systems. Furthermore, this observation leads to question what other mechanisms of post-translational modifications can retain memory of stimuli and implement relative sensing in inputs.

In the following subsections we will consider in detail several recent works on relative sensing across different organisms and different signaling systems.

Section 2.10.1: Example of Weber's Law at the organism level: scale-invariant nutrient searches in bacterial chemotaxis

First evidences of Weber's Law in bacteria emerged as early as 1988 (Mesibov et al., 1973). Bacteria were placed in solutions with meat extract. It has been noted that a fivefold increase in concentration of meat extract over the initial background concentration was needed to produce a detectable movement of bacteria. This effect was observed independently of the initial background concentration of the extract across two orders of magnitude of this concentration. Such phenomenon is referred to as *just noticeable difference* and is described by the original Weber's Law (see Section 2.10).

Next, bacteria have been shown to perform scale-invariant searches in nutrient gradients (Mesibov et al., 1973). In particular, bacterial responses to nutrients were quantified by the number of cells that over some time migrated to the sources of the nutrient. It was found that bacteria responses were independent of the basal nutrient levels across several orders of magnitude, for the constant ratio of concentration at the source to the basal level.

Subsequent studies focused on exploring molecular mechanisms behind scale-invariance in bacterial chemotaxis. In general, bacterial nutrient search relies on tumbling, or steps in random directions, and movements in one given direction (Keymer, Endres, Skoge, Meir, & Wingreen, 2006; Vladimirov & Sourjik, 2009). Tumbling is controlled by kinase activity of certain receptors, which upon binding of ligand decreases their methylation and reduce their kinase activity, thus keeping bacteria from tumbling and instead allowing it to move in one direction, along the gradient of the attractant (Tu, Shimizu, & Berg, 2008; Hart et al., 2013). This property allows bacteria to move towards the source of the nutrient, independently of the absolute level of nutrient at the source. Such scale-invariant could give bacteria survival advantage in environments with scarce or varying levels of nutrients.

Series of theoretical studies (Alon et al., 1999; Tu et al., 2008) suggested mechanisms behind relative sensing in bacteria and proposed that topology enables bacteria to produce fold change detection responses in nutrient sensing (Shoval et al., 2010).

Fold-change detection in E.coli chemotaxis recently has been experimentally verified both at the molecular level and at the organism behavior level (Lazova & Ahmed, 2011). The activities of key intracellular proteins regulating E.coli motility were measured in response to treatments with various doses of nutrient. To quantify the resulting displacement of the bacterial population in nutrient gradient with various background levels the authors report chemotactic migration coefficient that approximates average population-averaged spatial coordinate of bacteria. The authors identify regions of doses of stimuli in which both the reporter's responses and bacterial displacement are constant over at least one order of magnitude and with a schematic model explain the key regulatory features of bacteria chemotaxis system that allow scale-invariant searches.

Interestingly, mechanisms for both relative and absolute sensing in stimuli might be present in bacteria. For example, two-component systems with an absolute-response mechanism have been reported in bacteria (Batchelor, Silhavy, & Goulian, 2004; Shinar, Milo, & Alon, 2007). A classic example when an absolute amount of stimulus matters is a case of organismal embryo development (Rogers & Schier, 2011).

Section 2.10.2: Example of Weber's Law at the organism level: relative sensing of odorant levels in the drosophila olfactory system

Odorant discrimination and responses in insects presents an interesting example of encoding various cell functions through different response dynamics of the same signaling target. Recent work explored dynamics of olfactory neuron responses to different odorants (Martelli et al., 2013). Responses generated in such system usually last on the order of seconds. The authors find that shapes of such responses are stimulus-dependent. Further, they find that for certain odorants, although the magnitude of the responses to repeated stimuli changes, response dynamics is independent of background intensity over several orders of magnitude of ligand dose. Since in natural environments odorants usually occur on top of some background, it might be useful for olfactory

systems to detect stimuli changes relative to the background level. Moreover, in the odorant regions where response dynamics is background-independent, several neurons can detect changes in odorant levels and produce response with reliable dynamics, which could then be interpreted as the same cellular function. The authors suggest that such feature can allow insects to perform olfactory-driven navigation independent of signal intensity and background, reminiscent to bacteria chemotaxis movement in nutrient-gradient space.

This example further demonstrates ubiquity of relative sensing in biological molecular systems.

Section 2.10.3: Fold-change responses in nuclear phospho-ERK and an incoherent feed forward loop as a proposed mechanism behind it

Recent studies focused on fold-change detection in cell signaling cascades as a mechanism for robustness in cell-to-cell variation in gene expression, and not from the point of view of cells in populations responding to noisy levels of extracellular stimuli.

The possibility of phospho-ERK fold-change responses in sensing EGF has been recently investigated (Cohen-Saidon et al., 2009). Activated by EGF stimulations, phospho-ERK2 is known to translocate into the nucleus where it initiates transcriptional responses. The authors developed a clonal cell line with ERK2-YFP fusion protein and monitored single cells response to a constant EGF treatment. They found that despite the significant (up to four-fold) variability in pre-stimulus nuclear levels of ERK2 between individual cells, upon EGF treatment, ratio of peak to basal levels of ERK2-YFP was approximately the same across the cells. In addition, in 80% of the cells ERK2-YFP returned within 10% of the original, pre-stimulus level, thus exhibiting exact adaptation to the basal level.

However, the study was conducted using only a single dose of EGF, 10ng/ml. The dose is considered to be high physiological, at which phospho-ERK response has been shown to reach saturation in several other systems (Chen et al., 2009). It is plausible to assume that the basal level of ERK phosphorylation in a single cell in the absence of the EGF stimulus constitutes a constant fraction of the total ERK present in a cell; similarly, at saturation, the highest possible amount of

phosphorylated ERK is a constant fraction of the total expressed ERK in that cell compartment. These threshold values are likely to be preserved across single cells variance. It is not surprising then, that at saturating EGF dose, phospho-ERK responses across single cells reach the maximal thresholds of activation, thus showing the same ratio of the peak of the response to the initial basal phospho-ERK levels. In addition, the authors did not present any specific molecular mechanism which would be able to decipher pERK fold-changes downstream in the nucleus, although, as a possible candidate, they suggested an incoherent feed forward loop (IFFL) topology in transcription machinery.

Another view on fold-change responses in single cells suggests that growth factor concentrations in tissues can control probability of various cell fate decisions based on basal levels of expression of signaling proteins in individual cells (Birtwistle et al., 2012). For example, total ERK expression variability affects phospho-ERK levels in cells responding to EGF. In certain contexts, to maintain tissue homeostasis some subpopulations of cells in the same tissue might need to generate different population-wide responses to the same stimuli.

A theoretical study suggested a molecular mechanism able of fold-change detection at transcription level (Goentoro et al., 2009). Specifically, the authors investigated how an incoherent feed-forward loop (IFFL) network motif can realize fold-change detection. The authors claim the IFFL topology to be an "almost minimal circuit" to accomplish fold-change comparison in transcriptional networks. The authors consider several possible transcriptional regulation motifs in three-node topologies (X- activator, Y- repressor, Z – target or output). They conclude that to exhibit fold-change detection, IFFL motif does not require parameter fine-tuning, but relies only on some constraints in parameter ratios. The authors also demonstrate limiting cases in which other network motifs, such as "sniffer", can also exhibit fold-change detection. In addition, the authors suggest that some, but not all, networks with perfect adaptation might achieve fold-change detection.

However, in the main paper the authors refer to the fold-change detection property as one in which signals with same fold-changes over initial different background levels, can elicit the same response not only in magnitude but also in the temporal adaptation profile. Whereas, in the detailed description of the methods, to quantify fold-change detection in the parameter scan for the I1-FFL

model, the authors require that the responses to two identical fold-changes in signal match their peaks within 10%. The latter being exactly the requirement for Weber's Law.

Besides the possible benefits of providing robustness to cell-to-cell variation, the fold-change detection might be advantageous in discriminating meaningful stimuli changes over noise. More specifically, if noise strength is proportional to the standard deviation of the mean signal level, then, the higher the basal signal, the higher is its noise threshold level, and hence, mechanisms relying on fold-increases rather than absolute changes might be more suitable to function in such noisy conditions.

Section 2.10.4: Fold-change responses in b-catenin signaling in response to Wnt stimulation

 Another evidence of the possibility of encoding and reading cell information in relative, not absolute amounts was reported by Goentoro and colleagues (Goentoro & Kirschner, 2009). They investigated quantitative features of signaling through the Wnt pathway important in development. Wnt-induced signaling leads to accumulation of b-catenin in the cells which then leads to transcriptional activation of its downstream targets. This study suggests that ratio-sensing mechanisms might be present at the level of gene transcription.

The authors find that Wnt-induced fold-changes of b-catenin, rather than the changes in its absolute levels, are robust to variations in cell parameters and guide transcription of its downstream targets. A theoretical analysis of the Wnt-signaling model with b-catenin accumulation suggests that the absolute levels of b-catenin are sensitive to parameter perturbations, whereas fold-changes are robust. The fold-changes in responses are measured by a ratio between steady state in b-catenin response in cells stimulated with a certain Wnt dose compared to steady state levels in b-catenin in absence of Wnt.

The authors validate their predictions first, via series of experiments in mammalian cell lines: using inhibitors and overexpression constructs they show that absolute levels of b-catenin are sensitive to cell parameter perturbations; however fold-changes in b-catenin levels remain robust. Next, they show the biological relevance of robustness in b-catenin fold-changes in cell phenotype.

On the model of Xenopus embryos they show that the fold-changes in b-catenin level correlate with the developmental phenotype. Further, they also demonstrate existence of the transcriptional machinery possibly decoding relative changes in b-catenin levels.

Interestingly, this study investigated fold-change responses by considering the long-term steady state b-catenin levels, not taking into account the dynamics of immediate b-catenin response to Wnt, motivating it by the fact that to observe effects on phenotypes can take days. Finally, the authors emphasize that in order for the biological network to provide robustness in Wnt-induced fold-change in b-catenin it is important for network parameters to be fine-tuned, a point which we explore in Section 4.7 and Section 5.1.

Section 2.10.5: NFkB signaling in response to TNF stimulation

Lee et al (Lee et al., 2014) found that the TNF treatment controlled transcription of early genes in single cells via fold changes of nuclear NF-kB. The authors observe that the nuclear NF-kB levels vary between untreated single cells. They develop reporters and cell constructs enabling them to measure simultaneously in a single cell the levels of NF-kB translocated into the nucleus in response to a given TNF dose, and the transcription changes of a certain target gene initiated by this translocation, thus linking the NF-kB fold changes to the induced transcriptional response. The authors used several different TNF doses (0, 0.1, 1, 10ng/ml) and combined results in one set and found a correlation between different TNF doses and the induced fold-changes in single-cell NF-kB levels and the accompanying transcription responses.

The authors find that a model of transcription-level regulation by NF-kB levels with incoherent feed forward loop structure is able to recapitulate well the experimental data and predict existence of fold-change response detection. This study, however, did not investigate if the fold-changes in TNF levels were detected by the cells. Hence, similar to the previous example of nuclear phospho-ERK system, this paper concerns fold-change responses and detection in face of cell-to-cell variability in protein abundances.

Section 2.11: Previous quantitative studies of the emergent properties of receptors signaling systems through a combined computational and experimental approach

Combined computational and experimental approaches have been previously applied to study various receptor cell signaling systems (Becker et al., 2010; Tu et al., 2008). Such works often focus on gaining mechanistic insight on signaling networks through the consideration of signaling dynamics, or by understanding information processing of a signaling network. However, not many studies explored different emerging physiological properties of receptors sensory systems that are relevant to physiological context of the specific receptor system and that go beyond the cellular context. To the best of our knowledge, our work, for the first time relates EGFR cell signaling dynamics in response to EGF to a feature of cell memory essential for cells in their physiological environment.

As an example, a recent quantitative study of Erythropoietin receptor (EpoR) signaling was able to explain a peculiar property of Epo receptors enabling cells to sense Epo ligand concentrations across several orders of magnitude (Becker et al., 2010). EpoR receptors stimulate growth and survival of red blood cells in response to Epo. The authors combined experimental data and computational modeling to study signaling dynamics of EpoR and through it were able to derive broader properties of EpoR signaling cascade relevant to its ability to process wide range of Epo concentrations. In particular, it was found that by rapid ligand-independent replenishment of EpoR to cell surface, cells efficiently internalize and degrade the Epo ligand from the cell environment, while recycling EpoR to the cell surface, and thus the cells are able to sense broad range of Epo concentrations. The study thus links the biochemical properties of ligand and receptor interactions, to signaling dynamics and finally to the physiological effect on the cells. As implications, such study can also suggest desirable properties for design of potential stimulants of red blood cell proliferation optimized for their bioavailability and outcome on cell function.

Interestingly, the study (Becker et al., 2010) draws comparisons between the EpoR system and other receptor sensory systems, such as, EGFR and interleukin 3 (IL3) signaling. The authors describe various features characteristic to each specific receptor system. For example, similar to the

43

EpoR, IL3 receptors rapidly deplete ligand from cell surface. However, the key operational mechanism of these receptors relies on production of new receptors, not the ligand-induced mobilization of already synthesized receptors as in the EpoR case. As a distinctive property of EGFR signaling rapid EGF-induced internalization of the receptors is featured; whereas for the EpoR the rate of ligand-induced internalization is comparable to that of the basal internalization. It is interesting to study these distinct receptor signaling systems with a combined approach of experiments and mechanistic modeling, and be able to explain the emerging properties of these system relevant to the physiological context in which receptors function.

In addition, investigating the same cellular phenomenon, for example, the ability to sense a broad range of ligands or to remember doses of past stimulations, in different receptor systems can provide insight into different molecular mechanisms that can be used to realize same cellular function. In particular, it could be interesting to explore what other molecular mechanisms can implement cell memory of previous ligand doses, and how various properties of such memory affect cell functioning.

Another recent work presented theoretical analysis of information processing properties of receptor signaling systems emerging from their various biochemical features (Shankaran et al., 2007). The authors consider general properties of receptor internalization dynamics across several receptor sensory systems. In particular, they study how features of receptors ligand interaction and endocytosis influence function of the signaling system. The authors construct a unified mathematical model of receptors cell signaling and ligand transport, and instantiate each particular receptor signaling system with a specified set of kinetic parameters. They classify receptors systems into three types: those controlled mostly by efficiency of ligand capturing, second type is reliant on internalization of ligand from cell surface, and third type – utilizes both strategies. The authors suggest that the EGFR system is dual-controlled. In addition, they postulate that ligand-induced receptor internalization can help improve precision of signaling receptors.

Given multiple commonalities between the receptors sensory systems it will be interesting to investigate what general properties of signaling cascade would allow Weber's Law in in their responses to fold-changes in ligand concentrations and to explore what various mechanisms of Weber's Law can be realized in various receptors sensory systems.

44

Chapter Three: Methods

Chapter Three describes in detail the computational part of the thesis - the dynamical ODE model of EGFR signaling and the probabilistic methodology to sample the model parameters space and generate predictions of EGFR signaling behavior.

Section 3.1: Mechanistic Dynamic ODE Model

Section 3.1.1: Previous models of the EGFR signaling cascades

The EGFR signaling cascades have been modeled extensively for several decades now (Wiley, Shvartsman, & Lauffenburger, 2003; Kholodenko, Demin, Moehren, & Hoek, 1999; Chen et al., 2009; Schoeberl, Eichler-Jonsson, Gilles, & Müller, 2002; Shankaran et al., 2012; Kleiman et al., 2011). Various aspects of signaling have been addressed, for example, kinetics of ligand binding, dimerization of receptors, their activation and regulation by phosphatases, transport, interaction with scaffold proteins and activation of downstream signaling cascades. Often, for simplicity, studies focusing on immediate-early signaling (Kholodenko et al., 1999) omit implementation of receptor synthesis. Other models assumed an abridged view of ligand binding (Lazzara & Lauffenburger, 2009). In our model we focus on the most relevant features of EGFR signaling, while keeping the model as parsimonious as possible. Our model relates activation of EGFR to downstream responses of pERK and pAKT, yet also implements in detail EGF binding kinetics and receptors activation and downregulation. In general, mechanisms of signal attenuation present in the model are dephosphorylation of pEGFR, pERK, pAKT by their corresponding phosphatases and degradation of EGFR receptors. In addition, the model allows to attenuate signaling via EGF depletion from the media through receptor-dependent internalization of the ligand.

Section 3.1.2: Mechanistic Dynamic ODE Model of ErbB Signaling

Since in the MCF-10A and other cancer cells we work with, the abundances of signaling proteins in the cascade are on the order of at least one thousand molecules per cell, we do not consider stochastic modeling, but develop a dynamic ordinary differential equations (ODE) model. Our model relies on mass action kinetics. We use mass action kinetics, and not the Michaelis-Menten formalism, as the former might be more appropriate in the context of cell signaling *in vivo* (W. W.

Chen, Niepel, & Sorger, 2010). Reactions are encoded as simple unimolecular or bimolecular processes with parameters representing protein concentrations and kinetic rate constants of their interactions and of the first-order reactions (see Figure 2.1). Model equations and species descriptions are provided in the Section A.1.

Model scope encompasses both, detailed receptor-layer interactions (ligand binding, receptor dimerization, etc.) as well as simplified reactions leading to activation of downstream targets. At the receptor layer, we model EGF binding and unbinding, receptors dimerization, phosphorylation, receptors internalization, recycling, degradation, ligand-independent synthesis, and dephosphorylation by RTK phosphatases.

EGF is allowed to bind to and unbind from receptor monomers, unoccupied receptor dimers, and one- EGF molecule occupied receptor dimers, with distinct kinetic rate constants of binding and unbinding depending on the receptors species. Such encoding allows literature-suggested negative co-operativity in EGF binding to one-EGF-occupied dimers (Macdonald & Pike, 2008).

EGF-bound receptor monomer can reversibly dimerize with either unbound or bound EGF monomer. The two reactions are implemented with different binding rate constants to allow differential affinity between the dimers of two types. Dimerization between two unbound monomers is not implemented since its equilibrium constant is known to be about 2.5 orders of magnitude lower that the dimerization between receptors with at least one EGF bound (Macdonald & Pike, 2008).

After receptors form dimers they undergo phosphorylation. We assume that ATP is overabundant in the cells, so phosphorylation of dimers is implemented as a first order reaction. Both, one EGF-bound dimer and two EGF-bound dimers are allowed to undergo phosphorylation.

To keep the model simple, we assume that each monomer has only one virtual phosphorylation site, we do not distinguish between different phosphorylation sites on the receptor, and hence, the monomer can only be in one of the two states – it can be either phosphorylated or unphosphorylated.

Phosphorylated dimers are being rapidly removed from the cell surface by receptors internalization, also called endocytosis. Our model contains two cellular compartments – the plasma membrane and the cytoplasmic compartment, and the endosomal compartment. Endocytosis is implemented as a first order reaction turning plasma membrane receptor species into corresponding

endosomal receptor species. Signaling from endosomal phosphorylated receptors is allowed. In endosomes, receptors can be recycled back to the plasma membrane, or be degraded, or be dephosphorylated by endosomal phosphatase.

The endosomal RTK phosphatase is implemented as a separate species, distinct from the cytoplasmic RTK phosphatase. Both phosphatases act on phosphorylated dimers and are encoded as first-order reactions. Such implementation is supported by the findings that different phosphatases are overabundant throughout the cell and might have different catalytic properties. Hence, a first-order rate constant of dephosphorylation represents a unified dephosphorylation action of phosphatases in each of the two implemented compartments.

In addition to internalization and recycling of phosphorylated dimers, our model implements basal, ligand-independent internalization, degradation and recycling of inactive receptors. For the ErbB receptors, basal internalization is known to be an order of magnitude slower than the EGF-induced internalization (Becker et al., 2010). In addition, our model implements ligand-independent production of EGFR to the cell surface. The receptors production constant effectively encompasses both, receptors synthesis and their delivery to the cell surface.

Downstream of the EGFR receptors the model implements signaling to two targets, ERK and AKT. The kinase signaling cascade between EGFR and ERK is simplified, grouping details of receptors interactions with scaffold proteins and multiple intermediate kinases into one effective signaling component. Since measuring abundances and dynamics of signaling responses at scaffold and adaptor proteins is outside the scope of this thesis, the interactions with scaffold and adaptor proteins are encapsulated in effective rate constants of reactions between receptors and downstream kinases.

Similarly, signaling cascade leading to AKT activation is reduced to one effective intermediary component. Such simplifications still allow to fit well the experimental data with a substantially reduced number of model's parameters. However, we implement ERK activation via double phosphorylation of ERK in our model. Deactivation of both ERK and AKT is implemented via dephosphorylation by their own corresponding phosphatases.

47

To keep most essential features of the model and prevent the model from growing in the number of reactions, we assume several simplifying assumptions, supported in literature. First, phosphorylated dimers are not allowed to bind and unbind EGF, as we assume that phosphorylation and dephosphorylation, and dimerization and undimerization processes are relatively faster than the diffusion-limited EGF binding and unbinding. We assume that receptors phosphatase removes two phospho groups in one step (as opposed to two steps), as otherwise, dimers with one phosphorylated receptor would be rapidly re-phosphorylated.

The EGFR signaling model has been implemented in Jacobian, a reaction-engineering program from NumericaTech (Cambridge,MA). Jacobian implements numerical integration methods optimized to sparse systems, such as the EGFR signaling model. The model implementing EGFR and ERK dynamics consists of 54 reactions, encompassing 26 species and 15 rate constants listed in Section A.1.

Section 3.2: Bayesian sampling of model parameter space: Markov Chain Monte Carlo

Our mechanistic dynamic ODE model contains multiple parameters which are infeasible to measure experimentally *in vivo*. Literature derived estimates of parameter values may come from very different experimental biological systems, might have been obtained from assays performed outside of the living cells. In addition, given the large number of parameters (up to 29 non-zero initial conditions), it is not feasible to collect such experimental data set of signaling responses to constrain the model. Hence, the resulting models we develop are still vastly underdetermined.

Since the main goal of the developed models is to be able to reproduce the experimentally observed data, as well as to generate novel predictions, parameter unidentifiability presents a significant problem. Specifically, given that multiple different sets of parameter values might all fit the observed data equally well, yet, differ significantly in their prediction of unseen behavior, traditional single-point parameter estimate approaches are not applicable in light of parameter unidentifiability.

Hence, we develop and apply a principled probabilistic framework to address the problem of models' parameter unidentifiability. We use Markov Chain Monte Carlo (MCMC) techniques to

probabilistically sample models' parameters space regions which are most likely to have produced system behavior consistent with the experimentally measured data.

MCMC approach is used to sample a probability distribution that is difficult to sample directly (Gilks, Richardson, and Spiegelhalter, 1996). MCMC chain will perform a random walk over the parameter space, at each point evaluating how well the model with a given parameter assignment fits the observed data (see description of likelihood calculation in Section 3.2.2 and Section 3.2.3). Steps that improve the fit are always accepted, whereas, steps that decrease the likelihood of the fit can also be accepted, but with the probability of acceptance decreasing as the fit's likelihood decreases. Such transition rule allows to sample regions of high likelihood, as well as to move from one highly probable region to another through regions with lower likelihood.

Let $\pi(y)$ be the true underlying probability distribution over parameter space, y, that we want to sample. The true stationary distribution $\pi(y)$ we are interested in sampling from represents the probability of parameters space given the observed experimental data X :

$$\pi(y) = P(y \mid X)$$

Since estimation of $\pi(y)$ would require its evaluation over the multi-dimensional parameter space of the model, it is infeasible to compute it deterministically, which is why we use an MCMC framework to estimate it. To sample from $\pi(y)$ MCMC works with a proposal distribution, $Q(x)$. If the Markov Chain transition rules adhere to the detailed balance property, described below, MCMC will generate a statistically-dependent sequence of states from the target distribution $\pi(y)$ by sampling the proposal distribution $Q(x)$.

The detailed balance condition requires that the transition function $A(x,y)$ of the Markov Chain is such that the probability of being in point x and transitioning to point y at any given time is the same and thus satisfies the following:

$$A(x, y)\pi(x) = A(y, x)\pi(y)$$

The detailed balance property together with the conditions of aperiodicity and irreducibility guarantee that independently of the form of the proposal distribution, the MCMC chains will sample the stationary, target distribution $\pi(x)$.

Metropolis-Hastings algorithm provides a convenient way to construct the proposal distribution $Q(x)$ with detailed balance property. The algorithm starts with an initial state x_0 and generates subsequent samples according to a transition function, $A(x,y)$:

(*)
$$A(x, y) = Q(x, y) \min\left\{1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right\}$$

$Q(x, y)$ represents the probability to propose a step from point x to point y. In our work we use Gaussian proposal probabilities centered at the current point x and with variance σ specific to the particular parameter dimension. Choosing Gaussian proposal distribution favors generation of points closer to the current point x. In addition, Gaussian proposal function provides symmetry in transition probabilities from x to y and vice versa:

$$Q(x, y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-y)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}} = Q(y, x)$$

To estimate the transition probability given by the expression (*) we do not need to actually compute $\pi(y)$ at every point, but instead, we need to compute the ratio $\dfrac{\pi(y)}{\pi(x)}$ of its value at a proposed point y to the ratio at a current point x which using Bayes theorem can be expressed as follows:

$$\frac{\pi(y)}{\pi(x)} = \frac{P(y \mid X)}{P(x \mid X)} = \frac{\dfrac{P(y, X)}{P(X)}}{\dfrac{P(x, X)}{P(X)}} = \frac{P(y, X)}{P(x, X)} = \frac{P(X \mid y)P(y)}{P(X \mid x)P(x)}$$

Where $P(x)$ and $P(y)$ are the prior probabilities on parameter values x and y, and $P(X \mid y)$ and $P(X \mid x)$ are the likelihoods of the experimental data at these points in the parameter space. Likelihood score of the data in a given parameter point estimates how well the fit at that point in parameter space approximates experimentally observed behavior. $P(X \mid y)$ is usually estimated based on some error function, which, for example, could be represented as a Gaussian function:

$$P(X \mid y) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(S_i - X_i)^2}{2\sigma_i^2}}$$

where the product is taken over all experimental data points X_i fitted in the simulations, with their associated experimental errors σ_i, and S_i being corresponding solutions produced by model simulations.

The MCMC chains can have an additional parameter, referred to as temperature, T, which essentially can flatten out the likelihood function landscape by scaling down the difference between the probabilities of the current and the proposed steps:

$$P(X \mid y) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(S_i - X_i)^2}{T*2\sigma_i^2}}$$

If $T > 1$, then the ratio of the likelihoods between the proposed and the current step is decreased, compared to the case if $T = 1$ and therefore the step has a higher probability of being accepted. The MCMC simulations ran with $T > 1$ are referred to as heated.

To run MCMC chains over parameter space, one first needs to establish the prior probability distributions over the possible ranges of parameter values. We define the prior distributions of parameters based on generally accepted literature-derived estimates and extend that range by one or two orders of magnitude around the literature-derived values. In this way, we do not impose any specific point constraints on parameter values, but only limit their ranges to biologically relevant regions. The posterior distributions of parameter values we derive are not imposed artificially, but we derive them based on their likelihood to fit the experimental data. Next, the obtained posterior parameter distributions allow us to generate probabilistic predictions of unobserved model behaviors and postulate new hypotheses about the studied system.

Section 3.2.1: Parameter scan for the simplified and the detailed models

We apply the same probabilistic framework to simulate both the detailed mechanistic ODE models of the signaling network as well as the simplified analytical models derived for theoretical demonstrations presented in this work. The difference is that, in the case of the detailed models, we

fit the models to the experimental data, whereas, for the simplified models, we fit the model only to reproduce certain features of the signaling responses observed in the experimental data. For example, the list of features might include perfect adaptation, sensitivity to consequent EGF stimulations, non-zero amplitude above a certain threshold over the basal level, signal decay within a specific time frame. Hence, the application of the framework for the different types of models requires using different likelihood functions.

Section 3.2.2: Likelihood function in the detailed model

We apply our probabilistic framework to fit the detailed ODE model of EGFR signaling to the experimental time-course of phosphorylated targets and the total EGFR in response to treatment with constant doses of EGF over three hours.

MCMC chains explore the multi-dimensional parameter space by random walk guided by a set of Metropolis-Hastings transition rules. In this section we describe the likelihood function used to calculate the transition probabilities at each step of the MCMC chain. Each point in the parameter space is evaluated by a likelihood function describing how well the model fits the observed data given this parameters assignment.

Let $X = (X_1, X_2, \ldots, X_n)$ be a set of experimentally measured data points representing responses of a given target (e.g. pEGFR or pERK), to a given experimental condition (such as treatment with a certain EGF dose). To compute log likelihood score of each individual experimental data point X_i we assume the following generative model for our experimental data. Consider model's fit produced at a given parameter assignment θ_o. The fit, represented as $\mu^{\theta_o}(t)$, is a simulated time-course response of a target at a treatment with a given dose of EGF ligand with parameter values assignments θ_o. We then suppose that the fit represents a true biological behavior of the system in response to treatment with that dose of ligand, and that each experimental point, X_i, we observe has been generated randomly from normal distribution with mean $\mu_i^{\theta_o}$ and standard deviation σ_i. The mean $\mu_i^{\theta_o}$ represents an average of the experimentally-measured response at the

time point corresponding to the time of measurement of the point X_i, and the standard deviation σ_i is the error of that experimentally measured data point (Figure 3.1).

We compute the log-likelihood score of model's fit to each individual experimental data point. Each data point has a standard error associated with it. In case the model is fitted to a single cell distribution, the error represents the actual standard deviation of single cell responses, and the data point itself is the mean of that distribution; if the model is fitted to the ensemble averages data, the error represents the error in the mean, whereas the data point itself captures the average of the independent replicates of the experiment (see Section 4.1 for detailed description).

The probability of an individual experimental data point X_i given the parameter assignment θ_o can then be calculated as:

$$P(X_i \mid \theta_o) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(X_i - \mu_i^{\theta_o})^2}{2\sigma_i^2}}$$

Assuming independence of individual experimental data points, we can compute the probability of the entire experimental data set X at parameter assignment θ_o:

$$P(X \mid \theta_o) =$$

$$= \prod_{i=1}^{n} P(X_i \mid \theta_o) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(X_i - \mu_i^{\theta_o})^2}{2\sigma_i^2}} \quad (1)$$
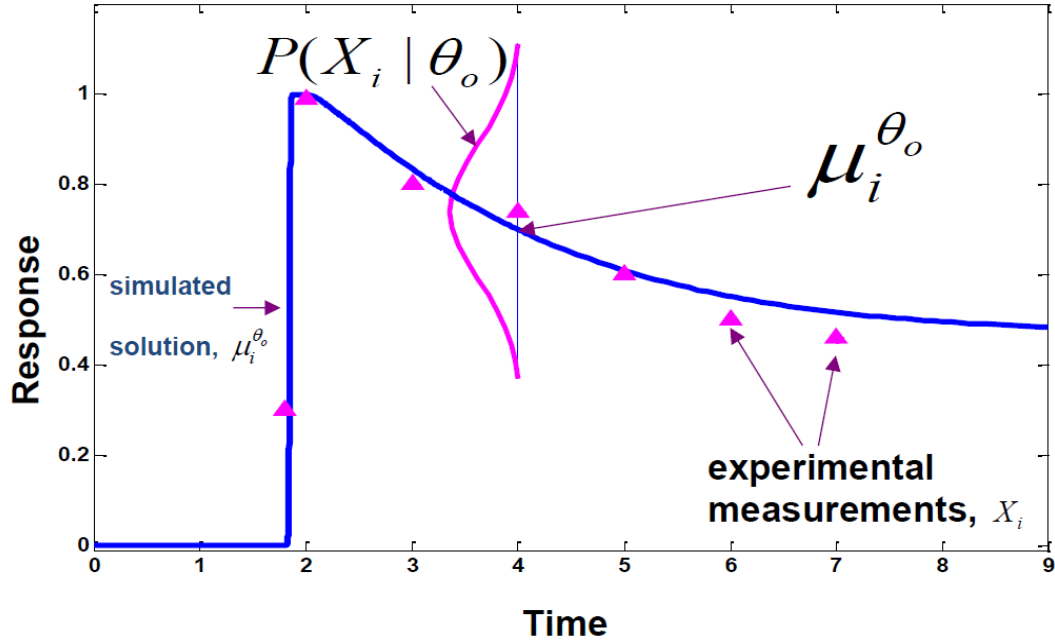
Figure 3.1: Computing likelihood of the experimental data given a model fit

In this derivation we assumed the normality of distributions from which the experimental points were generated. Such approach is applicable when, for example, experimental points X_i represent an average ensemble behavior of the cells, and σ_i are the corresponding errors in the mean. If, however, we chose to fit the model to a distribution of single cell responses, which in their shapes might deviate from the normal distribution, we could compute the likelihood function of the data assuming other than normal probability function. For example, the gamma distribution could be appropriate for representing skewed distributions (Figure 5.7). In other cases, the likelihood could be computed based on the directly derived probability density function of the single cell distribution at each experimentally measured condition.

We assume that individual targets, pERK, pAKT, pEGFR and tEGFR are independent of each other; therefore, to compute the overall log-likelihood of the data we can sum up together all their individual log-likelihood scores. However, in reality, the assumption of the targets independence may not necessarily hold, as, for example, both pERK and pAKT activation depend on pEGFR signal.

To compute the overall log-likelihood score for each target, we sum up the individual log-likelihood scores of this target's responses to each of the applied EGF doses fitted in the simulation.

54

For example, if in a certain simulation we choose to fit the model to pEGFR responses at high, medium and low EGF doses, at every point in the parameter space we will separately simulate model responses to the three given EGF doses, separately compute the log-likelihood of each response and report the sum of the three log-likelihoods. Such summation of the log-likelihoods relies on the assumption that target's responses to different EGF doses are independent between themselves, which may not necessarily be true, as given some observed responses to several doses may affect the probability distribution of responses at a different dose.

To compute the log-likelihood score of a time-course of a target's response to a given dose of EGF, we sum up the log-likelihood scores of individual time points fitted in the data, relying on the assumption of their independence from each other. However, in reality, phosphorylated responses at a given time point are probably dependent, and even determined by, phosphorylation responses at preceding time-points.

To address a concern that the experimental data points for different conditions might not be strictly independent, we consider the following argument. The experimental data in this work is obtained from separate individual wells treated with a specified condition and measured at a specified time. Each experimental measurement has a compounded error associated with it consisting of a measurement device error and a human-based error, which cannot be controlled absolutely, but can be limited with careful experimental practices. Still, if the measurement device error has a predictable distribution and is expected to be the same in measurements of different experimental data, the error from the experimental handling is not guaranteed to be the same on different days. Therefore, even though, the underlying biological quantities, such as phosphorylation signal at consecutive time points might represent dependent quantities, their measured values are treated as independent estimates due to the impossibility to reliably quantify the human-based error.

Using a large number of experimental data points in computation of the likelihood of the entire data set can over-constrain the sampling of the underlying parameter space. For example, a change in parameter values only slightly affecting the solution of the model could have a significant added-up change in the likelihood of the data, and therefore, unlikely to be accepted, resulting in a high rejection rate in the MCMC chains, and preventing efficient mixing of the MCMC chains in the

parameter space. We partially address such concern by applying established heuristic approaches, such as adaptive-step size MCMC chains and increasing the temperature at which the chains are run (Section 3.2).

Section 3.2.3: Likelihood function in the simplified model

In the simplified model described in Section 4.7 we do not fit the experimental data itself, but impose constraints on certain features of model responses, such as requiring adaptation of model responses to a pre-stimulus level within a given time-frame, or requiring the peak responses to be monotonic in EGF doses, and others. Depending on the goal of simulations and the imposed conditions, likelihood functions can very.

In the simplest case, we require that system's responses to stimulation with a constant non-zero dose of EGF over several hours are non-trivial and decay. Specifically, we require that the maximum response rises above a certain threshold over its pre-stimulus level and that the response at the end of the observation is lower than the maximal response. For example, Figure 3.2 shows hypothetical responses of the simplified model to the step input $u_o \rightarrow u_1$ produced at two different parameter values assignments. The dark blue fit has $y_{max} > y_o$ and $y_{end} < y_{max}$ – producing a non-trivial, decaying signal response, and thus, the parameter assignment which produced such fit will be accepted. Although the light blue fit produces an increase in the response to the step input ($y_{max} > y_o$), it will be rejected as the fit does not decay by the end of the simulation ($y_{end} \sim y_{max}$.)
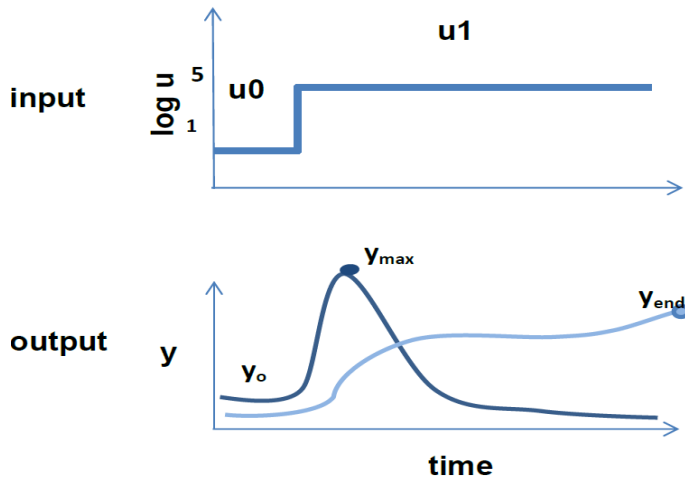
Figure 3.2: Diagram of two possible fits produced by the simplified analytical model.

Additional constraints on model responses to be included in the computation of the log-likelihood function are the monotonicity in the peak of the responses. For example, if we consider responses to two step inputs $u_o \rightarrow u_1$ and $u_o \rightarrow u_2$, where $u_1 < u_2$, we could require that $y_{max1} < y_{max2}$.

In another type of simulations we impose perfect adaptation on model responses. At each point in parameter space system responses to three doses u_1, u_2, u_3 are simulated. The doses are fixed for all simulations and are such that $u_1 < u_2 < u_3$, are equally distanced in the log space, i.e. $\frac{u_2}{u_1} = \frac{u_3}{u_2}$, and cover at least one order of magnitude $\frac{u_3}{u_1} = 10$. Such requirement on the input doses insures that the fitted effect (for example, perfect adaptation or Weber's Law) holds over a significant range in input doses and is not just a trivial artefact due to similarity of the initial inputs.

To impose perfect adaptation, first, we require that responses to all u_i significantly decay from their corresponding peaks at the end of ligand treatment (empirically derived threshold of $y_{end} < 0.2 * y_{max}$). In addition, we minimize the distance between the responses at the end of simulations by minimizing the coefficient of variations (CV) of the responses with a squared error function. However, the similarity of the responses to three different doses at the end of ligand treatment does not guarantee that the responses have reached steady state and are thus indeed adapted to their perfect adaptation levels. Hence, in addition, we numerically solve the system for steady state responses to each of the three doses and impose that by the end of ligand treatment

57

(three hours) the response at each dose is within 10% of the corresponding to that dose steady-state response value.

In Section 4.7.4 we describe parameter distributions obtained from fitting Weber's Law in the simplified model. To impose Weber's Law constraint in the simplified model, at a given parameter assignment p_i we simulate model responses to the three step inputs $u_1 \rightarrow \alpha u_1, u_2 \rightarrow \alpha u_2, u_3 \rightarrow \alpha u_3$ with same fold change, α, and different initial background levels of stimulus u_1, u_2, u_3, as described for fitting perfect adaptation in previous paragraphs. We simulate system responses pre-adapted to stimulations with u_1, u_2, u_3 and impose that the produced responses are non-trivial, adapt within the time of simulation, and have similar amplitudes, despite an order of magnitude difference in the basal input levels u_1, u_2, u_3. We quantify the similarity of the three maximal amplitudes by their coefficient of variations, thus, not constraining the model to produce outputs of any specific absolute levels, but allowing them to vary (Figure 3.3). The smaller the obtained CV, the more similar the peaks are to each other. An empirically derived CV threshold of 0.1 suggests that the responses exhibit approximate Weber's Law (Goentoro et al., 2009). In addition, comparing the amplitudes of the responses to step inputs with CV as a metric provides a distance metric applicable to assess the degree of Weber's Law across a variable number of responses. For example, Weber's Law could be imposed more strictly on an interval of EGF doses by minimizing the CV of responses from more then just three points from that interval.

All points in the parameter space which do not satisfy the conditions for perfect adaptation, non-triviality of responses, and monotonicity of the peak responses at the initial treatment with u_1, u_2, u_3 , are rejected. If the preliminary conditions are satisfied in a given point, the likelihood score of the peak amplitudes is computed and decision to accept or reject the proposed step is made.

Requirement that all model's responses at the three different doses decay to within 10% of the pre-stimulus level does not require that the responses decay within the same time, thus, allowing a flexibility in their rates of decay. Restraining the responses to all three doses to decay with the same dynamics, in addition to requiring that the responses have same amplitude and perfect adaptation, would generate fits with a so-called property of fold-change detection, which entails Weber's Law, but is rather a more stringent property (Shoval et al., 2010).
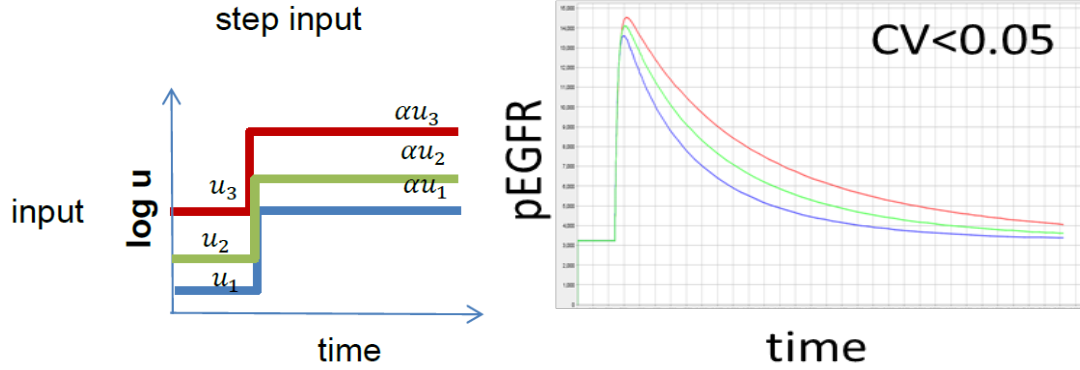
$$CV(p_i) = \frac{std(y_{max}u_1, y_{max}u_2, y_{max}u_3)}{avg(y_{max}u_1, y_{max}u_2, y_{max}u_3)}$$

$$Likelihood(p_i) = (CV(p_i))\hat{\,}2/(2*error\hat{\,}2)$$

Figure 3.3: Imposing Weber's Law on system responses to three step inputs with the same fold increase, α and different initial background levels of stimulus u_1, u_2, u_3. (A) Input time-courses. (B). Example of a fit with low coefficient of variation in amplitudes of the responses which also satisfies conditions for perfect adaptation and non-triviality of the responses. (C). Computing the likelihood score of the fit

Section 3.2.4: Prior distributions on parameter values

Our model parameters consist of the non-zero kinetic rate constants of reactions and the non-zero initial conditions of protein concentrations. At the beginning of the simulations prior to ligand addition, it is assumed that all receptors and downstream targets are inactive and therefore, abundances of the phosphorylated species have zero initial conditions and are not simulated as model's parameters. We use literature-based estimates of parameter values to derive the prior probability distributions on the parameters. In particular, we do not constrain model's parameters to the specific literature estimates, instead, we use those estimates as a basis for the range of possible values the model parameter can assume. We allow the range to be at least one or two orders of magnitudes around the literature estimates to accommodate various different literature estimates if they are available and to allow for parameter flexibility. We impose log-uniform prior distributions over the derived ranges. Parameters description and their prior distributions are listed in Section A.1.

Literature estimates of ErbB expression in mammalian cell lines are usually given within an order of magnitude precision. Experimental errors and variability in handling and media conditions to which cells have been exposed prior to receptor abundance measurements can result in fluctuations in the reported receptor abundances. However, there is a consensus in the literature that MCF-10A

cell lines express on the order of 10^5 ErbB1, whereas SKBR3 – 10^5, and MCF-7 -10^4 receptors per cell. We use ELISA assays (Lequin, 2005) to quantify average per cell abundance of EGFR in a population of cells in a given well. Our experimental measurements suggested that MCF-10A expresses about 300,000-600,000, whereas SKBR3 had about 250,000, and MCF-7 about 51,000 receptors per cell. To account for the fact that receptor abundances are measured with precision of up to an order of magnitude, in our model we let the priors on ErbB concentrations vary within one order of magnitude of the experimentally measured values.

It is experimentally infeasible to directly quantify fractions of total number of EGFR receptors present throughout different cellular compartments. To the best of our knowledge, not many studies have addressed this question with direct experimental measurements of total EGFR. Hence, in our simulations, we assign all receptors to be on the cell surface in the inactive state, rather than simulating various possibilities of distribution of the receptors among cell compartments. Such approach does not present a problem, since for every fit shortly after simulating ligand addition receptors are being redistributed from empty cell surface receptors into other receptor species.

Since we reduce the details of signaling cascades between EGFR and its downstream targets into effective rate constants, we do not need to simulate abundances of various intermediate proteins, such as, for example, scaffold proteins. We simulate total initial abundances of ERK and AKT. Our previous experimental measurements indicated that there are approximately 700,000 ERK molecules per cell, and approximately 600,000 AKT molecules per cell in the mammalian cell lines MCF-10A we are working with. In the model, we allow 10^4-$10^{6.5}$ variability for both total ERK and total AKT abundances.

Although many literature estimates of enzymatic rate constants are available, most of these rates are derived based on *in silico* assays, and outside of the *in vivo* context, or based on computational simulations, and hence, might differ substantially from the actual values *in vivo* (Chen et al., 2010). Hence, as our priors for rate constants we take literature-derived estimates and allow up to two orders of variation above and below the literature values to account for the differences in the biological backgrounds from which the estimates were derived from and the experimental errors in such assays. The derived priors thus allow to encode sufficient flexibility on parameter values while

still retaining the information about relative differences between various model parameters. For example, the rate of degradation of phosphorylated EGFR is believed to be approximately an order of magnitude slower than the rate of receptors autophosphorylation (Chen et al., 2009), which is reflected accordingly in the shifted ranges of the prior values of these parameters in our simulations. Kinetic rate constants of the reactions related to ERK activation have been studied extensively within the context of understanding dynamics of ERK signaling; the literature-reported estimates of such rate constants are incorporated into the priors of ERK- related parameters in our model (Fujioka et al., 2006; Aoki, Yamada, Kunida, Yasuda, & Matsuda, 2011).

We assume the prior distributions are uniform over the log range of parameter values, to avoid biasing the priors towards any particular experimental estimates derived from the literature. Alternatively, we could impose log-normal priors over the parameter space, or fix the parameter values to very specific point estimates if such precise measurements were available.

Section 3.2.5: Simulated Annealing to find the starting points for the MCMC chains

To optimize the overall running time of the MCMC chains, we perform simulated annealing (SA) searches to find starting points for the chains. SA allows efficiently exploring the model parameter space and finding the points with good likelihood scores. Hence, starting the MCMC chains from the SA-found points allows reducing the simulation time necessary for the MCMC chains to mix. The prior probabilities on parameter values, the likelihood function of the experimental data, and the transition rule for SA chains are the same as for the MCMC chains we run. For each sampled condition, we run sixty different simulated annealing chains. Each SA chain starts from a randomly generated point in the parameter space. Each chain is first run at a high temperature (in our case T=10) for 100-200 steps per parameter, followed by a long run at cooler temperature, T=1, for 1000-2000 steps per parameter. The cooling schedule was derived empirically, based on the properties of our system. With such cooling schedule, SA chains usually find reasonable fits to the data. Each chain reports its best-likelihood point in the parameter space and the points across all chains are ranked. Top 5 to 10 points with best likelihoods are selected as starting points of the MCMC chains for sampling.

Starting multiple SA chains from randomly generated over-dispersed starting points allows a reasonable coverage of the overall parameter space. Running multiple chains with heating and cooling ensures the chains do not get stuck and explore various regions in the parameter space.

Steps in the SA chains and in the MCMC simulations are made in one parameter at a time, in a fixed order of parameters. In all chains, the steps are performed in log-space of parameter values to ensure an efficient coverage across parameter space. At every step, the magnitude of the proposed step size is first generated from a normal distribution with the same variance for all parameters. If the proposed position is outside of the prior range for that parameter, the step is rejected and the chain moves to the next parameter. Otherwise, the likelihood of the data is evaluated at the proposed point and accepted or rejected according to the chain transition rules.

Section 3.2.6: Running MCMC chains

We start several independent MCMC chains from each of the top 5 to 10 point found by SA. The starting points may vary in their likelihood scores up to one or two orders of magnitude. We run multiple MCMC chains in parallel until approximately 2,000-5,000 steps per parameter have been simulated. The first 500-1000 steps per parameter are discarded as a burnoff during which the chains are "forgetting" their starting points and mixing in the parameter space. We then thin each chain by reporting only every Nth step, where N is the number of parameters simulated. To reduce autocorrelations within the chain, we report running average on the reduced chains. If the autocorrelation is still significant, we further thin the chains by increasing N, until autocorrelation is reduced but chains still contain sufficient number of points.

The acceptance rate of each chain is monitored for each individual parameter, and not for the overall chain, and is aimed to be within 20-30%. The chains are started with the same step size for all parameters. Initially, for each parameter, the acceptance rate is recorded and step size is adjusted in order to maintain the acceptance rate at its target level. The acceptance rate is calculated every 20 attempted steps, and the step-size is increased by 25% if the acceptance rate is too high, or decreased by 25%, if it is too low. If the acceptance rate is within the set limit, step size is kept unchanged. Through step size adjustments the step size may decrease to very low values. Hence,

there is a lower limit on the step size for all parameters –it is set to 0.075 to prevent chains from being stuck in limited areas in parameter space. Initially, every parameter is allowed to have up to six step size adjustments, to find a step size suitable for each parameter. After the six attempted step size adjustments the step size is being kept constant for the rest of the simulations. The segment of each chain in which parameters' step size adjustment is performed is being discarded as part of a burnoff. Customizing the step size for each individual parameter based on its acceptance rate takes into account features of direction specific to each parameter space and helps to improve mixing of the MCMC chains and to speed up their convergence.

Section 3.2.7: Metrics of Convergence of MCMC chains

We derive the posterior distributions of parameters based on samples obtained from running multiple MCMC chains. However, to ensure we obtain the true estimates of the parameter distributions we first need to make sure that the parallel MCMC chains have converged. In other words, we need to make sure that overall the chains have visited and sampled same regions in parameter spaces. Methods to assess convergence of MCMC chains mostly rely on heuristics (Brooks & Gelman, 2010). Given the parallel MCMC chains adhere to the same transition rule and use the same likelihood function, although they are started from points dispersed in the parameter space and are run independently of each other, after sufficient number of steps the chains are guaranteed to converge to the true underlying distribution due to maintaining the detailed balance requirement in the MCMC chains. In practice, we cannot run the MCMC chains indefinitely; hence, monitoring the convergence allows to assess if the chains have not converged yet and need to be run for longer time. If after multiple steps, the convergence of the MCMC chains is not changing significantly and the chains have sufficiently converged, we stop running the MCMC chains.

We monitor the MCMC chains convergence for one parameter at a time. First, for all chains we discard their burnoff regions. We then visually inspect the traces of multiple MCMC chains for every given parameter by plotting them on a graph with iteration step on the X axis and value of each chain at that particular step on the Y axis. We can thus identify the chains that are stuck in certain regions of parameter space, or chains which are not mixing well. In addition, we can detect if there is

a general trend for chains to converge, suggesting for how many more steps the chains need to be run for to achieve a better convergence score. In case that the chains appear to be well-mixed in certain parameters but stuck in others, we redistribute the attempted steps in each chain to be done in the stuck parameters, thus, possibly speeding up the mixing of the chains in such stuck parameters.

In addition to the visual inspection of chain trajectories we apply a frequently used Gelman-Rubin heuristic criterion to monitor the convergence of the MCMC chains (Brooks & Gelman, 2010). Briefly, the method considers within-chain and between-chain variances and computes the ratio of an over- and under-estimates of the variance of the true underlying distribution. When the obtained ratio is relatively small, the between-chain variance is small compared to the within-chain variances, thus suggesting that the chains are approaching convergence. Detailed description of the Gelman-Rubin criterion formula is given in Section A.3.

Section 3.2.8: Derivation of the posterior distributions of parameter values

Given the observed experimental data fitted in the model, X, from the converged MCMC chains we obtain the probability distributions of parameters θ using Bayesian approach. Overall, such distributions are more realistic than the literature estimates, since they are based on the real experimental data specific to the conditions of our system. Let

X − observed experimental data set
θ − point in parameter space
$P(\theta)$ − prior probability of the parameters
$P(X \mid \theta)$ − likelihood of the data
$P(X)$ − normalization constant (probability of the data)
$P(\theta \mid X)$ − posterior probability of the parameters

We can then write down the posterior probability of parameters as follows:

$$P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)} = \frac{1}{P(X)} P(X \mid \theta)P(\theta)$$

Therefore, up to a constant, the posterior probability distribution over parameter values is a product of the prior probability distribution of that parameter and the likelihood of the experimental data in that point in parameter space.

Although it is infeasible to compute the normalization factor $\dfrac{1}{P(X)}$, properties of the transition rules of the Metropolis-Hastings MCMC chains allow to sample the $P(\theta \,|\, X)$ by simplifying the computation of the transition probabilities from point in parameter space θ_i to θ_j by expressing them as ratio of likelihoods and priors of the corresponding points:

$$\frac{\pi(\theta_j)}{\pi(\theta_i)} = \frac{P(\theta_j \,|\, X)}{P(\theta_i \,|\, X)} = \frac{\dfrac{1}{P(X)} P(X \,|\, \theta_j) P(\theta_j)}{\dfrac{1}{P(X)} P(X \,|\, \theta_i) P(\theta_i)} = \frac{P(X \,|\, \theta_j) P(\theta_j)}{P(X \,|\, \theta_i) P(\theta_i)}$$

In practice, to obtain the posterior parameter distributions we combine the simulated points in the parameter space from all converged MCMC chains. For each parameter we compute a normalized histogram over the entire range of parameter values, which was initially specified in the prior of that parameter. The obtained histogram can then be interpolated into a continuous probability density function. In our analysis we can choose to work with either the original normalized histogram or with the interpolated, continuous distribution.

However, since the MCMC chains might not necessarily converge in all dimensions of the parameter space, to generate the posterior distributions we combine points from the MCMC chains taking into account the average likelihoods of the chains from which the points were obtained. Such approach relies on assumptions that the MCMC chains which do not converge perhaps sampled two distinct high-likelihood regions in parameter space, yet were unable to mix between the two different regions. We then combine the samples from each of the region according to the average chains likelihood score, which would allow rating the regions between themselves relative to their average likelihoods. To obtain the corrected MCMC sample we perform the following steps:

1) Consider all points from the thinned approximately converged MCMC chains with discarded burnoff

2) Normalize the likelihood of each point by the sum of the likelihoods of all points in the combined set of points, thus obtaining the probability distribution over the sampled points

3) Generate N random points from the collection of all the points according to the probability density function obtained in (2).

4) Report the probability distributions obtained from (3).

Reweighting the posterior distributions according to the likelihoods of the chains thus allows to correct for the possibility that the MCMC chains have not converged perfectly, and to obtain a better approximation of the true underlying posterior distribution of model parameter values.

Section 3.2.9: Derivation of the predictive distributions of the EGFR signaling model behaviors

Our method allows to conduct predictive inference on our model behaviors. Given the observed data, X, we want to obtain a probability density function of a new, previously unobserved data point, x. We marginalize the probability distribution of the unobserved data point given the fitted experimental data over the parameter space. We then derive that the probability of the unobserved behavior, up to a constant, is proportional to a product of the prior of the parameters, the likelihood of the fitted experimental data given parameter assignment and the conditional probability of the predicted value, itself, given the parameter assignment, this product is then integrated over the entire parameter space:

$$P(x \mid X) =$$
$$= \int_{\theta} P(x, \theta \mid X) \partial\theta =$$
$$= \int_{\theta} P(x \mid \theta, X) P(\theta \mid X) \partial\theta =$$
$$= \int_{\theta} P(x \mid \theta) P(\theta \mid X) \partial\theta =$$
$$= \int_{\theta} P(x \mid \theta) \frac{P(X \mid \theta) P(\theta)}{P(X)} \partial\theta =$$
$$= \frac{1}{P(X)} \int_{\theta} P(x \mid \theta) P(X \mid \theta) P(\theta) \partial\theta$$

X – observed experimental data set
θ – point in parameters' space
x – value to predict

Based on the above expression, we can generate the predicted distribution of the unobserved system behavior, x, from model responses simulated across the points in the parameter space, according to their posterior distributions obtained from fitting the model to the given experimental data X (Section 3.2.8). To actually obtain the predicted distributions of values of x, in our simulations we fit the model to a given set of experimental data X, and at each step in the parameter space we record the simulated model response to the experimental condition corresponding to x, which we want to predict. For example, if we fit the model to early phosphorylation responses and aim to predict model response three hour of constant EGF stimulation, then at each given point in the parameter space we simulate model response up to three hours, use the specified early responses to evaluate the goodness of the predicted fit (i.e. compute the likelihood of the data), and record the response at the three hour time point as a prediction. In case we predict model's response to a perturbation other than the treatment with a constant dose EGF to which we actually fit the model, at each step we simulate model responses to such perturbation after we have simulated the necessary constant EGF dose condition. Therefore, the simulated MCMC chain would not only produce parameter values assignments at each step, but also, the values of the response being predicted.

In addition to predicting system's responses at various times of ligand treatment, or to inhibitor perturbations, we can also generate predictions of any other variables in the system, including ones which were not measured experimentally or not fitted in the simulations. For instance, we can generate probabilistic distributions of the initial abundances and the time-courses of

intermediary signaling components, such as, for example, the activated or total MEK, or time-courses of inactive receptor dimers, and others (see Section 4.2 for examples).

The MCMC chains generate predicted distributions of values of experimental points to which the model is actually being fitted. Although, the model likelihood is computed directly based on the fit in these points, the predicted distributions shows some deviation from the experimental point, rather than being exactly at the imposed experimental value. This observation can be explained by the fact that we fit multiple experimental data points simultaneously, computing the likelihood score based on the fit to all points: while some fits may have similar overall likelihoods, they may vary between the likelihood scores to the fits at certain points. In addition, the fact that the model's multidimensional parameter space is vastly underdetermined could also contribute to the variability in the predicted values of the fitted points.

Section 3.3: Statistical method to compare parameter distributions and to identify parameters sensitive to specific experimental data points

Traditional approaches to assess model parameters sensitivity often rely on analyses ran at several points in model parameter space by varying one parameter at a time at those points and examining the induced change in system responses (Kohberger, 1978; Gutenkunst et al., 2007). We develop a novel statistical method to investigate global parameter sensitivities to specific features of model responses. Our probabilistic framework allows to globally assess model parameter sensitivity to specific experimental data points across the multidimensional parameter space.

More precisely, by fitting the model to the experimental data set with and without the points for a given condition of interest, we can identify parameters which have changed their distributions to accommodate fitting the additional experimental data points. The rationale behind this approach is that fitting additional data should impose more constraints on parameter space which is then reflected in changes in parameter distributions. It is important to note that additionally fitting points which do not necessarily provide new information about system behavior compared to the main set of experimental data, would not necessarily result in detectable shifts in parameter distributions; therefore, this

method is applicable to cases when additional data adds substantial qualitative constrain on system responses.

We run the MCMC simulations and fit the model to a set of the experimentally measured data points $X = (X_1, X_2, \ldots, X_n)$. Next, we add the X_{n+1} point and rerun the MCMC chains. From each of the two types of simulations we obtain probabilistic distributions of parameter values and identify the ones which significantly shifted in their distributions to fit the X_{n+1} point.

To compare parameter distributions obtained from fitting different conditions, we use a metric based on the Kullback-Leibler divergence measure (Kullback, 1987) and several other techniques. Comparing parameter distributions requires that the MCMC chains ran for each of the two compared conditions have converged. The metrics can then be applied to the two distributions obtained by combining individual chains for the same condition together. In case the chains in one condition have not converged, the combined distribution might appear multiple-peaked and heterogeneous, possibly substantially varying from the corresponding distribution from the second condition, in which the chains have converged. We apply a chosen comparison metric to each pair of the distributions for the same parameter, and then rank all parameters by their difference and, if applicable, their significance, corrected for multiple hypothesis testing.

The Kullback-Leibler divergence and Kolmogorov-Smirnoff tests allow to compare the actual distributions between themselves, and not just comparing their statistics. The Kullback-Leibler divergence is an asymmetrical measure which can be used to quantify the difference between two probability distributions Q and R. In its essence, the measure divides the range of values of the two distributions into small intervals; at each interval it then computes the difference between the two distributions and sums them up over the entire range. The higher the Kullback-Leibler divergence of R relative to Q, means that Q is more distinct from R. To obtain a symmetric distance measure from the Kullback-Leibler divergences we take a sum of the two distances from R to Q and from Q to R.

The Kullback-Leibler based metric does not take into account the relative order of columns of the histogram being compared. However, some important information on parameter distributions might be present in the relative order of the columns, especially, given that the histogram-based

distributions we derive might be dependent on binning applied to compute them. Therefore, we also apply the two-sample Kolmogorov-Smirnov test (Corder & Foreman, 2014) to compare parameter distributions taking into account the relative order of the columns of the compared distributions. The two-sample Kolmogorov-Smirnov test is a non-parametric test to compare the cumulative distribution functions of the two distributions. The advantage of the Kolmogorov-Smirnov test is that it is sensitive to differences in both location and shape of the cumulative distribution functions of the two compared samples. When Kolmogorov-Smirnov test is applied to a list of parameters, the returned p-values are corrected for the number of tests performed. The Kolmogorov-Smirnov test is a non-parametric test which does not operate on the averages or the standard deviations of the distributions that are being compared.

To identify parameter distributions which are distinct between the two fitted conditions, in addition to the distribution-based metrics, namely, the Kullback-Leibler divergence and the Kolmogorov-Smirnov test, we use several other measures which rely on comparing statistics of the distributions. This statistic also allows in a more intuitive way to identify parameter distributions which changed between the two conditions. First, we visually inspect parameter distributions from the two types of conditions. We can rank the parameters by absolute differences in the means of their distributions in the two conditions, thus identifying parameters with most shifted distributions between the two conditions. We can also rank parameters by the ratio between their variance of distribution from a condition with the additional fitted experimental data to the variance of distribution with less data fitted: if a given parameter distribution is constrained by the additional experimental data, it may be reflected in the smaller variance of the parameter distribution.

Although, comparing parameter distributions solely by their statistics is a more intuitive way to understand and interpret the differences in model parameter space, such methods may disregard important information offered by the distribution-based comparison methods we mentioned above. For example, in case the two distributions are symmetrical, their averages would be the same, whereas, their distributions would be scored as significantly different by the Kullback-Leibler based metric.

By comparing parameter distributions between two conditions with several types of statistical techniques one can not only gain an insight as to what parameters are important for a given experimental data, but also, an understanding of how the additional constraints can affect the underlying model parameter space. For example, we can find parameters for which a given experimental point shifts their distributions while preserving the shape of the distribution; alternatively, we can find parameters variance of which is significantly constrained by the additional data. Therefore, applying several distribution-comparison techniques we can identify various qualitative aspects of the structure-function relationship between model parameter space and its responses (see Figure 4.12 for examples of comparing parameter distributions).

Importantly, analyzing changes in parameter distributions resulting from imposing additional data constraints and simulating all parameters could also allow to identify global trends in the parameter space which could have been missed in the traditional single-point parameter sensitivity analyses..

The statistical method presented here allows to find parameters important to certain experimental data and is sufficiently versatile. The method can be applied in multiple different ways. For example, we can identify parameters responsible for phosphorylation signal decay by comparing parameter distributions obtained from fitting only initial signaling response to ones obtained by fitting initial and the long term response capturing phosphorylation signal decay (see Section 4.4). Alternatively, we can use our approach to compare parameter distributions in conditions which contain the same set of experimental data, yet in one condition certain points of interest would have more weights contributing to the overall likelihood function. For example, we can identify parameters most sensitive specifically to the early dynamics or only to the peak of the phosphorylation response, by running the second condition with increased weight on those specific points of interest. Moreover, our statistical approach could be useful in case the model is fitted to a combination of experimental data of different types. For instance, fitting receptors degradation data (measured by the total EGFR abundances time-courses) in addition to fitting the phosphorylation time-courses could suggest parameters important to fitting signal decay with a specified level of receptors degradation (see Section 4.2.2). Lastly, instead of finding parameters sensitive to certain measured experimental data

points one could impose various other, not necessarily experimentally detected, features and simulated behaviors and explore parameter structure important for fitting such hypothetical conditions.

Chapter Four: Results

In this section we present the main results of this study. First, we describe in detail the processing of the phosphorylation and receptor abundance time-courses experimental data used to fit the model. We then report various types of experimental data to which the model is fitted and the resulting changes in parameter distributions imposed by each type of data, thus exploring relationships between the model's parameter space and the observed behavior of the system. All experimental measurements described in this section have been performed and collected by our collaborator, Dr. Mario Niepel at the Laboratory of Prof. Peter Sorger at the Department of Systems Biology, Harvard Medical School.

Section 4.1: Processing of the experimental data fitted in the model: methods for scaling, normalization, and combining the data

Section 4.1.1: Cell lines

Experimental data for this work has been collected in mammalian cell lines. Specifically, MCF-10A, MCF-7, and SKBR3 cell lines derived from human mammary epithelium were selected for their high levels of ErbB expression and/or their responsiveness to EGF. MCF-10A cell line was derived from normal, non-cancerous mammary tissue. The MCF-10A cell line is considered to be a reasonable representation of normal mammary tissue, without any known mutations in the ErbB-related signaling cascades. However, since the MCF-10A cell line is an immortalized cell line, it still differs from the original normal epithelium cells from which it was derived. MCF-10A cells express EGFR on the order of 10^5 receptors per cell. We also work with two breast cancer cell lines, SKBR3 and MCF-7. SKBR3 cells express a significant amount of EGFR, estimated to be on the order of 10^5 receptors per cell, and overexpresses ErbB2 receptors, on the order of 10^6 receptors per cell. MCF-7 is known to express ErbB receptors, but the EGFR levels are very low, on the order of 10^4 receptors per cell.

Cells are seeded into the 96-well plates, each plate with 8 rows and 12 columns. Measurements from the wells at the edges of each of the 96-well plate are discarded due to possible plate edge effects, thus, each plate yields 60 wells: 6 rows and 10 columns. The cell density in each

well is determined by the type of the cells incubated in each well and is based on cell volume and confluence levels specific for each cell type derived from our empirical observations. MCF-10A cells are seeded at approximately 10,000 cells per well, SKBR3 at 20,000, and MCF-7 cells at 10,000 cells per well.

In the experiment's run on a single day, the number of cells in each well can differ between 2,000 and 20,000 cells per well. The difference in cell number is due to various factors, such as, deviations in number of cells pipetted during their seeding into the wells, uneven rates of cell loss due to staining with antibodies, and others. The distribution of the cell counts from the wells in a same day experiment usually has a coefficient of variation within 0.08-0.18. However, the average cell count across all wells for a given day can differ more drastically between the days.

The cells are incubated in the standard media for 24 hours, followed by media washout. The media washout is performed to eliminate any possible autocrine secreted ligands. Next, the cells are incubated in the starvation media for another 24 hours, followed by the second washout of the media. Such cell handling protocol has been suggested by empirical work with these cell lines and is thought to better synchronize the cells and induce them into a quiescent state. Following this experimental protocol the MCF-10A cells are at about 70% confluence.

Often, studies of cell signaling use ligand treatment profiles unnatural to the physiological conditions of the cells. For example, in certain studies bolus stimulations with very high doses of ligand are instantaneously added into the cell media (Becker et al., 2010). Although biologically not very likely, such stimulations can still lead to a useful insight about the functioning of the system by elucidating its behavior at maximal activation. In this work, however, we treat cells with various EGF doses, ranging from low physiological to high un-physiological doses, and maintain the EGF dose relatively unchanged over several hours. Such ligand stimulation profile evokes features similar to chronic ligand stimulation which cells might encounter *in vivo*.

Section 4.1.2: Microscopy Imaging

After the cells have been incubated in wells for 24 hours, starved for another 24 hours, had their media washed out, one hour after the final washout, cells are being subjected to experimental or control treatment: each well is being treated with a ligand dose or a control (no ligand) - for a

specified amount of time. ELISA immuno-histochemistry (IHC) fluorescence techniques used in this work to collect data need careful processing of the read-outs (Lequin, 2005; Weemen & Schuurs, 1971). IHC is used to quantify abundance of a target, such as the phosphorylated form of a protein, present in cells. The emitted fluorescence intensity from each well is captured by a microscopy device, quantified by an image analysis algorithm and reported in arbitrary units of fluorescence. At the time of the measurements, primary antibodies for the measured target, pEGFR, pERK, pAKT, total EGFR, or no primary antibodies for controls, are being added, followed by the addition of secondary antibodies. Primary antibodies bind to the target, whereas, secondary antibodies bind to the bound primary antibodies and emit fluorescence. To minimize possible perturbation effects due to experimental handling, - all wells on each plate are being measured at the same time point. Each row on the plate measures a single target; so plates can have rows with different targets measured. Different doses of ligand treatment are administered by columns, usually, with the control no EGF treatments being in the first one or two leftmost wells, with increasing EGF doses in the wells to the right from the controls. In each experiment run on a given day, we aim to have at least two repeats of same the experimental condition.

After the staining with primary and secondary antibodies, wells are imaged with an automated microscopic device. The images are then processed by an image analysis software package to quantify data from the experimental results. The algorithm detects individual cells in the image of each well, computes average fluorescence in the area corresponding to each cell, and for each well a distribution of values of single cell fluorescence is reported (see Section 4.1.4). To distinguish different cells, cell nuclei are stained with Hoechst dyes and are then detected based on the emitted Hoechst blue fluorescence.

The image analysis algorithm uses empirically derived thresholds for nuclei detection levels, to count only viable cells and avoid counting cell debris. Upon nuclei detection, the cell area is computed by expanding cell boundaries from the nuclei until bounds of neighboring cells are reached. The total fluorescence from this cell area is then measured and normalized by the detected cell area. Thus, the average fluorescence per cell area is reported to account for variations in cell sizes. In addition, the algorithm allows separate estimation of the fluorescence in the cell plasma membrane

bound region, the cytoplasm or the nucleus; however, in this work we use only the total cell fluorescence averages due to their higher reliability. The coordinates of the box area detected of each cell are also reported by the algorithm. The quantified fluorescence readout is reported in arbitrary fluorescence units (A.U.).

The image analysis algorithm also implements a quality control check to avoid counting fluorescence from overlapping cells multiple times. However, occasionally, due to possible optical anomalies of the microscopy or error in cell boundary detection due to nuclei staining, average fluorescence from a region with multiple neighboring cells might be reported as read-out for one cell. This generates very high values of fluorescence read-outs in the single cell distributions. Thus, discarding the top 3% of highest outlying read-outs in the distribution corrects for this effect. Overall, single cell distributions after such corrections have smaller variances and appear consistent from day to day.
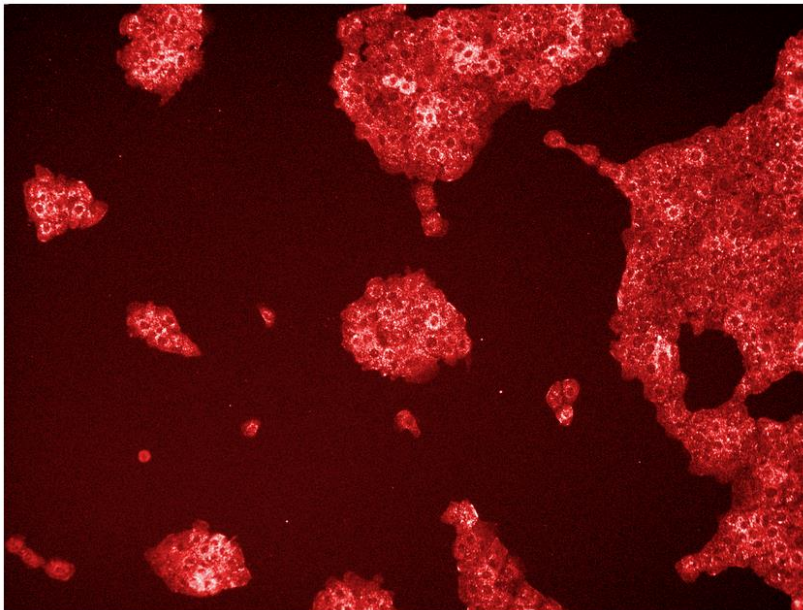
Figure 4.1: Microscopy imaging of MCF-10A cells treated with high dose EGF and phospho-EGFR detecting antibodies
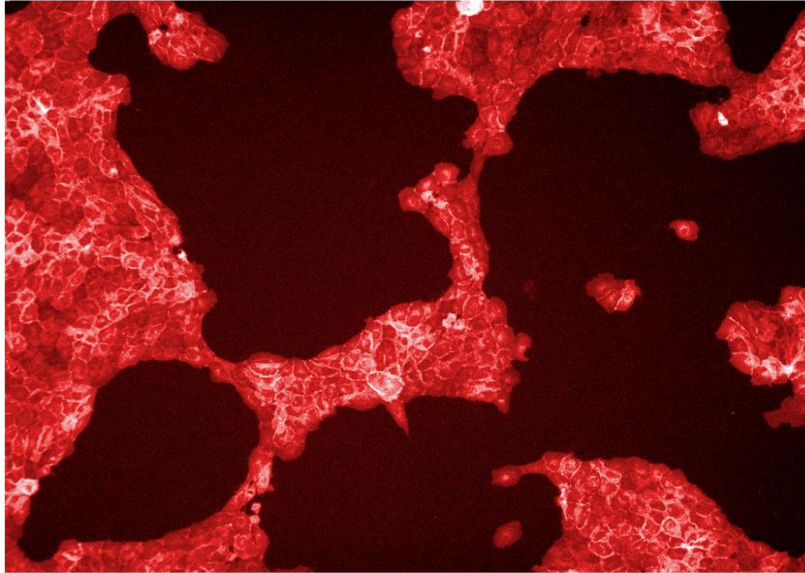
Figure 4.2: Microscopy imaging of MCF-10A cells treated with high dose EGF and total EGFR detecting antibodies.

Section 4.1.3: Background fluorescence offset subtraction in microscopy imaging

Despite technological advances in biology it is still challenging to quantify dynamic time-courses of protein abundances in mammalian cells (Teeffelen, Shaevitz, & Gitai, 2012). Measurements of *in vivo* fluorescence in the cells can be obscured by various levels of noise and artifact. There are still no reliable methods to convert fluorescence reported in arbitrary units into actual numbers of detected target molecules, although some techniques have been developed for bacterial cells (Rosenfeld, Perkins, Alon, Elowitz, & Swain, 2006). In our work we first pre-process reported fluorescence imaging data by correcting it for background fluorescence offset. Next, in our simulations we fit a parameter that corresponds to the fluorescence scaling factor and converts the experimental data into the number of molecules simulated in the model.

Each single cell fluorescence readout encompasses fluorescence actually emitted by the detected target, and also fluorescence emitted by non-specific binding of primary and secondary antibodies and device-related fluorescence background. To correct for the fluorescence offset not related to the detected target, background fluorescence needs to be quantified and subtracted from each single cell measurement. Many studies report fluorescence data with fluorescence background subtracted automatically by the image analysis algorithm (Shankaran et al., 2012). Often, such

77

automatic background subtraction can be unreliable, and hence, more carefully developed techniques should be used instead.

We apply a customized background fluorescence subtraction algorithm. First, we measure the time=0 fluorescence in wells with the cell population treated only with secondary antibodies, omitting any primary antibodies and EGF. Since there are no primary antibodies in such wells, the fluorescence readout should reflect only the compounded fluorescence due to the use of the microscopic device and the fluorescence offset due to secondary antibodies off-target binding. We refer to the measured quantity as "no primary" control or "background fluorescence offset". Since the primary antibodies are specific to each target, pEGFR, pERK, pAKT or tEGFR, but the detecting secondary antibodies are the same for all targets, the background fluorescence offset is the same for measurements of all targets across the experiments run in the same cell line on the same day. We observe that the background fluorescence can vary from day to day, and might depend on the condition of the measuring device (e.g. the last date when the device was serviced).

Even in absence of ligand, some small fraction of target protein can be phosphorylated, a phenomenon referred to as basal phosphorylation. We need to quantify the abundance of phosphorylated target induced specifically due to ligand treatment, thus excluding any fluorescence due to background offset or due to basal phosphorylation. Therefore, for each measured condition we also have a 0 EGF control that entails basal phosphorylation of the target at each given time point in the absence of EGF treatment as well as off-target binding of the primary antibodies, and the background offset. This no EGF control is then used to normalize and scale read-outs from the non-zero EGF wells (see Section 4.1.4).

We find that, in MCF-10A, the no primary and the 0 EGF controls are very similar on a given day, with the 0 EGF controls being only slightly (about 5%) higher than the no primary controls. This suggests that in MCF-10A the fluorescence due to the basally phosphorylated target is low compared to the background fluorescence offset. Hence, in case no primary control is missing in MCF-10A data, it is justifiable to use the 0 EGF controls instead to account for both device-related fluorescence offset and primary, secondary antibodies off-target binding in measurements of phosphorylated targets.

For total EGFR measurements, the no EGF control readout of total EGFR at the time=0 represents fluorescence due to total number of EGFR detected in the cells, off-target binding of primary and secondary antibodies, as well as the device-related fluorescence background. After we subtract the background fluorescence offset from the no EGF measurement of total EGFR at time 0, in the remaining fluorescence we cannot quantify relative contributions to the total fluorescence due to off-target binding of primary antibodies versus fluorescence due to the actual measurements of total EGFR. We can assume that the fluorescence due to off-target binding of primary antibodies to tEGFR is negligible in the fluorescence remaining after no primary control subtraction, and thus the remaining fluorescence corresponds mostly to the total EGFR abundance.

However, to get a more precise assessment of total EGFR in the cells, a more appropriate control to correct for off-target binding of primary antibodies for tEGFR would be a no EGF measurements of tEGFR in MCF-10A cells expressing no EGFR. If all other treatment and handling conditions are the same, the MCF-10A cells expressing no EGFR should be identical to the normal MCF-10A levels of off-target binding of primary, secondary antibodies and the same device-related fluorescence offset. However, generating such MCF-10A cells with knock-out EGFR is outside the scope of this study, and so, as an alternative, one might use the control no EGF tEGFR measurement from cells expressing very low levels of EGFR, for example, the MCF-7 cell line. In this case, the no EGF readout would encompass device-related offset, and fluorescence due to off-target binding of primary and secondary antibodies, and low levels of fluorescence due to detected EGFR. Such control measured in MCF-7 due to differences in gene expression between MCF-7 and MCF-10A, might have different levels of off-target binding of primary and secondary antibodies. However, in our case tEGFR readouts from MCF-7 cell lines are very similar to the no primary control measurements in MCF-10A, and thus might serve as an approximate control for tEGFR measurements in MCF-10A.

To compute the actual value of the fluorescence background control we use all measurements from single cell distributions from wells containing only cells treated with secondary antibodies added (no primary antibodies) and no EGF added at time=0 and all measured on the same day. We combine all single cell distributions from such wells, and report the average of the resulting distribution as the background fluorescence offset for that specific day and cell line. The fluorescence

background offset is then subtracted from each single cell value in the entire dataset for that cell line on the specific day.

Section 4.1.4: Single cell distributions data

We collect single cell measurements of pEGFR, pERK, pAKT and tEGFR response to EGF treatments and their no EGF controls in MCF-10A cells. Data reported from each well is the distribution of single cell responses of a given target at a given time point after treatment with specific EGF dose or the no EGF treatment control. We observe that the obtained single cell distributions are mostly bell-shaped.

To correct single cell distributions for day-specific instrumental offset, from each value in each single cell distribution we subtract background fluorescence, estimated as described in Section 4.1.3. To reiterate, the background fluorescence offset captures the effect of device-related fluorescence and fluorescence due to off-target binding of secondary antibodies, and is the same for all measured targets on a given day. After the background offset subtraction, to each single cell distribution we apply scaling and normalization protocols according to the type of the measurement in that distribution - phosphorylation or total EGFR. We present detailed descriptions of each of these protocols throughout Section 4.1.

The no EGF controls for a given phosphorylation target, after background offset subtraction, produce an estimate of the compounded fluorescence due to off-target binding of the primary antibodies and the basal phosphorylation of the target. The no EGF control wells measuring a given target and run on the same day have no perturbation applied (i.e. no EGF treatment), and thus are expected to have lower variation between single cell responses distributions across the wells. We find that for most days, the no EGF distributions within same day have similar bell shapes and similar variances, although they can have somewhat different means. Even the two no EGF controls on the same row can have slightly different means, most likely due to variations in experimental handling, such as different number of cells in wells.

We check the quality of the no EGF control distributions prior to using no EGF control single cell distributions to normalize the data. Specifically, we overlay the z-transformed no EGF control distributions from the same day, measuring the same target, and identify the distributions with shape

or variance significantly different from the majority of the distributions. Rows, in which no EGF controls distributions are abnormal, might be discarded from the data set, as this abnormality indicates that possible errors in handling the wells most likely occurred during treatment or imaging steps. In addition, identifying abnormal control distributions can help to detect possible inconsistencies in cell handling and to improve the experimental techniques thus yielding a better quality data.

Next, for each individual single cell distribution from a non-zero EGF treatment well, we combine the corresponding to it no EGF control single cell distributions and use the average and the standard deviation of the combined distribution to z-transform the single cell distributions for each individual non-zero EGF treatment. Therefore, by subtracting the average of the corresponding no EGF control distribution from the non-zero EGF responses, we obtain fluorescence value corrected for off-target binding of primary antibodies and the background phosphorylation, and hence, corresponding specifically to only the target response to the applied EGF treatment. Such transformation allows us to account for row-to-row variations between the wells and to obtain single cell distribution in target phosphorylation specifically elicited due to EGF treatment. To summarize, the obtained single cell distributions of phosphorylation responses are first corrected for the background fluorescence offset, and then z-transformed to their corresponding no EGF control distributions.

For total EGFR measurements, unlike for the phosphorylation measurements, we cannot separate fluorescence due to actual abundances of total EGFR from the fluorescence due to off-target binding of the primary antibodies. However we can assume the latter one is negligible compared to the former one. For a given day, we combine all single cell distributions measuring tEGFR at time=0, no EGF added and after the background offset has been subtracted from them. We then set the average of this distribution to represent fluorescence corresponding to 100% of total initial abundance of EGFR in the cells on that day. We then use this value to scale each single cell measurement of tEGFR on that particular day. To normalize the obtained tEGFR single cell distributions to the variance of the combined 0 EGF, time=0 single cell distribution of tEGFR, we divide each single cell value of tEGFR by the standard deviation of the combined no EGF, time=0

tEGFR distribution. Hence, we obtain tEGFR measurements expressed as fractions of initial tEGFR abundance on that day, and standardized to correct for the well-to-well variations.

We then can examine the properties of single cell distributions across multiple replicates all performed on the same day. The observed variability in the replicates done on the same day is most likely due to experimental handling and instrumental errors, as the underlying biological fluctuations in each repeat should be minimal.

Section 4.1.5: Combining the normalized experimental data obtained on different days

The goal of the experimental part of our work is to obtain a representative average behavior of the EGFR signaling system, which is independent of instrumental handling. Whereas multiple repeats of the same experiment conducted on one day allow to account for instrumental errors, repeating the same experiment on different days allows to capture true biological variability in the system behavior. To filter out instrumental errors from true biological behavior of the system, we need to be able to integrate and analyze together multiple independent replicates of the same condition measured on one day, as well as to combine them with replicates measured on different days. For this we assemble an experimental data set derived from multiple replicates of experiments conducted on different days, with every replicate containing several repeats of the same-day measurements of the condition. In this section we describe our methods to process repeats from different days to correct for their day-specific variations and render the data comparable between the days.

Analysis of single cell distributions for a given condition and a given target across different days allows to uncover the true biological variability separated from instrumental noise. We find that single cell distributions for most responses are consistently bell-shaped across the days. The means of the distributions might vary from day to day, even after correction for the background fluorescence, most likely due to accidental cell-count variations in the wells across the experiments (see Section 4.5). To be able to directly compare single cell distributions for the same condition, measuring the same target on different days, we need to scale each distribution for the day-to-day variability. We present such scaling methodology in the next section. Throughout this section, by response we mean

single cell distribution obtained from a well measuring a specific target at a specific condition, which is defined by EGF dose applied and time when the response is measured.

After correcting the data for background fluorescence on different days, reported fluorescence responses to the same conditions for a given target, in principle, should be directly comparable across the days. However, we observe that the scales of fluorescence readouts for the same phosphorylation target can vary significantly between the days. For example, we find that whereas for pEGFR and for pAKT, the values of their corresponding maximal signals usually are similar between the days, for pERK the maximal fluorescence readout can vary several folds on different days. One possible explanation is that primary or secondary antibodies used to detect the same target on different days can have some batch-specific variations, and thus emit somewhat different fluorescence signal in response to equal amount of the detected target. Hence, we perform an additional step of scaling the entire data for a given phosphorylation target on a given day and target-specific scaling factor described in detail in Section 4.1.7.

For tEGFR measurements since our within-the-day normalization procedure converts tEGFR data in arbitrary fluorescence units (A.U.) into the fractions of total initial EGFR present, no further scaling is needed to render the data comparable between different days.

Section 4.1.6: Working with the averages versus the single cell distributions data

After the responses for each target are normalized and scaled in each individual day-specific repeat, the responses for the same condition can be directly compared and combined together to be fitted in the model. This data set represents the behavior of the biological system separated from the instrumental noise and accounting for the day-to-day variation. In this section we present two approaches to combine the processed data for the use in model fitting: the ensemble averages approach and the combined single cell distributions approach.

There are two fundamentally different approaches to combine the experimental measurements from our experimental system. One approach combines responses from single cell distributions; second approach deals with averages from the individual single cell distributions. In the first approach, processed individual single cell distributions, replicates from both within and across the

83

different days, are all combined together in one distribution. If the resulting distribution is bell-shaped, its average and standard deviation can be used to fit the model. Alternatively, if the resulting combined distribution substantially varies from the normal distribution, one might attempt to fit the model to the combined distribution itself, and thus to better capture the biological information about the single cell population encoded in such distribution. Fitting the model to the combined single cell distribution would then allow us to simulate a distribution of individual single cell responses in cell populations. For example, such approach can be useful to study variability in single cell responses in a given cell population.

Some studies may aim to understand a collective cell response, or the average most probable response of cells in cell population. Such approach may be applicable, for example, in studies of cells behavior in tissue context, in which case, it would be more appropriate to fit the model to the averages of ensemble responses rather than to the combined single cell distribution. Hence, the context and goals of the study would determine which of the two data interpretation and combining approaches should be used.

In practice, fitting the data processed by either of the two presented approached may require slightly different MCMC sampling techniques. For example, ensemble averages data can have smaller variances, thus significantly constraining the model parameter space, and hence restricting MCMC chains movements across it. On the contrary, single cell distribution data is likely to have higher variances and thus could help MCMC chains move faster across the model's parameter space, sampling it more efficiently.

In this work we present the results of fitting the model to the combined ensemble averages of the individual distributions of responses. In Section 5.4 we describe how the model can be fitted to the combined single cell distributions of responses. We can analyze and compare parameter distributions obtained from each of the two approaches using similar methods.

Section 4.1.7: Normalizing and scaling phosphorylation data: the averages of single cell distributions

We measure pEGFR, pERK, pAKT responses to continuous treatment with different doses of EGF (0 -100ng/ml) within 0 to 3 hours of the treatment and replicate the experiment on several

different days. In this section we first present a method of scaling and combining the averages of the single cell distributions, and then describe a method of working directly with the single cell distributions. For both of the cases the first step of data processing is the same: for each target on a given day the measurements from each well are subjected to quality check based on the variances of their distributions, background fluorescence subtraction and z-transformed to the corresponding no-EGF control distribution within a given day for a given target.

Then, we take the averages of the single cell distributions from each replicate of same condition for a given target on a given day. To control for the day-specific effects, we need to rescale all same-day replicates to a day-specific scaling factor. For instance, the data can be scaled by an average fluorescence of the dose response of this data at the time when the signal reaches its peak; or the fluorescence at the saturating response of this target on a given day. Alternatively, more nuanced quantile-based techniques can be applied, although, it is not clear if such elaborate techniques, that may also require corrections for differences in distributions of single cell responses, would produce qualitatively very different data.

In some cases, for each individual target we often observe that the averages of single cell distributions in responses to several highest doses are very similar, reaching the maximum average of single cell distributions observed on that day. This peak represents the saturation response of the target on a given day. Since the correspondence between the absolute levels of phosphorylated proteins and the arbitrary unit of fluorescence of the measurements can change from day to day, relating fluorescence scales on different days by comparing the saturating fluorescent values may be more robust to instrumental errors. In a way, fluorescence readouts of a saturating response represent a more distinctive and specific biological feature (marker) of the cell population, which should be easier to capture in a day-to-day variability rather than the more variable between the days average of dose response at the time when maximum is reached.

However, if the observed responses to higher EGF do not appear to reach saturation, we cannot rely on the scaling by the maximal value. Instead, we can scale the data using average value of a dose response at time when maximum response is detected. Specifically, we can consider dose response curve at 10 minutes, time point at which in most detected responses highest amplitude is

reached. This approach could be more noise-proof as it relies on an average value from multiple measured repeats, as opposed to just one measurement of the maximal point.

Finally, the entire data for a given target is divided by the scaling factor of our choice – an average of the dose response or the saturating response. Then, we report the average and the standard deviation of the set of all reported means for single cell distributions for the same target and the same condition, combined across multiple days. Hence, the resulting data then represents ensemble averages of cell populations.

Next, we describe a method to combine single cell distributions across the replicates from different days. We use the same day and target-specific scaling factor as derived for the ensemble averaging to correct individual data points in single cell distributions. We then combine together all single cell distributions of a given target and condition measured across the days. The average and the standard deviation of the obtained combined distribution is then reported for each condition. If the obtained distribution appears normal, the model can be fitted to the points represented by such average and standard deviation. Alternatively, the model can be fitted to the entire single cell distribution corresponding to a specific target and time point.

Since our data processing algorithm involves division by values which themselves have associated with them errors, we apply the error propagation formula to correctly compute the standard deviation in the reported data points.

Section 4.1.8: Normalizing and scaling the total EGFR data: ensemble averages vs single cell distributions

Traditionally, changes in EGFR abundances over time are measured through quantifying the ability of cells to bind radio-labeled EGFR ligand (van de Poll et al., 2005). Such measurements do not directly measure receptor mass, but infer it indirectly through fraction of unbound ligand, remaining in the media of cells incubated with some dose of ligand over time. Hence, such techniques can incorporate errors due to various factors, for example, often neglected heterogeneity in ligand-binding properties of receptor monomers and dimers. Directly measuring the total EGFR abundances in cells over time with ELISA assays (Lequin, 2005; Weemen & Schuurs, 1971) has been challenging due to the absence of agreed upon standardized procedures of normalizing and

processing the tEGFR readouts: until recently there were no homogenous and reliable algorithm to correct for the background fluorescence offset and the off-target binding effects in measuring total EGFR abundance.

We use ELISA to experimentally measure changes in the total EGFR abundance in response to continuous treatment with different EGF doses over three hours of EGF treatment. We develop and apply a proper, in our view, background fluorescence subtraction algorithm for the tEGFR measurements.

Our experimental measurements of the total EGFR expression include levels of functional receptors present on the cell surface, as well as in other intracellular compartments. Specifically, our tEGFR measurements can include newly synthesized receptors en-route to the cell surface, or, receptors just internalized through endocytosis, or receptors in lysosomes targeted for degradation. We assume that most of the measured EGFR are actively participating in signaling, and that there are no reservoirs of receptors inaccessible to signaling. Some of our experimental observations suggest that in the absence of ligand stimulation, most of EGFR receptors reside near the cell surface: the levels of plasma-membrane-localized EGFR receptors are very similar to the total EGFR levels measured in MCF-10A cells. Therefore, we can assume that the tEGFR fluorescence measurement reported in the no EGF time=0 readout, after correction for the background fluorescence corresponds to the total number of receptors actively participating in cell signaling.

Converting the tEGFR fluorescence readouts into the absolute number of EGFR receptors present in the cells is experimentally demanding for several reasons. First, the scale of fluorescence, determining how many receptors are detected in a unit of reported fluorescence, can change from day to day. Second, the abundance of EGFR expressed in cells can not only vary between the days, but also, can vary within the same day in cells in different wells. Hence, to avoid necessity to quantify the absolute number of EGFR on every day the experiment run, on each day, we convert the fluorescence measurements of tEGFR into fractions of the initial tEGFR abundances. We then combine such fraction data between different repeats and days. We thus fit our model to the tEGFR data represented in fractions of initial abundance of EGFR receptors.

Our previous ELISA measurements show that MCF-10A cells express between 3x 10^5 and 6x10^5 EGFR molecules per cell. We expect that the average EGFR expression in cell populations with no EGF stimulation at time=0 in different wells on a given day to be similar. Plating cell population of several thousand cells in each well, in theory, should cancel out most of the variability due to stochastic fluctuations in cell populations, such as possible differences in cell cycle stages and gene expression.

However, the average EGFR expression in cell populations is expected to vary from day to day, reflecting slight variations in experimental handling of cells which can occur between different days. For examples, variations in cell densities in the wells plated on different days, the age of the media in which the cells are pre-incubated, and other conditions may contribute to the variation in the average EGFR expression in the cells on different days. Interestingly, incidental exposure of a batch of cells to residual doses of EGF can also affect the number of EGFR receptors expressed on the cell surface.

To summarize, after the background fluorescence subtraction and normalization, we represent changes in EGFR levels over time in fractions of initial amount of receptors present in the cells. Such approach then allows us to combine the total EGFR time-courses measured on different days and bypass the burden of experimentally measuring the average number of EGFR molecules in cells on each day the experiment is run. Repeating the same experiment on different days and combining the tEGFR data processed to represent the fraction of initial tEGFR allows to address biological variability in EGFR expression in cells between various days.

In our model simulations we incorporate the measured abundances of EGFR receptors in MCF-10A cells as a prior constrain on the range of values of parameter, which represents initial condition abundance of EGFR receptors in a cell. Then we probabilistically sample the values of the initial abundance of EGFR by fitting the phosphorylation and the tEGFR data. We thus obtain posterior distribution of the initial EGFR abundance which is derived based on the fitted experimental data.

Since we rely on assumptions that at the beginning of the experiment cells are in a quiescent state, at time 0, in the absence of EGF, tEGFR levels should be at equilibrium. Equilibrium in total

EGFR levels implies that the rate of EGFR production is balanced by the rate of the basal internalization, degradation, and recycling of receptors. For each given day, the tEGFR data is then scaled by the average of the combined single cell distribution of tEGFR at no EGF and at time 0, so that at initial time point the number of receptors corresponds to 100% of total EGFR actively participating in cell signaling.

We then divide the rest of the tEGFR data measured on that day by this value. After the data from each individual day have been processed and represented as fractions of total EGFR receptors, repeats from different days can be combined together. For each condition we report the average and the standard deviation of all the averages of single cell distributions from multiple wells.

To combine single cell distributions of tEGFR measurements, first, we use the day-specific scaling factor derived as in the above procedure of scaling and normalization of the averages to divide each individual point in the single cell distributions of tEGFR on that day. The single cell distributions from each condition then represent the fraction of total EGFR. In addition, within each day, we divide all points in the single cell distributions by the variance of the distribution obtained from combining the corresponding controls of no EGF time=0 tEGFR distributions. For each condition, we then combine together all single cell distributions across repeats from the same day as well as from different days. If the resulting distribution appears normal, the model can be fitted to the reported averages and standard deviation for each of the resulting distributions for a specified condition; alternatively, the model can be fitted to the very distribution at each condition.

Standard errors of the reported data are corrected by taking into account the division by scaling factors, which themselves have standard errors associated with them.

Section 4.2: Various types of experimental data constrain the model parameter space: exploring relationship between the structure of the model parameter space and system behaviors

To explore the structure-to-function relationship between the model's parameter space and model's responses to EGF stimulation we fit the model to different types of experimental data measured specifically in our biological system - MCF-10A cells. It has been previously suggested that to better constrain parameter space of a computational model it may be a more optimal strategy to

collect various complementary types of experimental data, rather than to collect multiple repeats of the same type of measurements (Apgar, Toettcher, Endy, White, & Tidor, 2008). In our approach, we first conduct experiments measuring the phosphorylation and the EGFR receptor abundance time-courses in response to treatments with various constant doses of EGF. We collect several independent replicates of the same experiment conducted on different days to ensure that the obtained data is reproducible and that it captures the true biological variability in the studied system (see Section 4.1). In addition, we conduct several other experiments of different types to further dissect the ErbB signaling system from multiple angles. In particular, we measure the EGFR signaling responses to constant-dose EGF stimulations with addition of other perturbations, such as, for example, EGF washout from the media, or addition of various kinase inhibitors. Fitting the model to such additional data allows to further constrain the model's parameters space and thus leads to a better understanding of the underlying molecular mechanisms realized in the ErbB signaling system.

Kinetic modeling of signaling responses to different EGF doses has been previously applied to understand the MAPK signaling cascade (Schoeberl et al., 2002). However, previous approaches relied on deriving their conclusions based on simulating various perturbations to the system, such as, for example, altering protein overexpression or rates of interactions, in a limited number of parameter assignments. Our approach allows to account for broad variations in the underlying model's parameter space and to derive a mechanistic insight about EGF signaling based on the distributions of parameters, rather than based on single point estimates of the parameter values.

This section describes our analysis on identifying model parameters which are sensitive to various additional types of experimental data. The experiments measure pEGFR, pERK, pAKT, and total EGFR in response to perturbations either by saturating doses of kinase inhibitors, or by ligand washout from the media of EGF pre-treated MCF-10A cells. Obtaining such complementary to the main phosphorylation responses data leads to a better understanding of the mechanisms shaping the ErbB signaling dynamics that are most likely to be actually realized in this biological system. Our model shows that same system behavior (i.e. phosphorylation signal decay) can be fitted through different combinations of available mechanisms (i.e. degradation of receptors or action of phosphatases). Additional perturbations experiments can help assess separate contributions from

individual mechanisms to shaping system dynamics. The main principle in the analysis presented below is that constraining model responses to additional experimental data should lead to changes in the model parameter space to accommodate the additional data points. Therefore, by comparing predicted distributions of parameters obtained from fitting the model to the fundamental experimental data set (the phosphorylation time-courses in response to constant-dose EGF treatment over several hours) to the distributions of parameters obtained from fitting the model to the same data set and additional data points, we should be able to identify parameters that have significantly shifted in their distributions to accommodate the extra data points, and thus are potentially important to the additional experimental conditions.

By fitting our model to the comprehensive data set of the phosphorylation time-courses we derive posterior distributions of parameters. The obtained distributions reflect constraints imposed on the model parameter space by experimental data representing most essential aspect of the ErbB signaling system – phosphorylation responses. However, as we show throughout Section 4.2, in addition to the phosphorylation time-courses we fit our model to other types of experimental data. Thus, the posterior distributions of parameters obtained by fitting only phosphorylation data serve as the fundamental distributions of parameters, to which we then compare the distributions obtained from the simulations with addition of other types of experimental data.

In the next sections, we describe how fitting the model to various additional experimental data affects posterior parameters distributions. More specifically, we first derive posterior parameter distributions to fit only the phosphorylation time-courses at different EGF doses (as in Section 4.3.1). We then refit the model to the phosphorylation time-courses and the additional experimental data and obtain a set of more constrained posterior parameter distributions. We apply statistical methods of comparing parameters distributions described in Section 3.3 and report parameters that were most shifted in their distribution by the additional experimental data. Note that the shift suggests the importance of these parameters and their sensitivity to the specific additional experimental data.

Section 4.2.1: The phosphorylation signaling time-courses of phospho-EGFR, phospho-ERK, and phospho-AKT in response to treatments with multiple constant EGF doses

Our main experimental data consists of the phosphorylation time-courses of pEGFR and its downstream targets, pERK and pAKT, in response to stimulations with constant EGF doses up to three hours (Figure 4.3). The data shown was obtained by combining and normalizing independent biological replicates of the experiments conducted on three different days (see Section 4.1). The time-courses span up to three hours after EGF addition and the earlier more transient dynamics is being measured at more frequent time intervals.
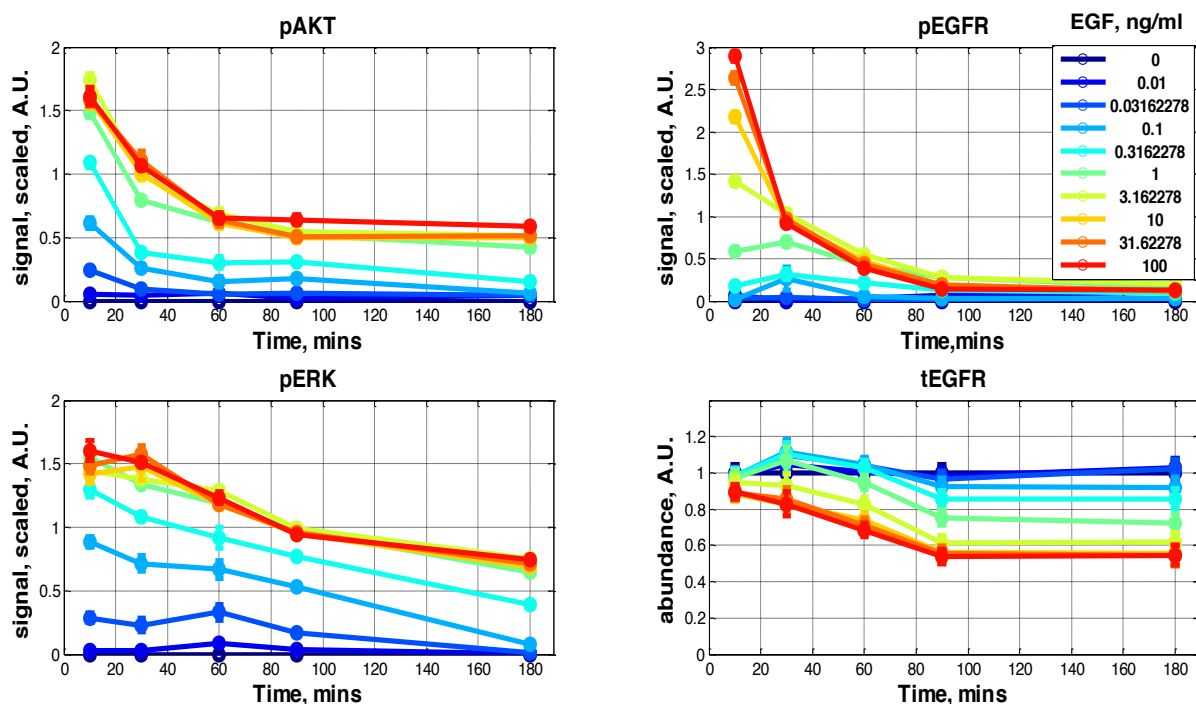
Figure 4.3: The experimental data: the three-hour phosphorylation and the total EGFR time-courses collected from MCF-10A cells treated with different EGF doses (0-100ng/ml).Each point represents a triplicate average and its corresponding standard error. Phosphorylation signal at time 0 is assumed to be 0

The phosphorylation time-courses data captures various important features of the ErbB signaling dynamics. For example, we observe differences in the signal decay rates on pEGFR and on its downstream targets: pEGFR at highest EGF dose decays to its basal level within 1.5 hours, whereas pERK and pAKT decay slower. Lower doses of EGF elicit pEGFR responses of a more sustained dynamics, whereas higher EGF doses produce transient pEGFR responses with well-defined pulse-like shapes. The dose-responses of pEGFR and pERK at early time (10mins)

92

demonstrate the effect of ultrasensitivity (Figure 4.4), a well-studied feature present in the MAPK cascade (Goldbeterl & Koshland, 1984). Ultrasensitivity describes a condition when signal at pEGFR is amplified by the kinase cascade leading to the downstream pERK responses, leading to pERK signal saturation even prior to the saturation at the pEGFR responses. Hence, the collected phosphorylation time-courses data at different EGF doses captures such key properties as the dose-dependent signaling dynamics (transient or sustained), differential sensitivity to EGF doses of each target (threshold of pERK activation is at lower EGF doses then that of pEGFR), as well as the non-trivial relationship between activation of pEGFR and its downstream targets (relative shifts in pEGFR and pERK dose responses).
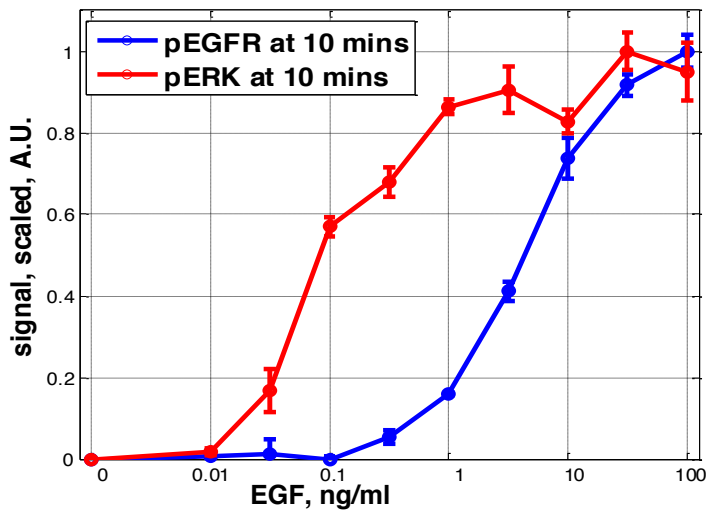
Figure 4.4: ERK ultrasensitivity: dose responses of pEGFR and pERK measured at 10 minutes after EGF treatment in MCF-10A cells.

Figure 4.5 shows a model-produced fit to the experimental data, in which the model was fitted to the phosphorylation responses of pEGFR and pERK and to the time-courses of the total EGFR abundances up to three hours of stimulation with EGF doses ranging from 0 to 100ng/ml EGF. The produced fit agrees reasonably well with the experimental data. However, due to a relatively large number of fitted conditions (10 EGF doses), number of targets (three), and extended time of observation (3 hours) fitted in this simulation, some experimental data points are inevitably worse fitted then others (see tEGFR at highest doses). In addition, our assumptions used for the normalization and scaling of the experimental data could have led to data with certain intrinsic inconsistencies and thus contributed to the fact that some points are not being fitted well by the

model. For example, our assumptions regarding the amount of the background fluorescence offset subtraction in total EGFR measurements could result in under-estimating the fraction of EGFR receptors degraded by three hours of EGF stimulation (Section 4.1.3): by subtracting the lower background fluorescence values from the tEGFR experimental data, we produce normalized and scaled tEGFR measurements with lower fraction of receptors degraded by the end of three hours.
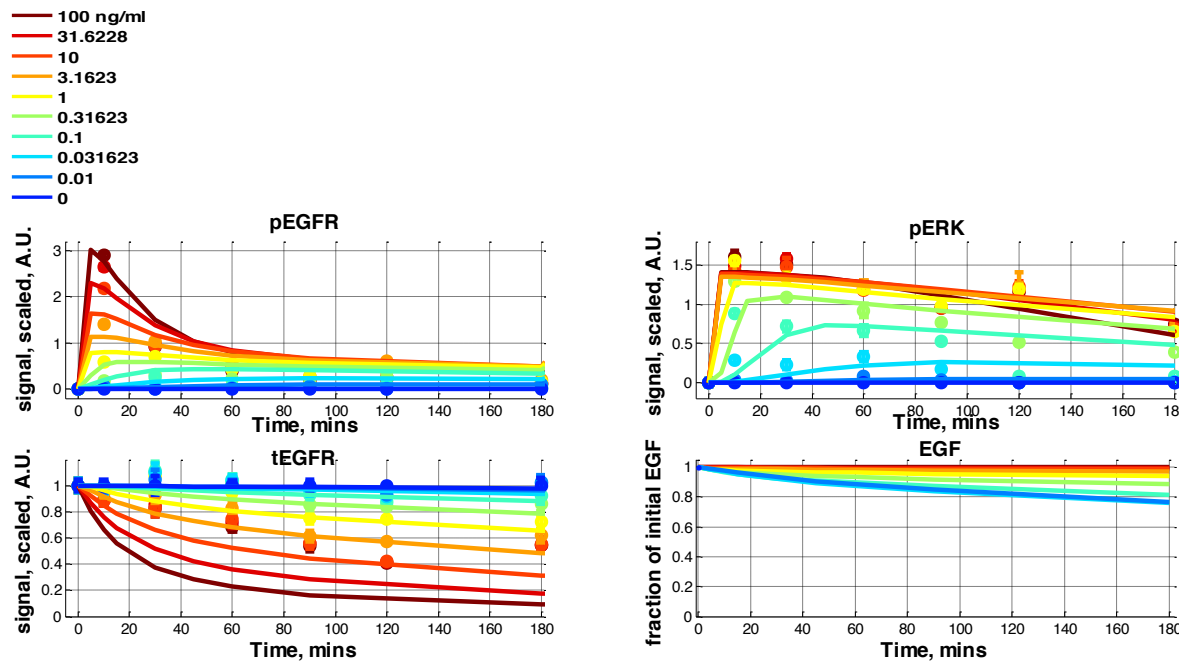
Figure 4.5: Model fits to the experimental data (phospho-EGFR, phospho-ERK and tEGFR time-courses) found by simulated annealing. EGF fraction data were not fitted. Solid lines represent one fit to the data. Points represent the experimental data points with their corresponding standard errors.

Section 4.2.2: The total EGFR degradation time-courses at multiple EGF doses

The results presented in this subsection were obtained by fitting the extended ErbB signaling model previously published in (Chen et al., 2009). Due to the size (159 non-zero initial condition parameters) and complexity of the model in described this section, we fit the model only to the experimentally measured responses to high EGF dose (100ng/ml) for up to 1.5 hours of stimulation (Figure 4.6). Among the key distinctions of the extended model over the simplified model is the implementation of two types of EGFR internalization, namely, internalization through coated pit proteins (cPP) and non-cPP dependent internalization. In addition, the extended model implements

separately species of cytoplasmic and endosomal RTK phosphatases of EGFR and represents their activity as bi-molecular reactions with their own separate rates. Lastly, the extended model encompasses a generally accepted view that phosphorylated receptors are preferentially destined for degradation in endosomes, whereas unphosphorylated receptors can be recycled to the cell surface (Sigismund et al., 2005).

To experimentally assess degradation of EGFR in our system, we quantify the changes in total abundances of EGFR receptors over several hours in response to stimulations with different constant doses of EGF. These experiments show that a higher EGF dose elicits more receptors degradation by 1.5 or 3 hours of EGF stimulation. In addition, the data shows that initially, at all EGF doses, the number of EGFR receptors does not change significantly, reflecting the established fact that prior to degradation receptors need to be internalized and shuttled into the lysosomes (Sigismund et al., 2005), a process requiring time on the order of minutes. As described in Section 4.1, we fit receptors abundances data using fractions of the initial number of receptors remaining in the course of cells treated with different doses of EGF over time. Although we are working with fractions, rather than with the absolute number of receptors, we simulate initial abundances of EGFR in the cells as a parameter.

Fitting receptors degradation data in addition to the phosphorylation time-courses imposes a constraint on model not only to fit phosphorylation signal decay, but also to fit this decay with a specified degree of receptors degradation. Therefore, we anticipate that such constraint could limit certain mechanisms and dynamic processes in the model, not necessarily directly related to degradation of active receptors.

We observe that fitting our model exclusively to the phosphorylation time-courses produces fits with a wide range of predicted fraction of receptors degraded in the long run of stimulation with EGF (Figure 4.6, fits in blue). We then plot the predicted distribution of fraction of EGFR remaining by 1.5 hours of 100ng/ml EGF stimulation (Figure 4.7A, distribution in blue). The distribution suggests that fitting phosphorylation data can be realized with various degree of receptors degradation. However, in the majority of fits (60%) phosphorylation signal is fitted with degradation of almost all EGFR receptors by 1.5 hours of 100ng/ml EGF stimulation. The model also predicts that in about

20% of the fits it is possible to fit phosphorylation signal with almost no degradation of EGFR receptors (>90% receptors remaining). Experimentally we have detected that in MCF-10A cells treated with 100ng/ml EGF by 1.5 hours approximately 55% of the initial number of EGFR receptors would remain undegraded in the system. Therefore, our model predicts a very different behavior from the one that is realized in the cells. We then re-simulate the model by imposing the experimentally observed EGFR degradation time-courses (Figure 4.6, fits in red). We see that the model can still fit well the imposed experimental phosphorylation data, although the predicted average time-courses with the imposed degradation time-courses show a slightly different amplitude of the pEGFR peak and early dynamics on all phosphorylation targets (5 minute time point) (Figure 4.6). Imposing the experimentally measured EGFR degradation time-courses results in significantly constrained distributions of the remaining tEGFR by 1.5 hours, with smaller variance (Figure 4.6, 1.5 hours time point and Figure 4.7A, dark red).
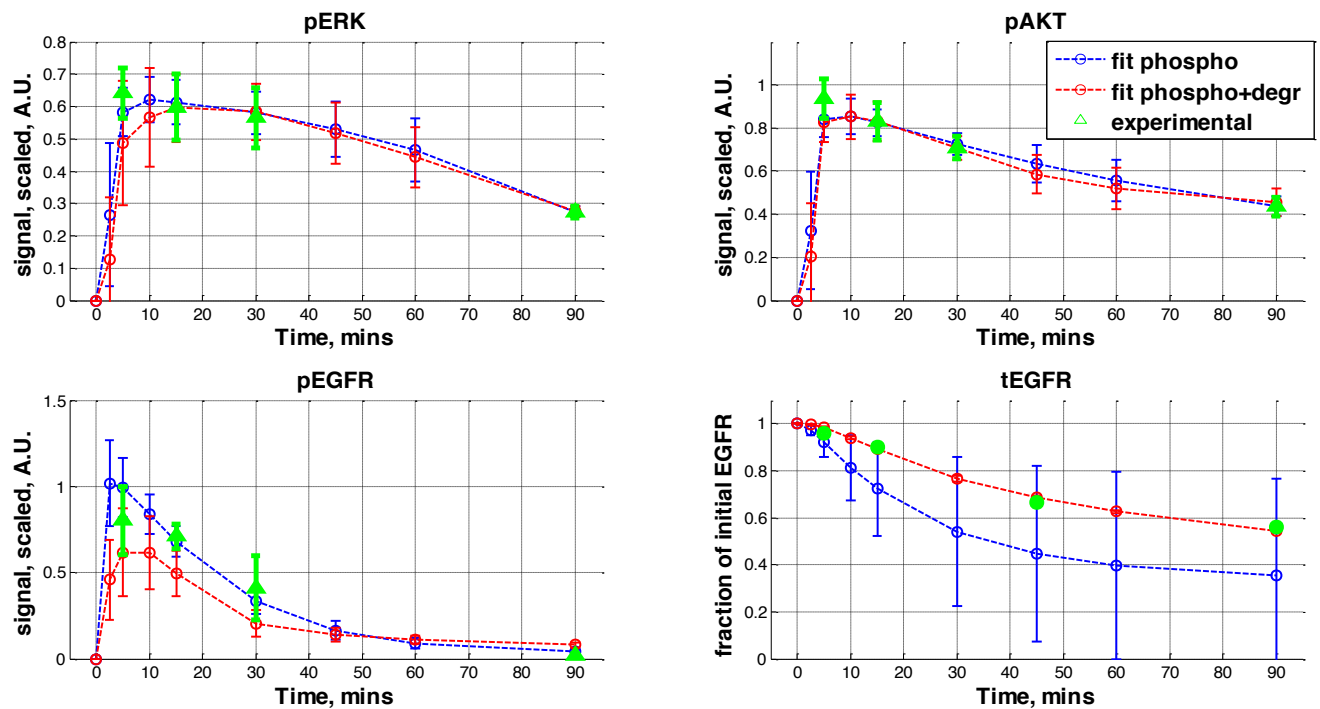
Figure 4.6: Predicted fits to the phosphorylation and the total EGFR responses to 100ng/ml EGF in MCF-10A cells. Model was fitted to the experimentally measured data points (green). Blue curves represent the averages and the standard deviations of the predicted probabilistic distributions from the fits to the phosphorylation data only; red curves represent the averages and the standard deviations of the predicted probabilistic distributions from the fits to the phosphorylation and degradation data.
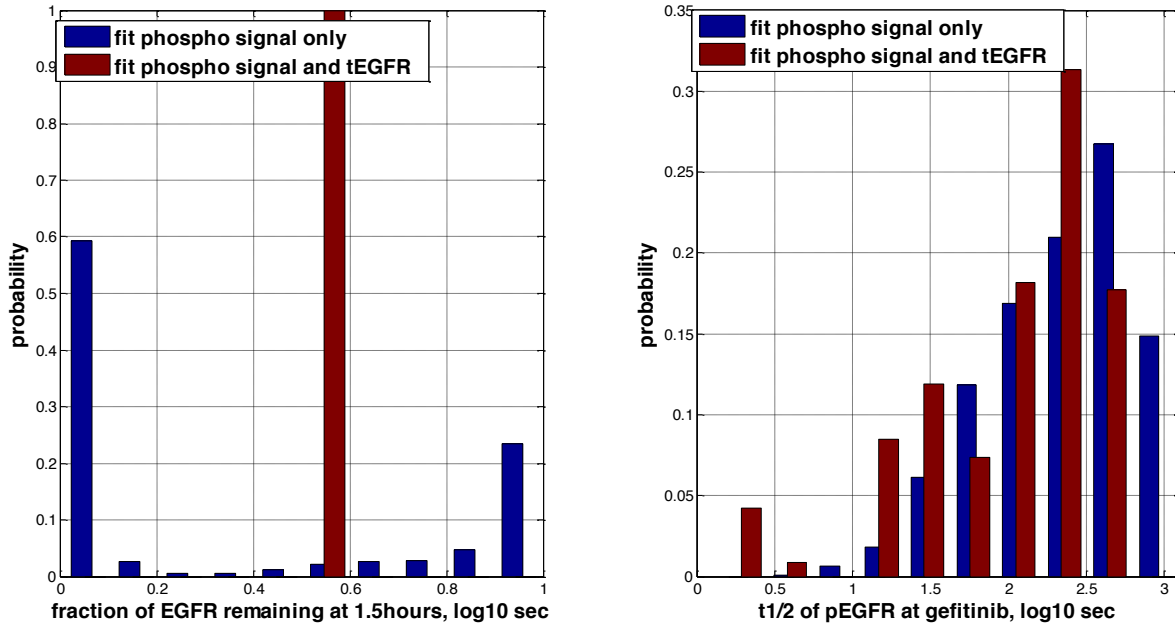
Figure 4.7: Probabilistic predictions: degradation of receptors and activity of EGFR phosphatase. (A) Predicted amount of remaining EGFR at 1.5 hours of stimulation with 100ng/ml EGF. Shown in blue are the distributions obtained from fitting the model only to the phosphorylation time-courses up to 1.5 hours of 100ng/ml EGF treatment in MCF-10A cells; shown in red are the distributions obtained from fitting the model to the same phosphorylation time-courses and the corresponding total EGFR time-course up to 1.5 hours of 100ng/ml EGF treatment in MCF-10A cells. (B) Predicted distributions of pEGFR t1/2 in MCF-10A cells treated with 100ng/ml EGF for 10 minutes followed by addition of saturating dose of pEGFR inhibitor gefitinib at 10mins.

We consider early pEGFR responses (2.5-5 minutes) to 100ng/ml EGF stimulation – predictions that appear to change once EGFR degradation data is included in the fit. In fits to the phosphorylation data only, pEGFR appears to have reached its peak by 2.5 minutes and is decaying by 5 minutes (Figure 4.8), a trend suggested by shifted leftwards predicted distribution of pEGFR at 5 minutes. The shift in the distributions from fitting the phosphorylation and the degradation data suggests that between 2.5 and 5 minutes, the pEGFR signal is predicted to mostly increase. Therefore, including the experimentally measured EGFR degradation data in the fits affects pEGFR activation dynamics by slowing it down.
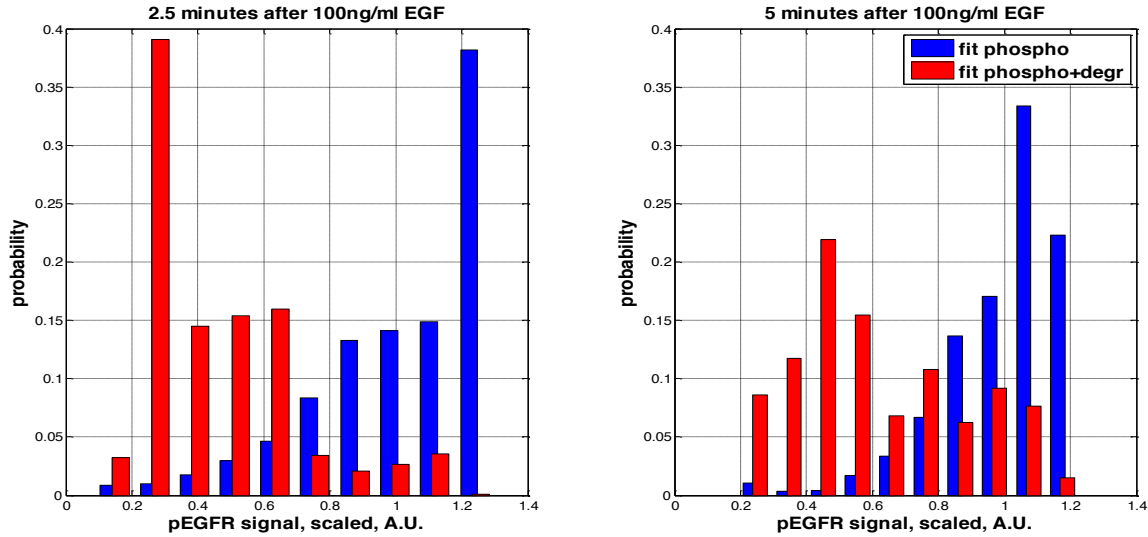
Figure 4.8: Predicted distributions of the early pEGFR responses in MCF-10A cells treated with 100ng/ml EGF.Model was fitted only to the phosphorylation time-courses alone, or to the phosphorylation and the tEGFR time-courses measured in MCF-10A cells treated with 100ng/ml EGF.

We then analyze the changes in the model parameter space that occurred to fit the EGFR degradation data. We consider the differences in the averages of predicted parameter distributions from each of the two fitted conditions: the first condition fitting phosphorylation signal only, and the second condition fitting the phosphorylation and the total EGFR time-courses (Table 1). We rank the parameters by the differences in the averages and find that most of parameters on the top of the list are related to EGFR internalization, dimerization, and phosphorylation. We include parameters related to activities of EGFR phosphatases and the rates of degradation of active receptors – reactions that directly affect tEGFR abundance by degrading phosphorylated receptors and determining what fraction of the total receptors is phosphorylated and thus subject to faster degradation. Fits with experimentally observed tEGFR degradation are predicted to have a more efficient rate of recycling of cPP proteins (k15), which facilitate internalization of receptors, and also have a faster rate of direct internalization of receptors (k6), possibly indicating enhanced removal of active receptors from the cell surface. However, changes in these parameters are also accompanied with a decreased rate of EGFR phosphorylation (kd123), and a faster *kcat* of endosomal RTK phosphatase (kd95), which together could suggest more efficient dephosphorylation of active receptors. Signal attenuation through receptors dephosphorylation rather than through their degradation thus could lead to preservation of receptors from degradation. Interestingly, we find that

parameters related to cytoplasmic RTK phosphatase (c5000 and k940) suggest a higher abundance of the phosphatase and a faster rate of its binding to activated cytoplasmic receptors. Since, in the model, phosphorylated surface receptors are internalized much faster than unphosphorylated ligand-unbound receptors, having a more active cytoplasmic RTK phosphatase could lead to more rapid dephosphorylation of cytoplasmic receptors, which thus could help to attenuate signal, with a decreased receptors degradation. Therefore, the model predicts the importance of parameters of RTK phosphatases to fit the experimentally measured EGFR abundance time-courses, suggesting a relationship between EGFR degradation and activities of RTK phosphatases, both of which attenuate pEGFR signaling.

Table 1: Top parameters related to EGFR signaling and degradation. The parameters are ranked by the shifts in the averages of their predicted distributions when the total EGFR abundance time-courses were additionally fitted.

	phospho signal only, *m1 (log10)*	phospho signal and tEGFR, *m2 (log10)*	*m2-m1 (log10)*
k15	-5.43	-4.37	1.07
k6	-5.14	-4.48	0.66
kd123	-0.26	-0.67	-0.41
kd2b	-2.60	-2.21	0.39
kd95	1.08	1.45	0.37
kd2	-2.78	-2.43	0.35
k940	-4.65	-4.34	0.31
c5000	4.96	5.25	0.29
k2	-5.10	-5.39	-0.29
kd4	-2.32	-2.60	-0.27
kd5	-0.72	-0.94	-0.22
k60	-3.36	-3.15	0.21

parameter name	Description
k15	recycling of internalized cPP proteins to the cell surface
k6	direct internalization of active receptors
kd123	phosphorylation of EGFR
kd2b	undimerization of EGF bound ErbB1-ErbB2 heterodimer
kd95	kcat of phosphoatase - stronger (in endosomes)
kd2	undimerization of EGFR homodimers
k940	binding of cytoplasmic RTK phosphatase to active EGFR dimers
c5000	concentration of cytoplasmic RTK phosphatase
k2	dimerization of two EGF bound EGFR monomers
kd4	unbinding of cPP from receptors
kd5	release of receptors into endosomes through cPP internalization
k60	rate of degradation of active receptors

To further investigate the relationship between EGFR degradation and RTK phosphatase activities predicted by changes in model parameters space, we compare system's responses assumed to be strongly affected by the corresponding parameters. For example, the amount of EGFR remaining in the system after 1.5 hours of stimulation with EGF can serve as a proxy for the values of parameters related to receptors degradation in the system: The more receptors remain undegraded, the slower is the rate constant of degradation. As an approximate metric of RTK phosphatases

activities on system behavior (phosphorylation responses of pEGFR), we consider the following quantity. In model simulations described in this section, along with the predicted time-courses responses to constant EGF that are used to fit the model, we also generate predictions of a scenario in which after 10 minutes of 100ng/ml EGF treatment signaling phosphorylation of EGFR is inhibited, thus imitating addition of saturating doses of EGFR kinase inhibitor gefitinib. The approximate metric of pEGFR RTK phosphatase activity is then estimated from such responses by measuring a half-life of pEGFR signal immediately after kinase inhibition. A lower pEGFR signal half-life ($t_{1/2}$) could suggest a faster activity of EGFR phosphatase. In most fits, the predicted pEGFR signal half-life is sufficiently fast, on average 5.2+/-4.1 minutes, thus such rapid signal decay is unlikely to be attributed exclusively to the more time-consuming process of receptors degradation. Hence, most probably such rapid signal decay is realized through RTK phosphatases. The predicted distributions of pEGFR signal $t_{1/2}$ show that imposing experimentally observed tEGFR degradation speeds up pEGFR signal decay at inhibitor treatment, as manifested through shifts in predicted distributions to the left (average predicted pEGFR signal $t_{1/2}$ shifted to 2.8+/-2.2 minutes). This suggests somewhat faster activity of the phosphatases (Figure 4.7 B). Since, the experimentally imposed degradation of EGFR was significantly lower than the predicted degradation in the majority of the fits, we hypothesize that RTK phosphatase activity is anti-correlated with the strength of receptors degradation.

To further check if the negative correlation is observed in the predictions of the model fitted only to the phosphorylation data, we consider two subsets of fits from these simulations: in one subset, phosphorylation responses decay is predicted to be fitted with most receptors degraded (Figure 4.7A blue distribution, left-most peak), and in the second subset, phosphorylation is fitted with insignificant receptors degradation (Figure 4.7 blue distribution, right-most peak). We then compare the predicted distributions of pEGFR half-lives at inhibitor treatment between the two subsets, which thus could allow to compare the average strengths of the RTK phosphatases in these two subsets (Figure 4.8). We find that in the fits with strong EGFR degradation, the pEGFR half-life was predicted to be on average 4.49+/- 3.43 mins, whereas in the subpopulation of fits with weak EGFR degradation the pEGFR half-life was predicted to be shorter, on average 1.85+/-0.83 mins. In the fits with EGFR degradation similar to the experimentally measured estimate (between 0.55-0.65 of initial

receptors remaining by 1.5 hours of EGF treatment) the pEGFR half-life was predicted to be even shorter, on average 1.15+/-0.16 minutes. The correlation coefficient between the amount of receptors remaining and the pEGFR half-life at gefitinib was found to be -0.3106 (p-value $<10^{-4}$). It follows that, in the model, faster phosphatase activity could indeed be correlated with lower receptors degradation (Figure 4.9).
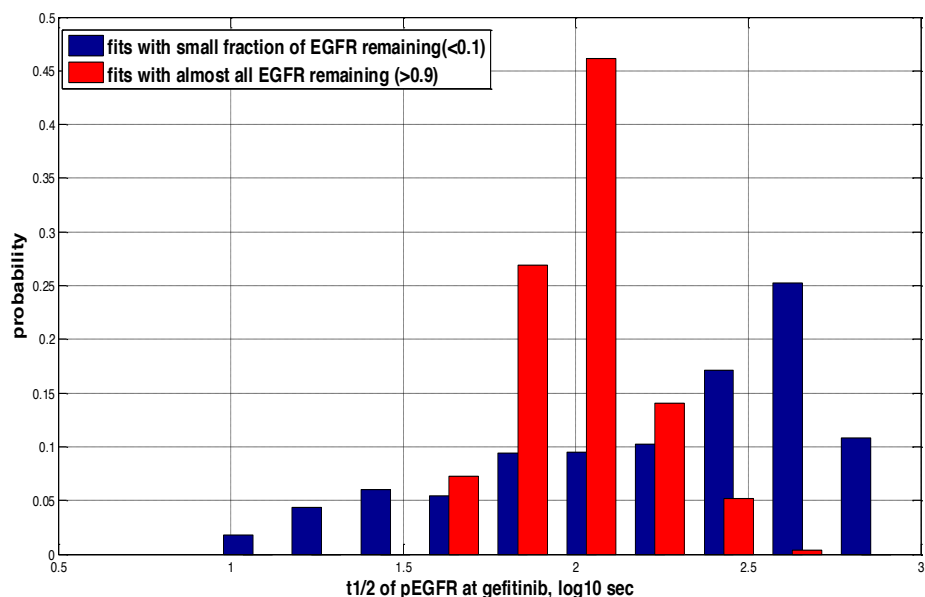
Figure 4.9: Predicted half-life of pEGFR signal at gefinitib inhibition in fits with different strength of EGFR degradation.

We demonstrated how experimental measurements of EGFR degradation could be used to infer constraints on the parameters related to other mechanisms of pEGFR signal attenuation, namely, receptors dephosphorylation by RTK phosphatases. Quantifying relative contributions of these two mechanisms to the attenuation of phosphorylation on EGFR, to the best of our knowledge, has not been previously addressed. It might be interesting to explore what possible combinations of different molecular mechanisms can realize the same cell function (e.g. phosphorylation signal decay) and predict their relative likelihoods to occur *in vivo*. For example, the choice of mechanisms to decay phosphorylation signal can depend on cellular context (e.g. levels of expression of RTK phosphatases) or on various features (e.g. duration or dose of ligand treatment) of the environmental perturbation to which the cell is exposed. Our approach could allow simulating multiple cellular

conditions of interest (i.e. various dose profiles of ligand treatment or perturbations) and gaining an insight about model mechanisms through the properties of model's parameter space.

Section 4.2.3: Estimation of EGFR phosphatase activity through the rate of phospho-EGFR signal decay in response to treatment with EGFR kinase inhibitor Gefitinib

Kinase inhibitors have been previously used to study dynamics of phosphorylation responses. Experimental use of kinase inhibitors has allowed to establish that the observed phosphorylation response is determined by the dynamical balancing of two opposing processes of phosphorylation and dephosphorylation in signaling cascades (Kleiman et al., 2011). Gefitinib is an ATP-competitive inhibitor of tyrosine kinase domains of ErbB. Applying saturating doses of gefitinib to EGF-treated cells promptly inhibits EGFR phosphorylation and results in a rapid drop in phosphorylation signal, on the order of seconds to minutes. Such drastic signal decay cannot be attributed to a slower-scale process of receptor internalization and degradation, and thus is most likely caused by action of RTK phosphatases (Kleiman et al., 2011).

We use experimental data collected in MCF-10A cells in response to EGF pulse gefitinib chase (pulse-chase) experiments reported in (Kleiman, 2010). MCF-10A cells are first stimulated with 100ng/ml EGF for 2 or 10 minutes, followed by addition of saturating dose of gefitinib. Phopshorylation of EGFR is then measured over time every 10 seconds after gefitinib addition. In addition, pERK and pAKT signals are measured as well (see Section 4.2.4).

In the model, we simulate the addition of inhibitor by completely stopping any phosphorylation of EGFR receptors at the time inhibitor is added. Such simplification, however, does not take into account possible time needed for the inhibitor to diffuse in the cells in the experiments. Fitting the model to the phosphorylation and the total EGFR time-courses as described in Section 4.2.2 we generate predicted distributions of the pEGFR half-life at gefitinib addition in cells treated with 100ng/ml EGF for 10 minutes. For comparison, the experimentally measured pEGFR half-life in MCF-10A cells is reported as ~ 43 sec in (Kleiman et al., 2011) (Figure 4.10). Although, the model predicts that the majority of fits have a much slower pEGFR half-life at gefitinib addition, the experimental estimate is predicted to have a non-zero probability of ~0.1. Therefore, while the experiments suggest

a fast activity of RTK phosphatases *in vivo*, the model predicts that phosphorylation signal can be fit even with a much slower phosphatases, possibly due to compensatory degradation of EGFR as a second signal attenuation mechanism.
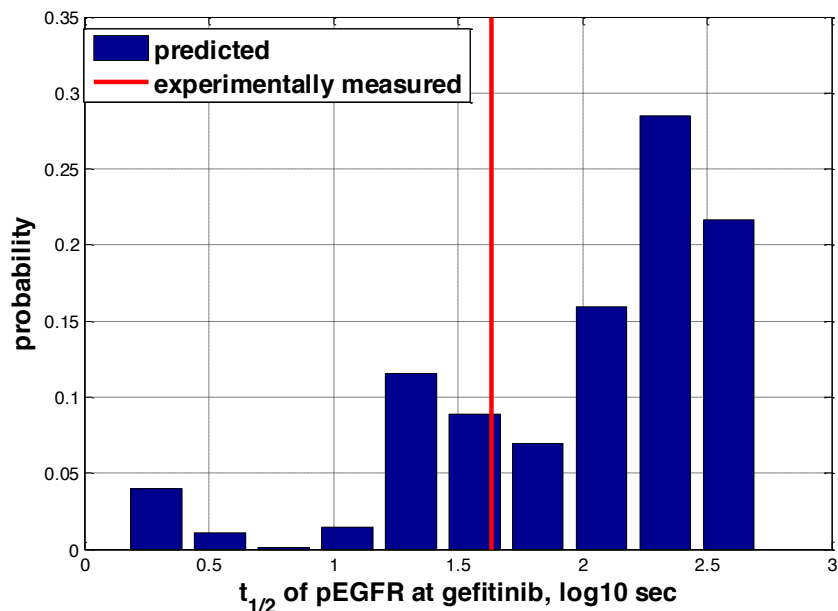
Figure 4.10: Predicted half-life of pEGFR signal at gefitinib addition, 10 minutes after stimulation with 100ng/ml EGF. Model was fitted to the phosphorylation and the tEGFR time-courses measured in MCF-10A cells. Shown in red is an experimental estimate of the pEGFR half-life after 100ng/ml EGF and gefitinib applied in MCF-10A cells and reported in (Kleiman et al., 2011)

We can use our model to explore the effect of gefitinib inhibition of pEGFR on its downstream target, pERK, and on the levels of EGFR abundance. Figure 4.11 shows predicted responses to the constant stimulation of 100ng/ml of EGF of pEGFR, pERK and total EGFR (bold lines), which were fitted to the experimental measurements, along with the predicted responses of these targets to gefitinib inhibition at 10 minutes after 100ng/ml EGF stimulation (shown in dashed lines). In the shown fit, the pEGFR half-life is very short, ~26 sec, and that of pERK is on the order of 3 minutes - values which reasonably well agree with the corresponding experimental estimates (pEGFR $t_{1/2}$~43 sec, pERK $t_{1/2}$ ~4mins (Kleiman et al., 2011)). This fit predicts that immediately after gefitinib addition the total EGFR would remain constant, suggesting that after the disappearance of phosphorylated receptors, degradation of receptors stops. Interestingly, such observation could suggest possible therapeutic implications of gefitinib treatment – by inhibiting phosphorylation of receptors in the presence of high doses of EGF, gefitinib still preserves a number of EGFR receptors, therefore,

potentially maintaining cell sensitivity to consequent changes in EGF doses. For example, for a sufficiently high increase in EGF dose, gefitinib action could be overpowered, leading to activation and phosphorylation of receptors. Since our model implements gefitinib action in a general, non-detailed, all-or-none manner, the model cannot be used for exploration of this effect. However, other studies have focused specifically on in-detail modeling of interactions of kinase inhibitors with receptors and their effects on signaling (Chen et al., 2009; Kleiman et al., 2011).
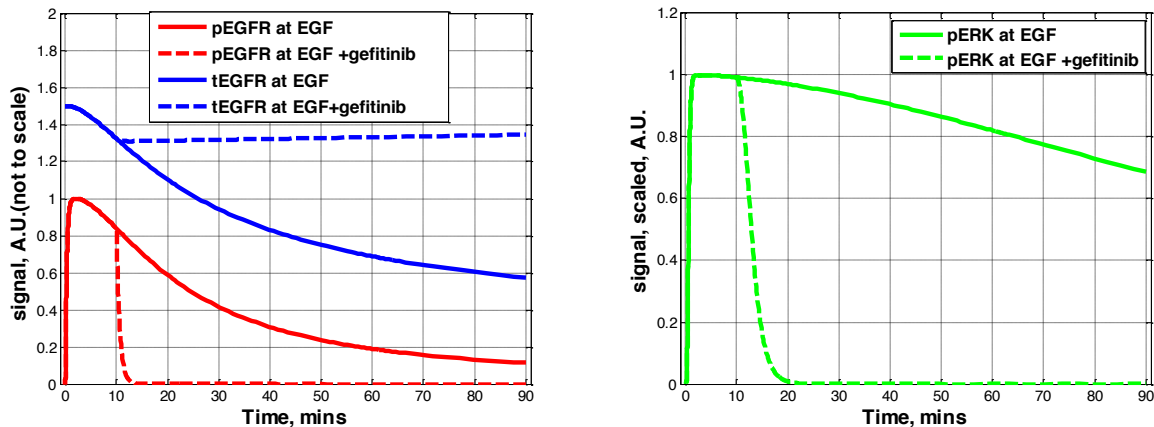
Figure 4.11: Simulated time-courses of MCF-10A cells responses to 100ng/ml EGF with addition of pEGFR inhibitor gefitinib at 10 minutes.

We proceed to use the predicted parameter distributions obtained from fitting the model to the phosphorylation and the total EGFR abundance at 100ng/ml EGF treatment for 1.5 hours to deduce information about parameter values in which the data can be fitted with similar to the experimentally measured *in vivo* strength of RTK phosphatases. Figure 4.12 shows predicted distributions of parameters related to EGFR activation and RTK phosphatase activity obtained from all found fits (in blue), and from a subset of fits with RTK phosphatase activity as measured experimentally (in red). We see that although not all parameters are predicted to differ in their distributions (for example, kd950, the rate of kcat of cytoplasmic RTK phosphatase) in the fits with fast RTK phosphatase activity, some parameters distributions seem to be shifted in a trend suggesting a more abundant RTK phosphatase (c5000) and a faster rate of its binding to active receptors (k940), which is consistent with the faster rate of pEGFR signal drop at gefitinib inhibition. Somewhat shifted distributions of parameters not directly related to RTK phosphatases, such as, for example, the rate of EGFR degradation (k60), internalization (k6, kd6) or dimerization (k2, kd2) of

receptors, suggest that not only dephosphorylation of receptors, but also other dynamic processes can affect the observed pEGFR behavior at gefitinib inhibition. Overall, the predicted distributions can be used to get an insight about values of parameter space in which the model is consistent with phosphorylation and total EGFR experimental data and fits this data with the experimentally observed strength of RTK phosphatase (Figure 4.12, in red).
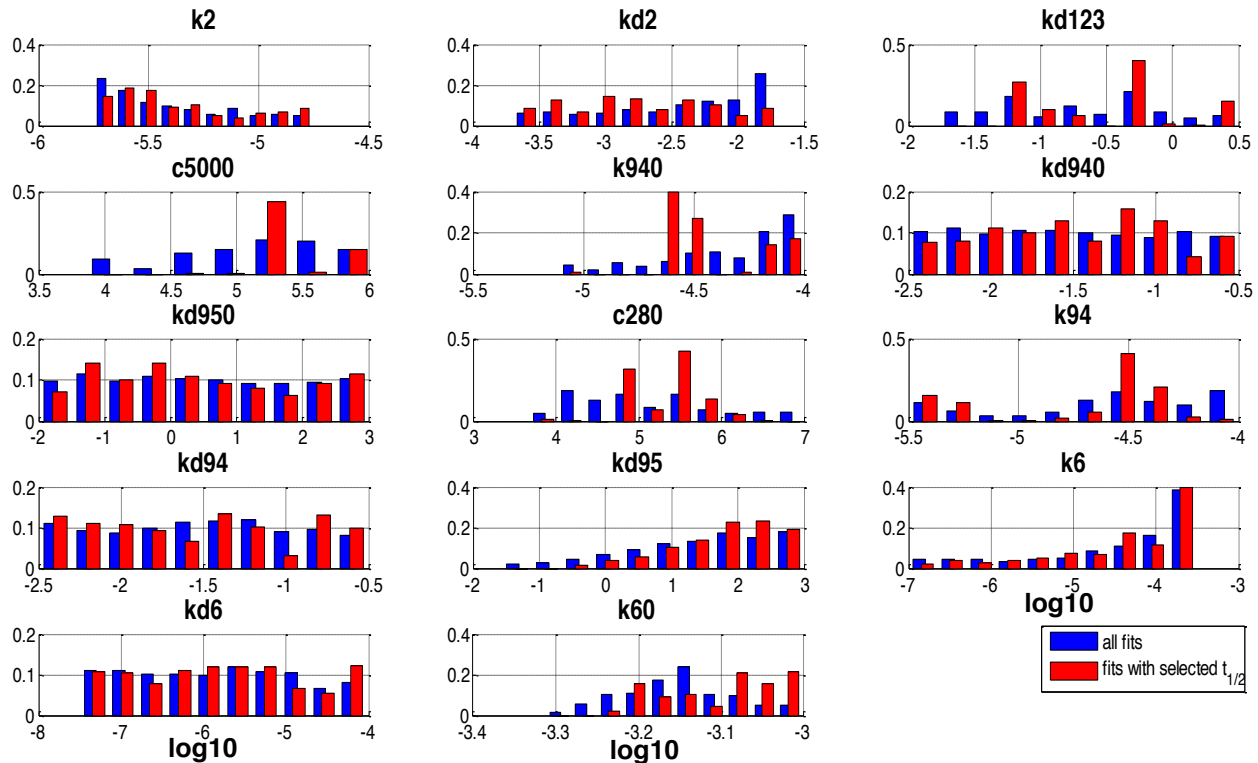
Figure 4.12: Predicted parameter distributions from model fitted to the phosphorylation and the total EGFR time-courses measured in MCF-10A cells. Shown in red are the parameter distributions from the subset of fits with the half-life of pEGFR at gefitinib being within the experimentally measured estimate of 10-40 seconds. For parameters description see Table 1.

Section 4.2.4: Estimating ERK phosphatase activity through the rate of pERK decay at MEK inhibition

In our model, pERK downregulation is realized through the action of pERK phosphatase, as well as through signal decay at the receptors level. To assess the rate of phospho-ERK dephosphorylation, we use an EGF pulse, MEK inhibitor chase (pulse-chase) experiment: EGF stimulation of cells is perturbed by the addition of inhibitors of MEK, the upstream kinase activating ERK (Figure 2.1). After the initial stimulation with 100ng/ml EGF for 10 minutes, cells are treated with

a saturating dose of MEK kinase inhibitor PD0325901 and pERK response is measured frequently for one hour afterwards. The iMEK treatment induces rapid decrease in pERK signal with a half-life of pERK signal on the order of minutes. Since in the iMEK experiments pERK is decoupled from regulation by its upstream kinase, the observed rapid pERK decay should be attributed solely to the action of pERK phosphatases. First, it is unlikely that ERK is degraded during EGF stimulation. Therefore ERK degradation cannot contribute to the observed rapid pERK decay at iMEK treatment. In addition, series of experiments reported in literature (Kleiman, 2010) show that inhibition of phosphatases with small molecule inhibitors leads to rapid increase in phosphorylation of their targets and suggest that the phosphorylated targets are under constant downregulation by phosphatases. Hence, pERK signal decay observed in the pulse-chase experiments is most likely caused by the action of ERK phosphatases.

We fit the model to the phosphorylation and the total EGFR time-courses and predict the distribution of the pERK half-life after iMEK addition after 100ng/ml EGF stimulation for 10 minutes (Figure 4.13). The obtained distribution varies across several orders of magnitude, suggesting that the model can fit pERK signal decay in response to persistent EGF stimulation relying on different degrees of strength of pERK phosphatase. The model predict the median pERK half-life at iMEK to be 0.98 mins; however, the experimental estimates of the pERK half-life in an EGF pulse-iMEK chase experiment in MCF-10A are somewhat lower, approximately 2.7+/- 0.3 minutes. The experimentally measured value has a non-zero probability in the predicted distribution (the area shown in red in Figure 4.12). It is possible that our model underestimates the half-life of pERK at iMEK due to our assumption that iMEK addition instantaneously prevents phosphorylation of ERK by pMEK. However, in reality, the iMEK molecules may need time to diffuse into the localization of their targets, therefore possibly slowing down the time to observe pERK signal decay. In addition, it is plausible that inhibitors of MEK are not specific and may have unintended off-target effects that could influence the MAPK signaling cascade not only at the level of MEK but possibly at other signaling components.

Here we demonstrate how our approach can be used to investigate the relationship between model parameters space and model's responses (e.g., pERK signal decay) through analyzing correlations between parameter values and quantities of interest (e.g., pERK signal half-life at iMEK

addition). To explore the relationship between regulation of phospho-ERK signal decay and model parameter space, we look at the correlations between values of the predicted pERK half-life at iMEK addition and individual parameter values. We find that the parameters highly correlated with the rate of pERK decay at iMEK are related to the action of pERK phosphatase. Specifically, effective rate of ERK phosphatase binding (represented by the product of the concentration of ERK phosphatase and the rate of its binding to phosphorylated ERK) is negatively correlated with the observed pERK half-life at iMEK addition (Pearson Correlation coefficient = -0.97, p-value < 10^{-4}); the rate of kcat of the ERK phosphatase also had a negative correlation (Pearson correlation coefficient =-0.32, p-value < 10^{-4}). Such findings are not surprising, as ERK phosphatase is the only direct negative regulator of pERK signal in our model. We also find that the rate of binding of MEK phosphatase to pMEK positively correlates with the pERK half-life at iMEK addition (Pearson correlation coefficient = 0.59, p-value < 10^{-4}). This may suggest that the faster MEK phosphatase binds to its target, the slower the pERK decay at iMEK, and thus the slower the phosphatase of pERK can be. This points to a possible compensatory relationship between actions of phosphatases in the kinase signaling cascade.

To better understand what molecular mechanisms were actually regulating pERK signal decay in MCF-10A cells, we examined the predicted fits with the pERK half-life at MEK inhibition being within the experimental estimate of 2.4-3 minutes. Figure 4.14 shows how constraining the model to the experimentally measured responses to iMEK perturbation allows to significantly limit the model parameter space. In particular, the possible range of values of the effective rate of ERK phosphatase can be narrowed down from about 4 to only 0.5 orders of magnitude based on the EGF/iMEK pulse chase experiment. Overall, this result suggests that to regulate phopho-ERK signal decay the system is more likely to use a pERK phosphatase of moderate activity.

In addition, this example demonstrates how applying our model simulations and additional experimental data could elucidate the relationship between the system's mechanisms (expressed through the model parameter space) and system's function (observed signaling responses). The parameter distributions predicted by fitting the model to a given set of experimental data can be used to generate specific hypotheses about the underlying mechanisms realized to produce the observed system responses. Fitting the model to the generic time-courses data can constrain the model

parameter space. However, imposing constraints from various perturbations experiments should constrain the model parameter space to a more narrow range of specific molecular mechanisms. Therefore, our approach may be useful in optimizing experimental design by suggesting most informative perturbations and conditions to elucidate various molecular mechanisms involved in cell signaling. This topic is an important area of research in system biology and has been previously explored in detail (Liepe, Filippi, Komorowski, & Stumpf, 2013).

We now briefly describe several other examples of how our model predictions can be used to explore the relationship between the model parameter space and its functions. For instance, we can consider two subsets of predicted fits with opposing pERK half-life at iMEK perturbation: one set with pERK half-life under 1 minute, and the second set – on the order of half an hour. We can then compare parameter distributions different between the two subsets and find the most distinct parameters between the two observed pERK behaviors. Therefore, such comparison can provide an insight into the regulation of pERK with various underlying mechanisms, yet resulting in similar pERK behaviors.

Another possible application of our approach can be used to differentiate between the two possible types of kinase activation and deactivation reactions. For example, activation of pERK relies on phosphorylation of two of pERK residues and is performed by the same enzyme, MEK. Such activation is processive if in one binding to its target, the enzyme performs the two phosphorylation steps; otherwise, if the enzyme unbinds and rebinds between the two steps, the reaction is referred to as distributive. Recent works suggested that the double-phosphorylation activation of ERK and its dephosphorylation *in vivo* are distributive processes. However, some studies reported that under certain molecular condition of the so-called molecular-crowding effect (Aoki et al., 2011) the double ERK phosphorylation reaction may become processive. pMEK activation also relies on two phosphorylation steps.

We can fit our model to the experimental data and by examining the obtained parameter distributions and dynamics of individual intermediate variables not measured experimentally, we can generate predictions as to which two-step reactions are processive or distributive in the ErbB signaling cascades. To allow a possibility of both processive and distributive mechanisms of ERK

109

activation and deactivation in our model, we implement each of the activation/deactivation steps as separate reactions with distinct kinetic rate constants. We then can simulate these parameter values by fitting the model to the phosphorylation and total EGFR time-courses and impose the constraint of pERK half-life at MEK inhibition to be within the experimentally observed range. The model simulation then allows to sample various parameter combinations in which pERK signaling is fitted through the activation/deactivation processes of distributive or processive types. To distinguish whether in a given model fit the ERK activation is realized via processive or distributive phosphorylation, we examine dynamics of intermediate variables, namely singly phosphorylated ERK (Figure 4.15). The model suggests that processive activation of ERK is characterized by a sharp increase in the double-phosphorylated ERK species and with low levels of the intermediate single-phosphorylated ERK, indicating that all ERK is being rapidly converted into double-phosphorylated ERK at the beginning of stimulation. Similarly, one can distinguish between processive or distributive dephosphorylation of ERK and MEK by analyzing dynamics of single- and double-phosphorylated ERK at later times, after the signal peak, when signal is decaying due to the action of ERK phosphatases and decreased activation from upstream kinase. Generating model fits with two different types of the two-step activation reactions and comparing their parameter distributions could then provide an insight into specific parameters that determine the type of the reaction.
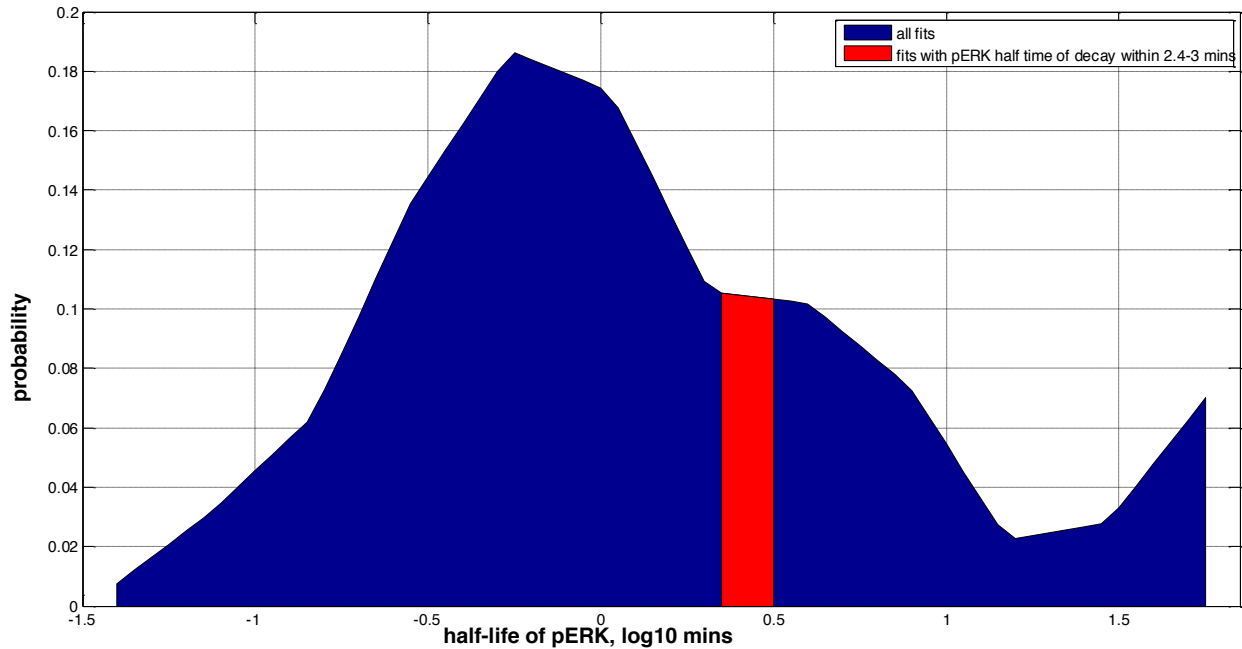
Figure 4.13: Predicted distribution of pERK $t_{1/2}$ after addition of MEK inhibitor in MCF-10A cells treated with 100ng/ml EGF. Shown in red is the corresponding experimental estimate measured in MCF-10A cells.
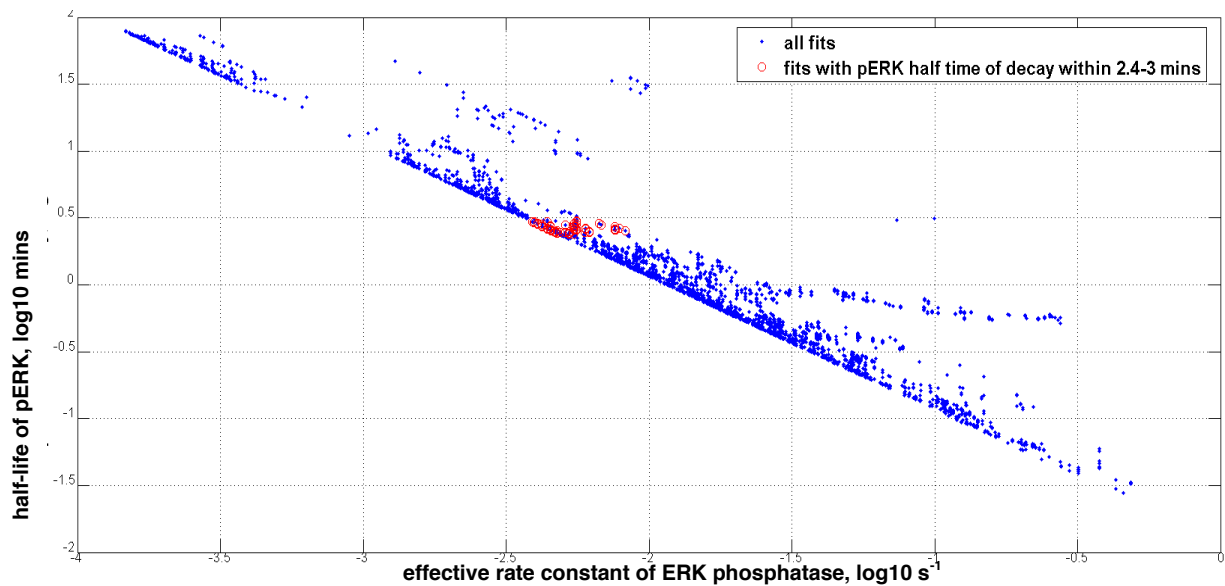
Figure 4.14: Correlation between the predicted half-life of phospho-ERK at MEK inhibition and the rate of ERK phosphatase parameters.

In addition to phosphorylation signal decay on pERK in response to perturbation with MEK kinase inhibitor, we also have experimental measurements of pERK signal decay caused by kinase inhibition of pEGFR with gefitinib. The data suggests that upon gefitinib addition at 10 minutes, pERK in MCF-10A cells decays with a half-life of approximately 3.5 minutes, and pAkt is dephosphorylated

slightly faster, with $t_{1/2} \sim 1.2$ minutes. Using half-lives of pERK in response to signal inhibitions at the upstream, pEGFR level, or the downstream, MEK level, can help to understand the information flow in signaling cascades. In particular, by fitting the model to the phosphorylation and the total EGFR time-courses and imposing constraints on the half-life of pERK signal after perturbation with either of the inhibitors, and then comparing the resulting changes in the model parameter space, one might assess information content of each of the perturbations. Furthermore, including experimental data from both of the perturbations to be fitted in the model can help better constrain signaling cascade parameters relevant to signal propagation between pEGFR and pERK.
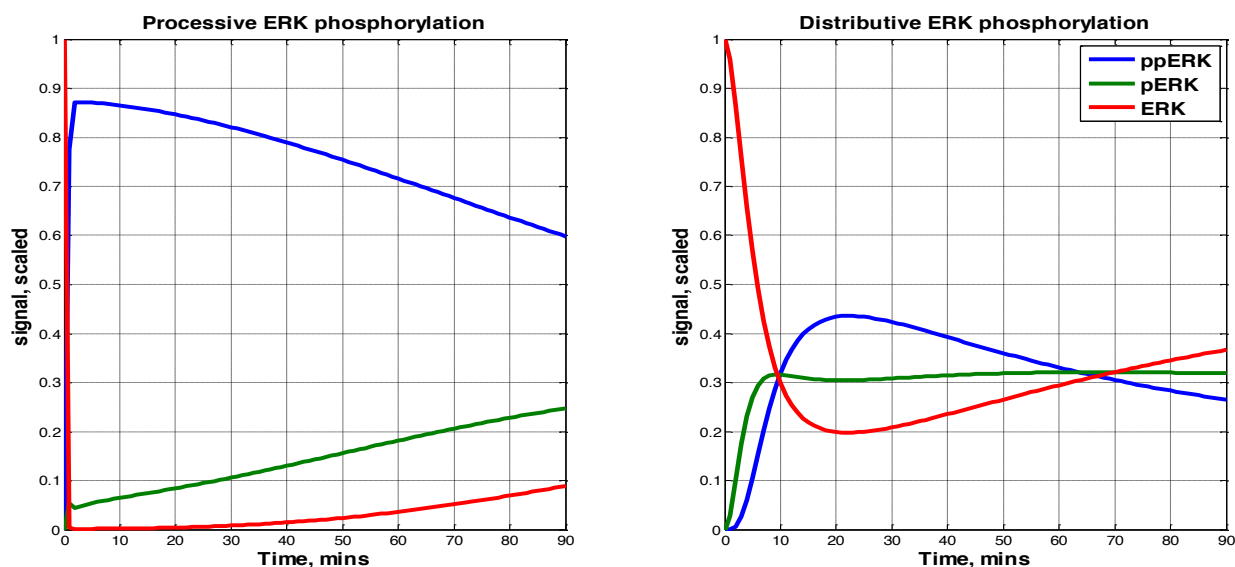
Figure 4.15: Examples of model-predicted time-courses of ERK species in the two types of phospho-ERK activation: processive and distributive. Time-courses of ERK, and intermediate species - single phosphorylated pERK and final product - double phosphorylated ERK.

Section 4.2.5: Phosphorylation signal decay and EGFR abundance recovery in EGF washout experiments

In this work we consider signaling responses to stimulations with constant doses of EGF over several hours. However, we also use the results of two types of perturbations - addition of kinase inhibitor (described in Sections 3.2.3 and 3.2.4) and EGF washout experiments presented in this section. In ligand washout experiments, cells are being treated with various constant doses of EGF for ten minutes, followed by replacement of the cells media with a ligand-free media.

112

Ligand removal from the cell surface is a physiologically plausible scenario, in which, for example, growth factors can be rapidly removed from the cell surface by blood, or alternatively, slowly diffuse away. Hence, it may be useful to understand what mechanisms are shaping cell responses in the conditions of ligand washout. Furthermore, it may be interesting to study how rapid changes in extracellular stimuli are propagated towards the downstream targets and if buffering or time-delay effects are present in the cascade.
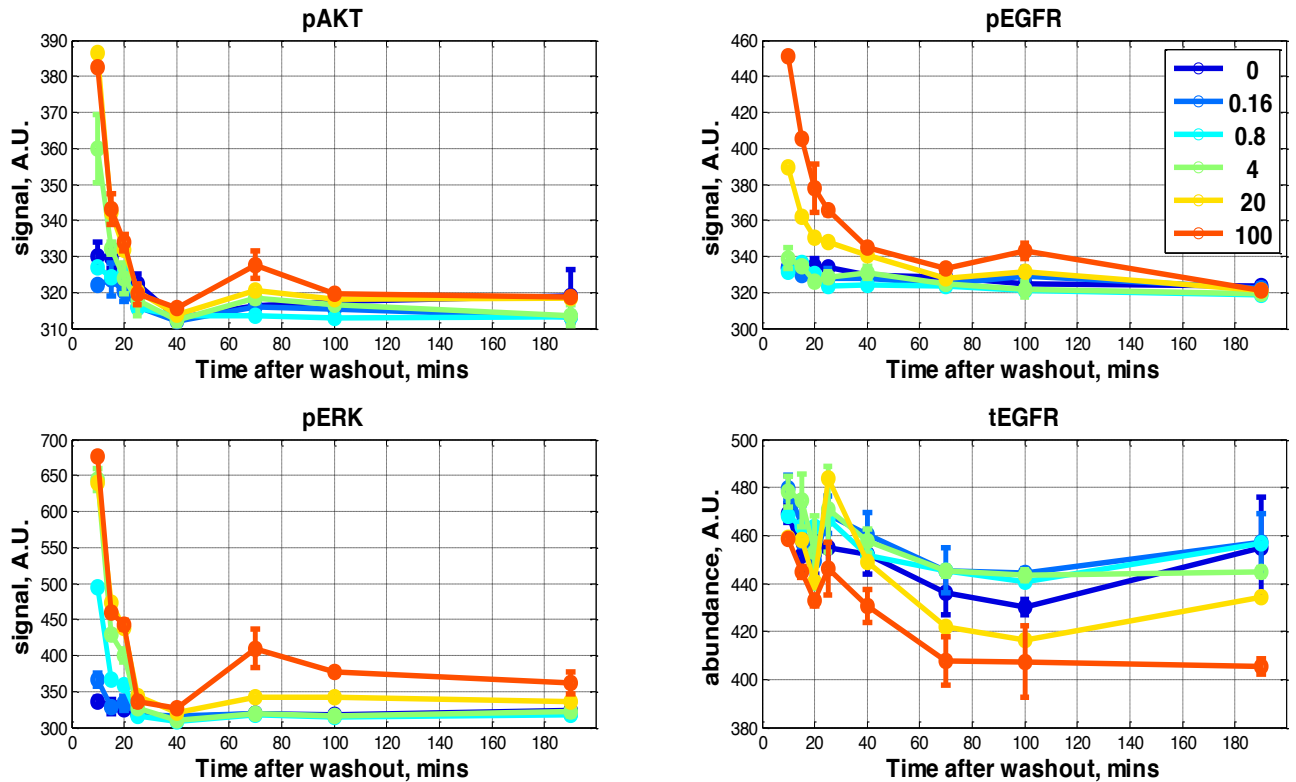
Figure 4.16: Ligand washout experimental data: phosphorylation and total EGFR time-courses measured in MCF-10A cells exposed to EGF for 10 minutes and subjected to EGF washout at time=10mins.

Figure 4.16 shows the experimental measurements of the phosphorylation responses and the total EGFR time-courses in MCF-10A cells after they were treated with a range of EGF doses for 10 minutes and subjected to EGF washout (at t=10 minutes in the graph). Phosphorylation signal and the total EGFR abundance are measured after EGF washout for up to three hours. These experiments show that phosphorylation signal drops after EGF is removed, with the rate of decay apparently dependent on the dose of EGF in the initial 10 minute treatment: at EGF doses between 4 and 100ng/ml, the half-life of pEGFR is between 5 and 10 minutes (estimated as 6.75+/-1.75

minutes), whereas at lower doses (EGF < 0.8 ng/ml) the half-life of pEGFR appears to be between 10 and 15 minutes (Figure 4.16). Phosphorylated ERK and AKT signals also appear to decay with relatively fast dynamics.

Interestingly, pEGFR drops slower at EGF washout ($t_{1/2}$~ 8 mins after 100ng/ml EGF) than at kinase inhibition ($t_{1/2}$~ 43 sec minutes after 100ng/ml EGF), possibly due to the fact that at the time of washout EGF bound to cell surface receptors remains in the media and thus stimulates some remaining signaling. Moreover, in the kinase inhibition experiment, EGF is constantly present in the media, and gefitinib competes with ATP for binding to the receptors. Gefitinib bound to receptor temporarily inhibits its activation, until it unbinds. This demonstrates that although both the EGF washout and the pEGFR kinase inhibition lead to pEGFR signal decay, these two perturbations lead to phosphorylation signal decay through very different mechanisms. Hence, using experimental inhibition from both types of such perturbations should better constrain various parameters in our model and provide more precision about the mechanisms used in the model to fit signaling dynamics.

The total EGFR time-courses data shows that even after EGF is washed out, receptors continue to degrade over some time – possibly reflecting the degradation of the remaining phosphorylation receptors in the system. However, by three hours after EGF washout, total EGFR receptors appear to recover to nearly their pre-stimulus levels at most EGF doses, and possibly at the 100ng/ml as well, although at a slower rate. Since the experimental measurements count all total EGFR receptors, the observed recovery in the number of receptors is not likely due to mobilization or transport of previously inactive in signaling receptors, as those would have been detected in the total receptors abundance. Therefore, it is most probable that the observed recovery was due to de novo synthesis of receptors. In addition, this observation suggests that relatively short, on the order of 10 minutes, pulses of EGF may be forgotten by cells within several hours after their removal, depending on the dose of the treatment. However, pulses of higher EGF doses might change the number of receptors on the cell surface for several hours after removal of the pulses.

Additionally, we conduct an experiment in which MCF-10A cells are stimulated with 100ng/ml EGF for 10 minutes, followed by EGF washout, incubation for 60 minutes, and restimulation with 100ng/ml for 10 minutes (Figure 4.17). Although this EGF re-addition experiment was not used in

fitting the model, it serves as an additional confirmation of the model predictions. The experiment showed that the amplitude of pEGFR responses at the second EGF addition is significantly lower than the amplitude of the initial response, even though the EGF doses in both stimulations are the same. Therefore, the cells have reduced their sensitivity to EGF, possibly due to the fact that one hour after EGF washout is not sufficient to restore the system's sensitivity after a high dose EGF treatment. The time-courses of tEGFR measurements in the experiments where washout is applied 10 minutes after 100ng/ml EGF (Figure 4.16) show that, even after ligand washout, tEGFR continues to decay, most likely due to the presence of phosphorylated receptors. Recovery in tEGFR is thus delayed. The data shows that by 1 hour after EGF washout in the 100ng/ml stimulated cells, tEGFR abundances decrease to their lowest values, and start to recover from that time. Moreover, since this is an estimate of the total EGFR in the cell, it is not clear what fraction of these receptors has returned to the cell surface, where it can bind ligands, as opposed to still being transported to the cell surface in some ligand-inaccessible intracellular compartments.
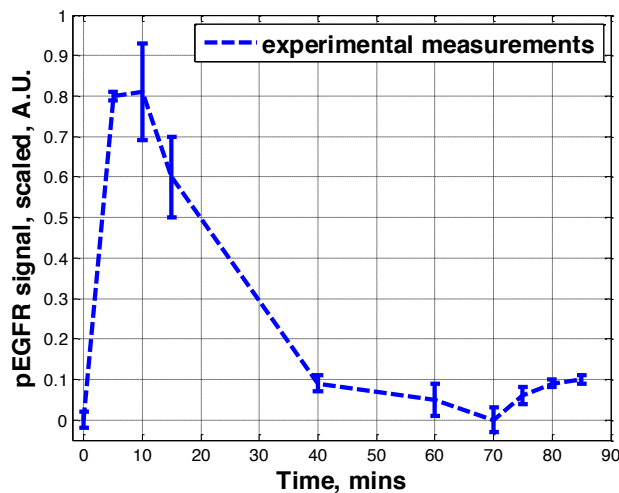
Figure 4.17: Experimental confirming of decreased sensitivity to EGF in MCF-10A cells previously exposed to EGF .MCF-10A cells were treated with 100ng/ml EGF for 10 minutes. At t=0, 100ng/ml EGF is applied, at t=10 minutes, EGF is being washed out from the cell media, at t=70 mins 100ng/ml EGF is added to the media. Reported are the averages of triplicates and the standard errors.

We fit the model to the phosphorylation data and the total EGFR time-courses in response to continuous stimulation with various EGF doses and predict the distribution of pEGFR half-lives at EGF washout after 10 minutes of 100ng/ml EGF. The model predicts an average pEGFR $t_{1/2}$ of 13.5

+/- 0.3 minutes, whereas, the experimental estimate of the pEGFR half-life in the washout after 100ng/ml is approximately 6.75+/-1.75 minutes, and is substantially faster. The model predicts that on average, tEGFR by three hours after EGF washout would recover to approximately 70% of its initial levels (Figure 4.18, dashed blue line).

We then refit our model to the same phosphorylation and the total EGFR time-courses. In addition, we impose that the pEGFR $t_{1/2}$ at EGF washout after 100ng/ml is within 6.75+/-1.75 minutes, according to the experimental data. The model predicts the fastest pEGFR signal drop at EGF washout at 100ng/ml (Figure 4.18, red dashed line) and pEGFR $t_{1/2}$ on average to be 8+/-0.41 minutes. In addition, tEGFR is predicted on average to recover to its initial levels by three hours after EGF washout
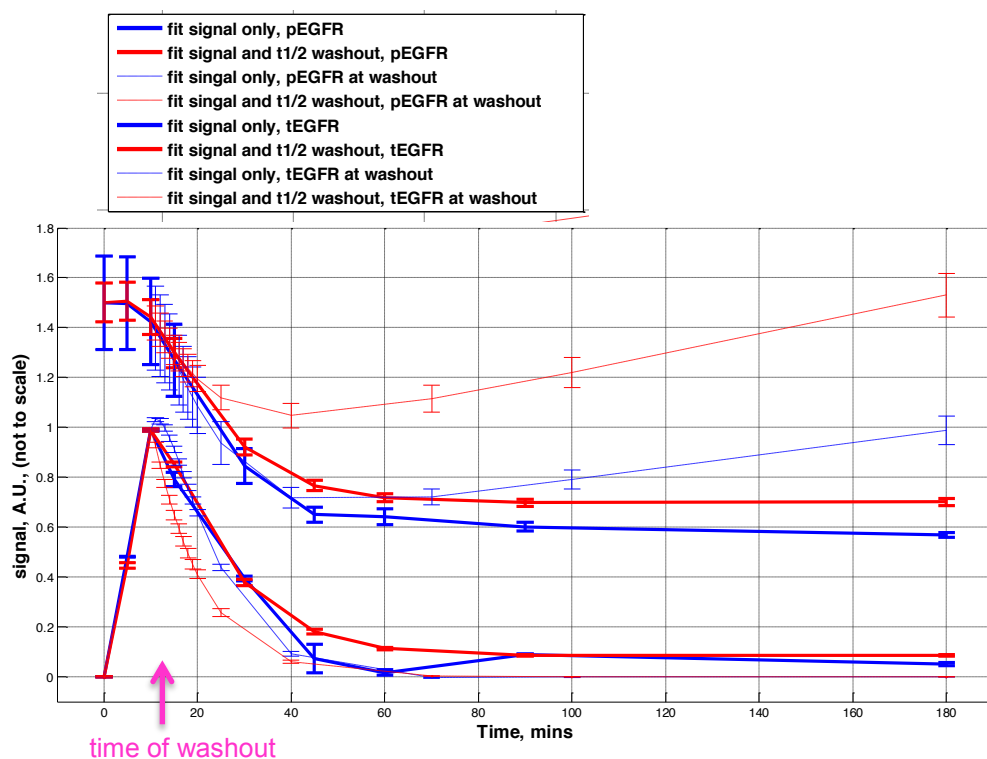
Figure 4.18: Distributions of time-courses predicted by model fitted with or without the experimental estimate of the pEGFR half-life after 100ng/ml EGF followed by EGF washout in MCF-10A cells.Shown are the averages and the standard errors of the predicted distributions from model fits with (red) and without (blue) the pEGFR half-life at washout estimate included.

Table 2: Comparison of the averages of the predicted distributions of parameter values fitted to the phosphorylation and the tEGFR data versus the averages from the distributions obtained by additionally imposing the experimentally measured pEGFR half-life at washout.

parameter name	phospho and tEGFR data only, *m1* (log10)	phospho, tEGFR and washout data, *m2* (log10)	difference between the averages, *m1-m2*
kd31	-2.80	-0.73	-2.07
kd2a	-0.17	1.75	-1.92
kd1	-2.73	-1.22	-1.51
kd2	-1.69	-0.68	-1.01
k2a	-7.32	-6.37	-0.95
k60	-3.18	-2.59	-0.59

predicted pEGFR $t_{1/2}$ at 100ng/ml washout	13.58 +/- 0.33 mins	8 +/- 0.4 mins	

kd31   koff of EGF unbinding from two-EGF bound dimer
kd2a   undimerization of two-EGF bound dimer
kd1   koff of EGF unbinding from monomer
kd2   undimerization of one-EGF bound dimer
k2a   dimerization of EGF-bound dimers
k60   degradation of active receptors

We then compare the distributions of parameters that have shifted to accommodate faster pEGFR decay at EGF washout. We find that the most shifted in distributions parameters are related to processes of EGF unbinding, receptors undimerization, and degradation of active receptors. Specifically, the rates of EGF unbinding from monomers and fully occupied dimers were faster by 1.5 and 2 orders of magnitude, suggesting faster dissociation of EGF from receptors. In addition, rates of undimerization and active receptors degradation were also predicted to be faster. Therefore, using EGF washout data in our approach we can better constrain parameters important to EGFR ligand kinetics, undimerizaiton and receptors degradation.

Note, that in this example we fitted our model to the phosphorylation and the tEGFR time-courses and pEGFR $t_{1/2}$ at washout only at a single, highest EGF dose. Potentially, we can fit our model to the washout experimental data obtained at different initial EGF pre-treatment and not only

the highest EGF dose (Figure 4.19, Figure 4.20). This, however, would require an additional model simulation step at each dose at which we choose to fit the EGF washout measurement, which would substantially slow down the simulation times compared to model fitting with only one washout condition. We find that fits to the phosphorylation, the tEGFR and the washout data only at high EGF dose can still reproduce reasonably well the features of the washout data at other EGF doses (Figure 4.19, Figure 4.20). In addition, simulating the washout experiments at the found model fits could allow a direct computation of pEGFR $t_{1/2}$ at EGF washout at various pre-treatments with EGF doses, unlike the case with the experimental data, in which pEGFR $t_{1/2}$ at washout can be determined only within the errors of time intervals at which the experimental pEGFR signal was measured after EGF washout (in our case this interval is 5 minutes).
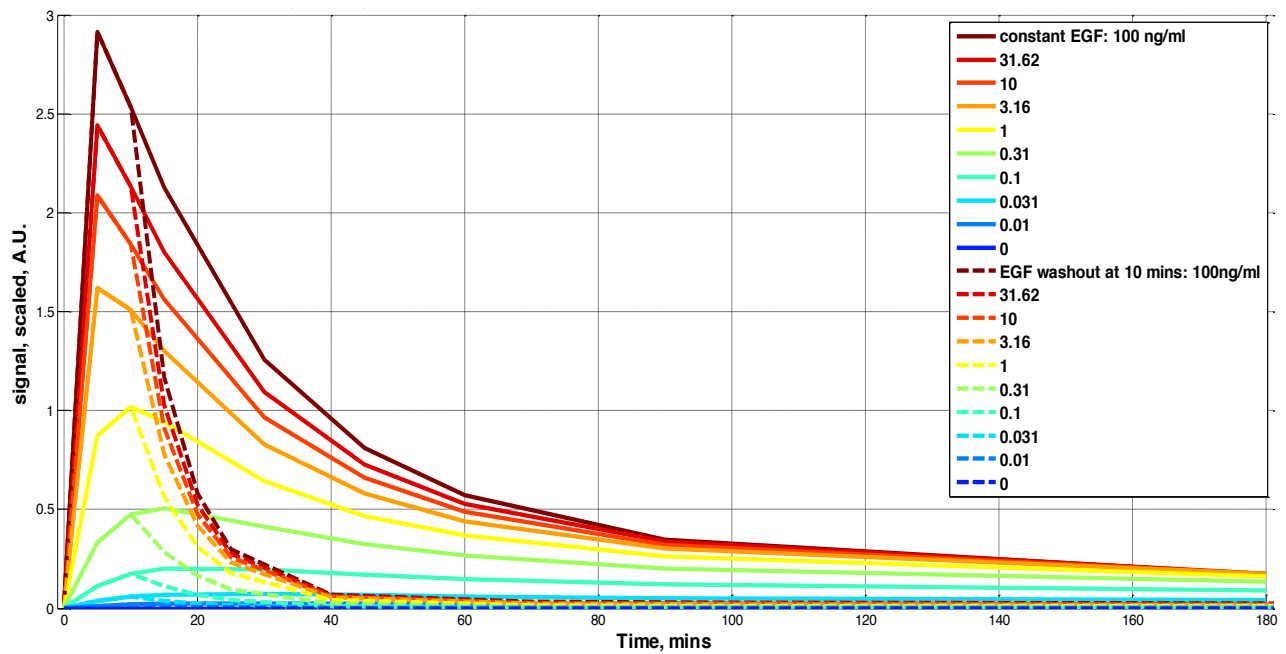
Figure 4.19: Example of model predicted pEGFR responses to stimulations with continuous dose EGF and to stimulations with 10 minutes constant dose EGF followed by EGF washout.
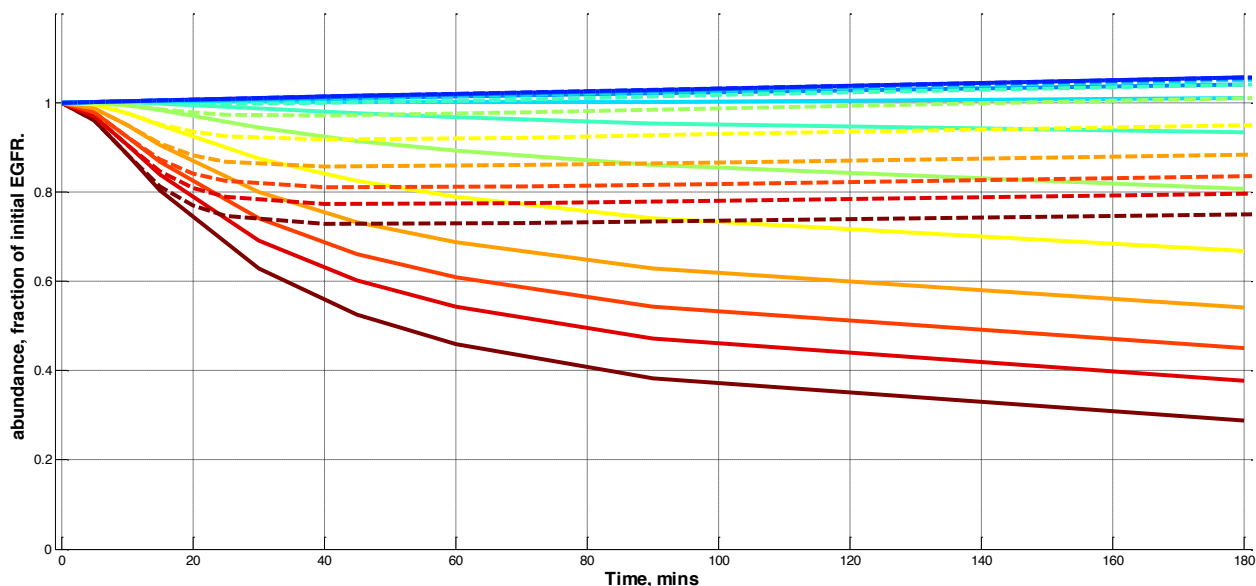
Figure 4.20: Example of model predicted total EGFR responses to stimulations with continuous dose EGF and to stimulations with 10 minutes constant dose EGF followed by EGF washout.

Section 4.3:Sampling results: predicted distributions of parameters values

Section 4.3.1: Predicted distributions of parameter values

We fit our model to the phosphorylation time-courses data measured between 0 and 1.5 hours of treatment with EGF in MCF-10A cells (described previously in Section 4.2.1). We then run MCMC simulations and obtain probabilistic distributions of parameters that were most likely to have produced the observed data. The prior distributions imposed on the ranges of the parameter values were derived based on the literature estimates when available (see Section 3.2.4), and depend on the type of reaction that the parameter represents. To demonstrate our methodology with regards to various features of the parameter space of a large, complex signaling cascade model, in this section we use a previously developed extended EGFR signaling model (Chen et al., 2009) (see Section A.1.1) with 159 parameters. The model encompasses EGFR, ERK and AKT activation and has detailed implementation of EGFR receptors interactions with adaptors and scaffolding proteins.

We generate probabilistic predictions of the protein concentrations and of the kinetic rate constants of reactions (see Section A.1.1). Such probabilistic predictions could provide a valuable insight about likely ranges of *in vivo* protein abundances or rate constants of reactions *in vivo*, quantities which are still challenging to measure experimentally in the mammalian cells (Beck et al.,

2011). Our model simulations predict the levels of various phosphatases that affect cell signaling responses and can be important in downregulating the phosphorylation signal. The prior probability distribution on values of parameter for ERK phosphatase abundance was allowed to vary within three orders of magnitude, 10^3-10^6 molecules per cell; the predicted distribution is skewed towards 10^4-$3x10^4$ molecules per cell (Figure 4.22). The predicted distribution of the abundance of PTEN (c279), phosphatase negatively regulating signaling upstream of AKT, has two peaks and is concentrated either between $3x10^4$ and $3x10^5$ molecules per cell (Figure 4.21). Being able to predict the levels of PTEN in cells may be biologically useful, as loss of PTEN has been implicated in cancer (Sulis & Parsons, 2003).
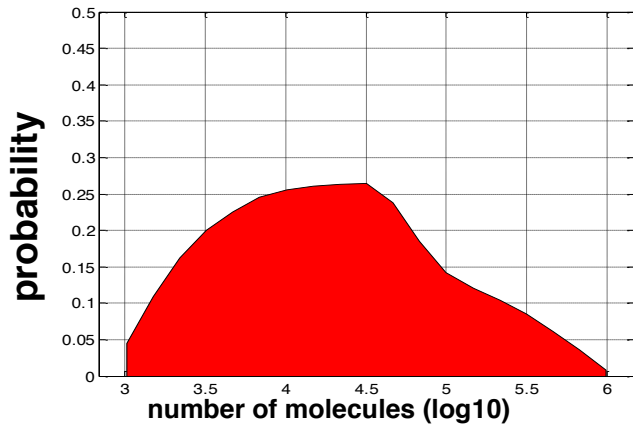
Figure 4.21: Predicted distribution of ERK phosphatase abundance. The model was fitted to the MCF-10A phosphorylation data.
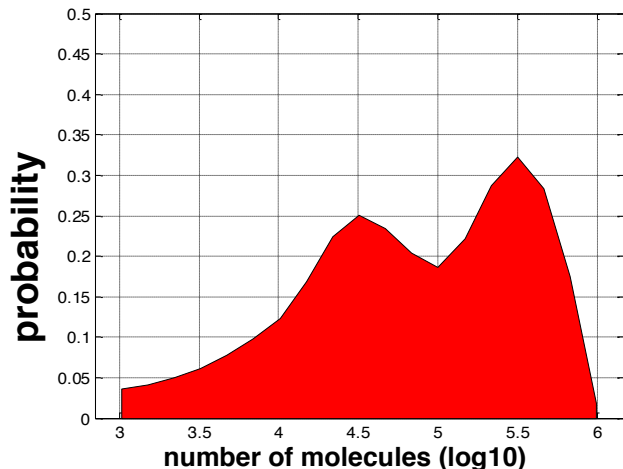
Figure 4.22: Predicted distribution of PTEN phosphatase abundance . The model was fitted to the MCF-10A phosphorylation data.

We can also predict the distributions of initial abundances of various scaffold and adaptor proteins, which bind to the activated EGFR receptors in the cytoplasm and may alter signaling dynamics (Levchenko, Bruck, & Sternberg, 2000). Predicted distributions of initial levels of Sos(c30), Grb2(c22), and Gab1, a Grb2 associated protein 1, (c426), although spread within several orders of magnitudes, still are helpful by suggesting more probable regions of approximately 1-2 orders of magnitude for the abundances of these signaling components in cells in which the experimental data has been collected (Figure 4.23 A-C).

Along with the abundances of scaffold and other proteins, we predict the distribution of initial levels of PIP2 (c444) (Figure 4.23 D), a phospholipid which serves as a substrate for other signaling proteins and is important for activation of the AKT branch (Manning & Cantley, 2007).
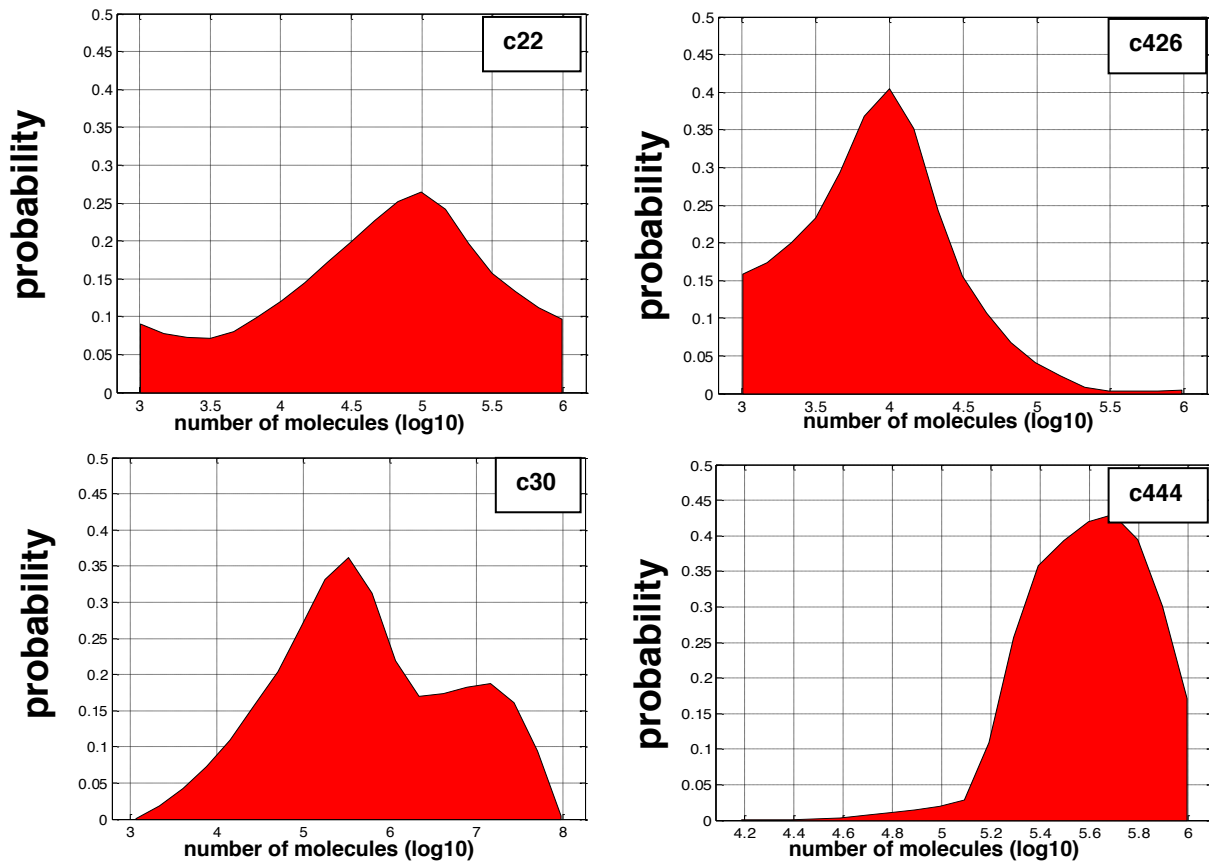
Figure 4.23: Predicted distributions of the adaptor and scaffold proteins abundances: (A) Grb2 (c22); (B) Gab1 (c426); (C) Sos (c30); (D) PIP2 (c444). The model was fitted to the MCF-10A phosphorylation data.

The rate of degradation of active phosphorylated EGFR receptors is an important parameter controlling EGFR signaling (Sigismund et al., 2005). However, since the effects of EGFR receptors degradation are entangled with other dynamical processes of receptors trafficking, such as, for example, receptors recycling, it may be experimentally challenging to correctly establish the rate of receptors degradation. The obtained distribution (Figure 4.25) suggests that within a range of two orders of magnitudes of 10^{-5}-10^{-3}, the degradation rate constant most likely lies in the interval $10^{-3.5}$-10^{-3} sec^{-1}.
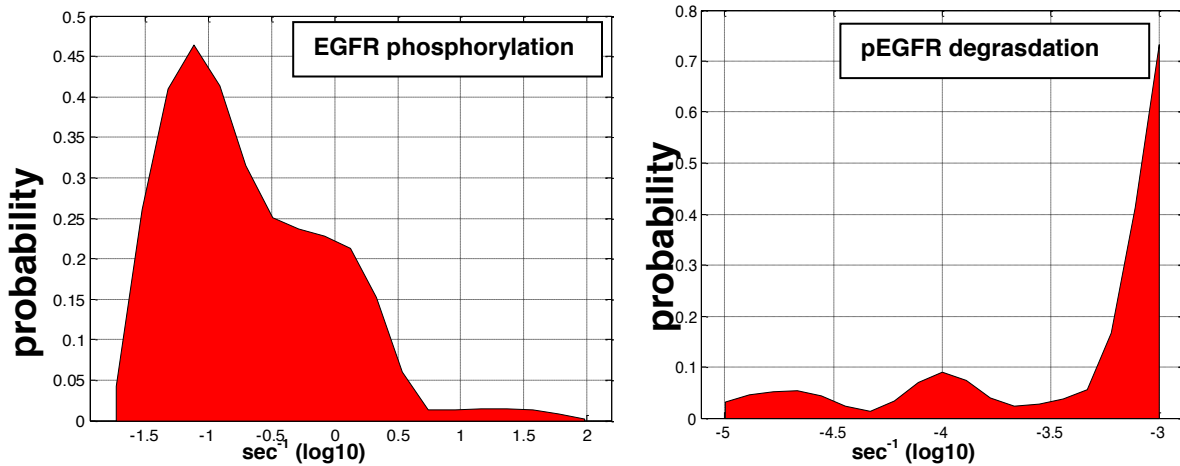
Figure 4.24: Predicted distributions of the rate of EGFR phosphorylation

Figure 4.25: Predicted distribution of the rate of degradation of phosphorylated EGFR The predicted distribution of the rate of EGFR phosphorylation (kd123) although varies across four orders of magnitude, is concentrated around 10^{-1} sec^{-1} (Figure 4.24), which agrees with other estimates of the EGFR activation rate (Chen et al., 2009).
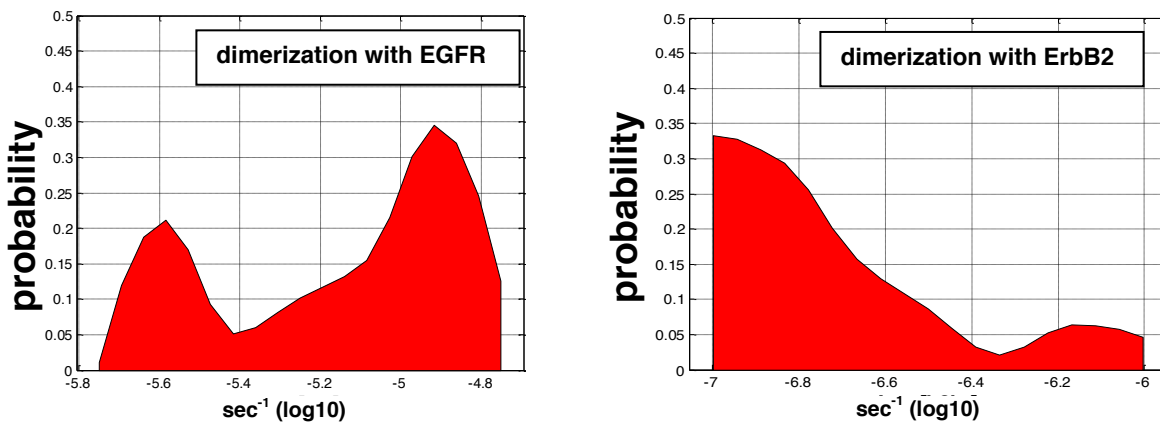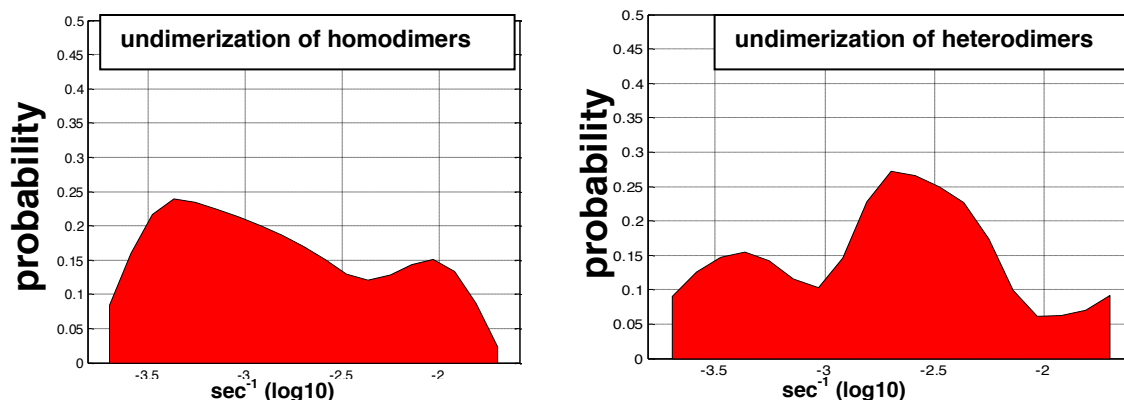
Figure 4.26: Predicted distributions of the rates of EGFR dimerization with EGFR and ErbB2: EGFR homodimerization (A) and undimerization (B), EGFR-ErbB2 dimerization (C) and undimerization (D).

Our model can predict the distributions of rate constants of reactions of dimerization between different species of EGFR receptors (Figure 4.26). For example, in this run of an extended model, EGFR receptors are allowed to form homodimers with other EGFR or heterodimers with ErbB2 receptors. The corresponding rate constants of homodimerization (k2) and heterodimerization (k2b) are allowed to vary within one order of magnitude, and the relative shift in their ranges of values reflects a prior belief that EGFR can have varying affinities to different dimerization partners. The predicted parameter distributions suggest that the rate of undimerization of EGFR homodimers (kd2) is skewed towards lower possible allowed values, whereas, the rate of undimerizaiton of the heterodimers is skewed towards higher values, suggesting faster undimerization of the latter dimers and therefore their reduced stability relative to EGFR homodimers. Although such prediction can be challenging to verify experimentally, it might provide some understanding on how various assumptions about the biochemical properties of receptors, such as their dimerization preferences, can be encoded as the priors on the models parameters and whether such assumptions are supported by the underlying experimental data.

The distributions for some parameters are predicted to be double-peaked, for example, k2 in Figure 4.26 A. It might suggest that there are two different modes in the multi-dimensional parameter space in which k2 is likely to be to fit the given experimental data; the correlation analysis between values of k2 and other parameters could possibly help to explain the double peaked distribution of k2. However, it is also plausible that the double-peaked distribution is a result of slow convergence of the MCMC chains in that parameter dimension, in which case, running longer MCMC chains in that

parameter could allow the chains to cross the barrier between the two regions of parameter space corresponding to each peak.

Certain parameters are predicted to have flat distributions, suggesting that based on the fitted experimental data, the value of the parameter within a given prior range is not important to fit the given data. Several possible conclusions about such parameters could be made from their flat distributions. It may be the case that the imposed range of the prior probability is outside the region where the parameter can have an effect on model responses. For instance, the rate of internalization of inactive receptor (k6c) with very low imposed prior range of 10^{-8} to 10^{-7} is predicted to have a flat distribution in this range (Figure 4.27 A) and suggests this parameter has no significant effect on systems response to which the model is fitted. Such scenario is plausible, given that the experimental data we fit represent phosphorylation dynamics in response to EGF, a condition in which available EGFR receptors upon EGF ligand binding rapidly form dimers and get activated. Therefore, the rate of internalization of inactive receptor monomers, although could be important in determining system's steady state in the absence of EGF, may not be relevant to fit the system phosphorylation responses to EGF treatment.

To check if a parameter with a flat distribution is indeed not important in the model, one might rerun the simulations with a wider range of prior parameter values. Alternatively, one might choose to remove the process represented by this parameter from the model, thus simplifying and reducing the model, while not compromising its ability to fit the experimental data.

For example, we find that the rate constants of EGF unbinding from EGFR monomer in the endosomes (kd10) has a relatively flat, thus, uninformative distribution across about three orders of magnitude, and hence, can be removed from the model (Figure 4.27 B). Biologically, it has been suggested that unlike other ligands, EGF does not dissociate from EGFR receptors in the endosomes, but is degraded with the receptors (Roepstorff et al., 2009), therefore, corroborating the fact that the process of EGF unbinding from monomers in the endosomes might not be relevant to fitting the phosphorylation data in this model. Similarly, we find that some rates of reactions describing adaptors binding to receptors also have flat distributions (k25), suggesting that certain

aspects of interactions between receptors and adaptors could possibly be simplified in this model (Figure 4.27 C).
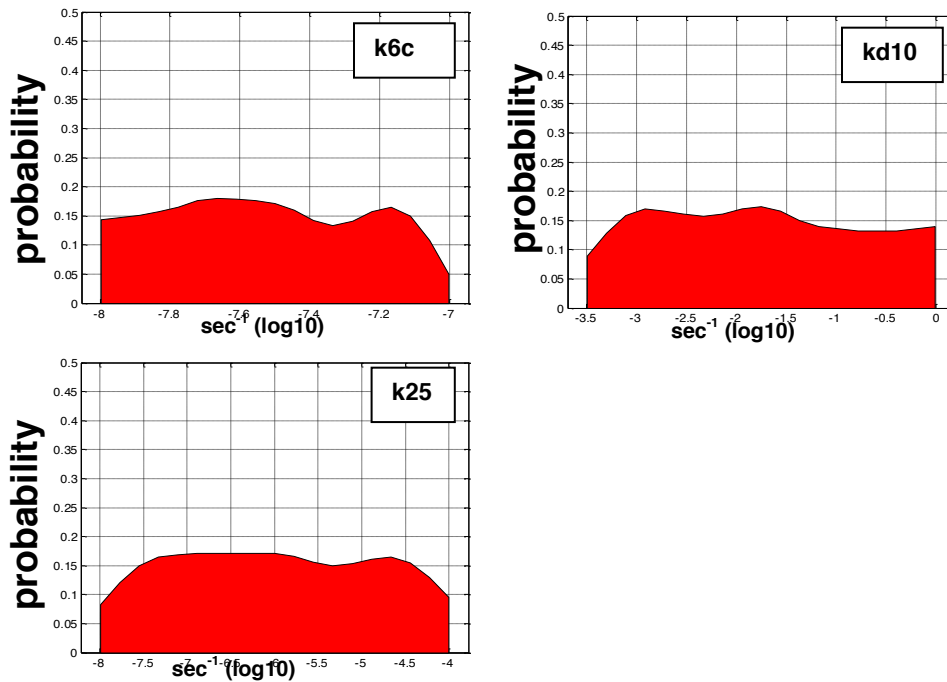
Figure 4.27: Predicted distributions of various model parameters (A) rate constant of internalization of inactive receptors (k6c); (B) rate constant of EGF unbinding from EGFR monomers in endosomes (kd10); (C) rate constant of adaptors binding to phosphorylated EGFR (k25)

Section 4.3.2: Identifying parameters sensitive to specific experimental condition

In this section we describe examples of the application of our approach to study the relationship between parameter values and systems behaviors. In particular, by comparing parameter distributions obtained from fitting two different subsets of experimental data (one set being inclusive of the other set) we can identify parameters important to fitting specifically the experimental data different between the two subsets. For example, we can detect parameters responsible for phosphorylation signal decay by comparing the parameter distributions obtained from the fits to only early dynamics of the responses to fits to the entire time-courses, including early and late responses, which demonstrate the long-term signal decay. The constraints on the model responses imposed by additional experimental data should be reflected in the changes in model's parameter space.

In Section 4.2.1 we described the procedure of fitting the model to the phosphorylation time-courses. In the analysis presented in this section, we first fit the model to the phosphorylation

signaling response of up to 30 minutes of stimulation with only high EGF, and plot the predicted distributions of pEGFR and pERK phosphorylation response at 1.5 hours of stimulation (Figure 4.28, experimental data shown in black).

Model simulations show that fitting the model to early dynamics predicts a wide range of possible responses by 1.5 hours. This suggests that the fits with different parameters assignments can all equally well capture amplitude and dynamics of signaling responses at early times, but have different underlying rates of signal decay, and thus vary in their predictions of the 1.5 hour signaling responses. We find that although the obtained distributions predict a non-zero probability over the experimentally observed values of pEGFR (0.0182+/-0.016 A.U.) and pERK (0.273+/-0.017 A.U.) at 1.5 hours, both distributions are predicting significantly higher average responses (Figure 4.28), suggesting that, according to the model, phosphorylation signal decay measured on pEGFR and pERK signals is not very likely based on the early dynamics of the responses. We then fit the model to the early phosphorylation responses and include the 1.5 hours responses as well. In this way, we impose that the signal on pEGFR and pERK decay by the 1.5 hours of stimulation with EGF(Figure 34, Figure 4.32). We find that the model can fit signal decay dynamics observed in experimental data, and compared to the fits to only early dynamics (three points) predicts the distributions of 1.5 hours behavior with substantially shifted averages (Figure 4.29, Figure 4.30).
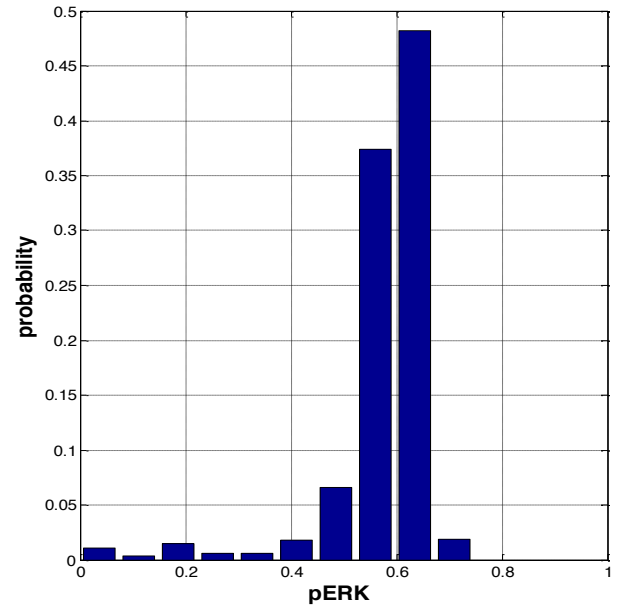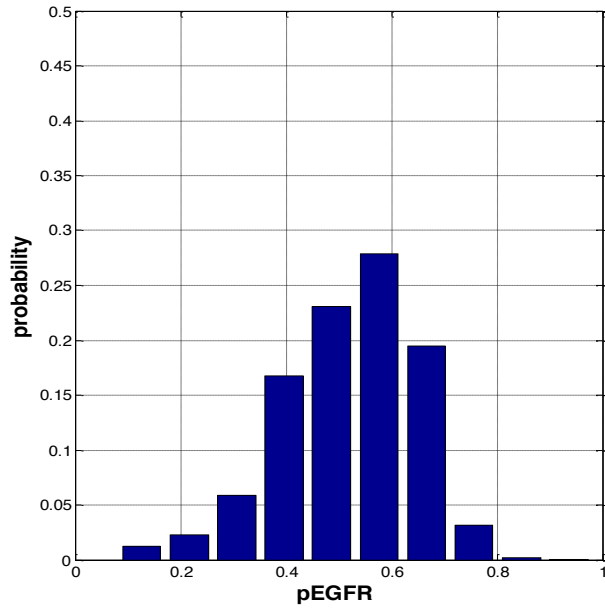
Figure 4.28: Predicted distributions of pERK and pEGFR responses in MCF-10A cells after 1.5 hours of stimulation with 100ng/ml EGF. The model was fitted to the experimental measurements of the corresponding responses in MCF-10A cells up to 30 minutes of simulation with 100ng/ml EGF.
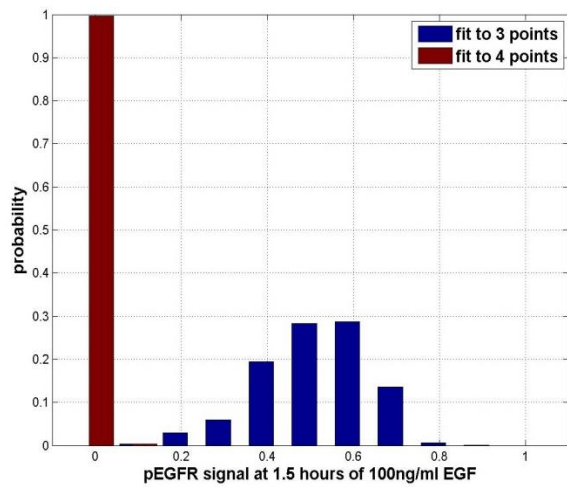
Figure 4.29: Predicted distribution of pEGFR response 1.5 hours after 100ng/ml EGF stimulation in model fit to the phosphorylation time-courses data measured in MCF10A cells.
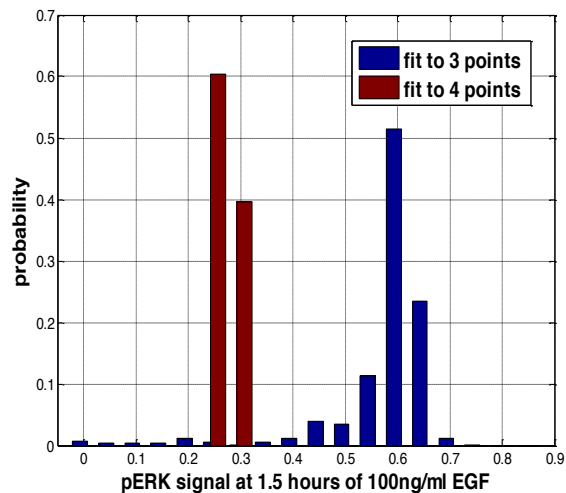
Figure 4.30: Predicted distribution of pERK response 1.5 hours after 100ng/ml EGF stimulation in model fit to the phosphorylation time-courses data measured in MCF10A cells
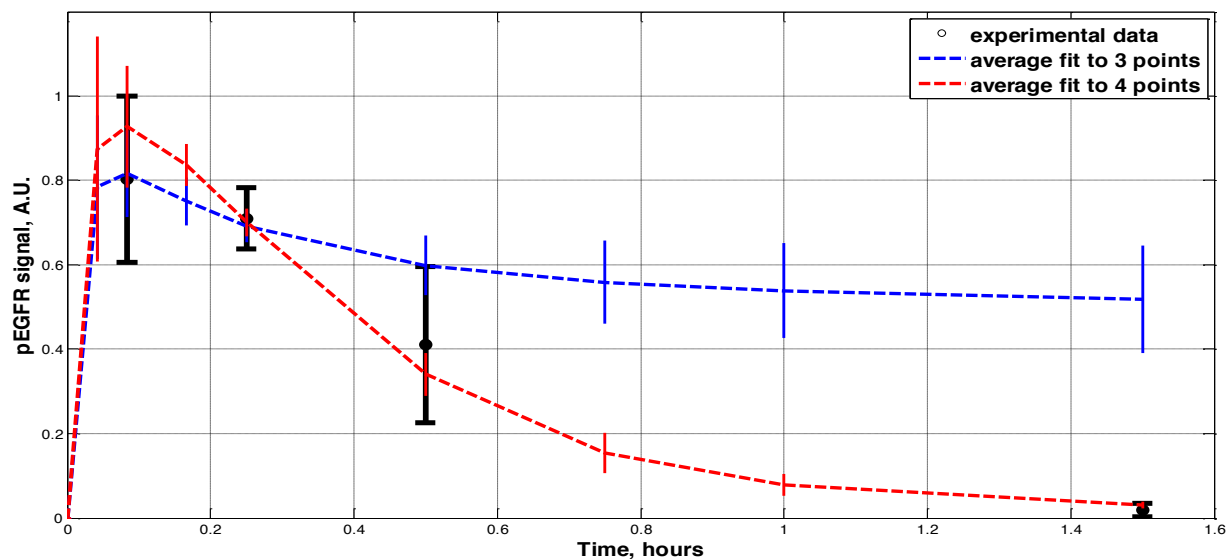
Figure 4.31: Predicted average time-courses of pEGFR response obtained from the model fitted to the early or to the entire phosphorylation responses in MCF10A cells.
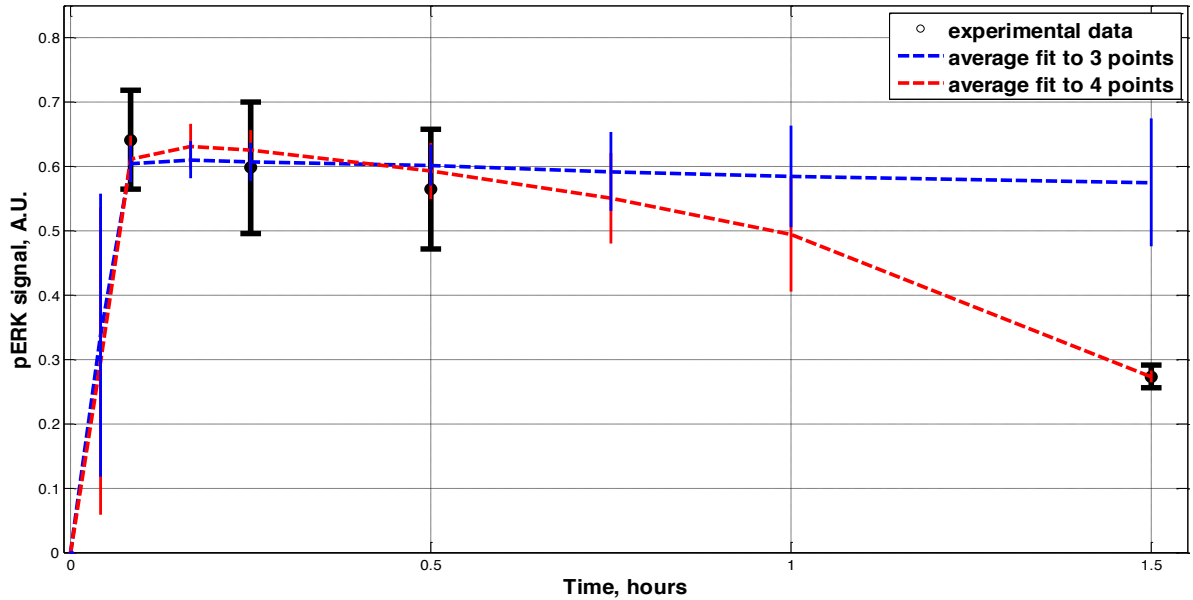
Figure 4.32: Predicted average time-courses of pERK response obtained from the model fitted to the early or to the entire phosphorylation responses in MCF10A cells.

We then apply statistical methods described in Section 3.2.9 to compare parameter distributions and find parameters important to phosphorylation signal decay. We find that the top ranked parameters differing between fitting the two conditions (early responses only or the entire time-course) are mostly related to activity of phosphatases and internalization and degradation of EGFR receptors (Table 3, Table 4).

parameter	description	sum of Kullback-Leibler divergence scores
k60	rate of degradation of EGFR	-1.92
kd123	rate of phosphorylation of EGFR	-1.68
c12	abundance of cPP protein (internalization of EGFR)	-1.28
c280	abundance of endosomal RTK phosphatase	-1.25
c5000	abundance of cytoplasmic RTK phosphatase	-0.76

Table 3: Top ranked parameters related to pEGFR signal decay at 1.5 hours of stimulation with 100ng/ml EGF in MCF-10A cells.

parameter	description	sum of Kullback-Leibler divergence scores
k56	rate of binding of ERK phosphatase	-3.07
kd55	kcat of ERK activation by pMEK	-2.29
kd57	kcat of ERK phosphatase	-0.87
c60	Abundance of ERK phosphatase	-0.27

Table 4: Top ranked parameters related to pERK signal decay at 1.5 hours of stimulation with 100ng/ml EGF in MCF-10A cells

In addition to comparing parameter distributions by the Kullback-Leibler based metric, we compare the averages and the medians of the distributions of the same parameter in the two different conditions. We find that indeed, the Kullback-Leibler criterion may miss distinctions in distributions of some parameters known to be important to fitting phosphorylation signal decay on pEGFR. For example, kcat of endosomal RTK phosphatase was found to be not significantly different in the distributions by the Kullback-Leibler metric (kd95 sum of the Kullback Leibler scores =-0.52), however the differences in the medians suggested that this rate is by approximately one order of magnitude faster in the fits to the entire response dynamics which entails decay of the phosphorylation signal. Abundance of the endosomal RTK phosphatase is predicted to be approximately 1.5 order lower in the fits with phosphorylation signal decay (c280 sum of the Kullback Leibler scores =-1.25). Although endosomal phosphatase can remove phosphorylation signal, it can also prevent receptors from degradation, as phosphorylated receptors are usually degraded faster than the unphosphorylated ones. Therefore, the role of the activity of the endosomal RTK phosphoatase can be ambiguous in respect to phosphorylation signal decay.

Since the rate constant of degradation of receptors was found to be important in accommodating phosphorylation signal decay, we considered system output variable which can serve to estimate the actual strength of degradation in the fit. Specifically, we consider predicted distribution of undegraded EGFR receptors at 1.5 hours of EGF stimulation (Figure 4.33). We assume that the number of receptors remaining undegraded in the long run after high EGF stimulation could reflect the rate of degradation of activated receptors – the higher it were, the more receptors would be degraded in the course of stimulation with EGF, thus less of the receptors remaining by 1.5 hours. We find that in the fits to early dynamics only, receptors degradation is predicted to be not substantial,

with on average 90% of receptors remaining undegraded; however, in the fits with imposed phosphorylation signal decay, most receptors are predicted to be degraded by 1.5 hours, indeed, supporting the predicted importance of receptors degradation process in fitting signal decay.
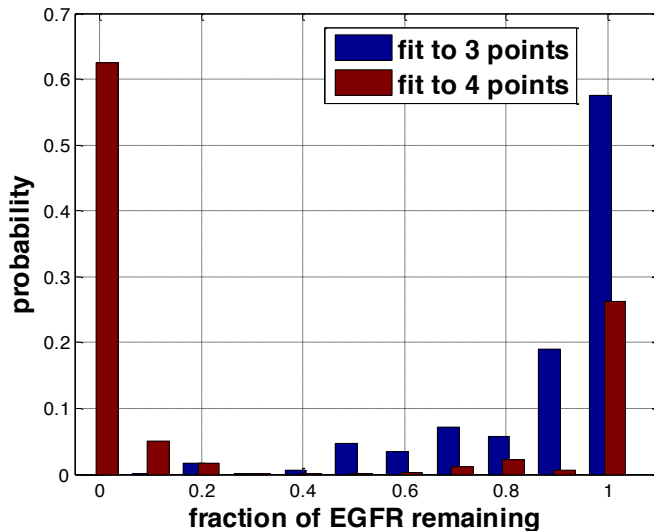
Figure 4.33: Predicted fraction of remaining, un-degraded EGFR receptors at 1.5 hours of stimulation with high dose EGF in MCF-10A cells.

In general, it is a non-trivial undertaking to identify parameters responsible for a specific feature of experimental data, such as, phosphorylation signal decay, as it can be realized through an interplay of multiple mechanisms, namely, action of phosphatases, internalization, and degradation of receptors, as well as ligand depletion. Relative contributions of these processes to signal decay are not known and could be elucidated through studying the properties of parameter space of the model fitted to the experimental data.

In the example presented here, we imposed signal decay on both pEGFR as well as pERK, and therefore, the detected parameters with significantly shifted distributions are relevant to fitting dynamics of both pEGFR and pERK together. However, if we were interested in identifying parameters responsible specifically for signal decay on either one the two targets, we could resimulate our model by including late response of only one target of interest, either pEGFR or pERK, and repeat the analysis.

Section 4.3.3: Identifying correlations between parameters

The generated probabilistic parameter distributions can provide an insight about the overall structure of the model parameter space by suggesting possible correlations between individual parameters. Since the distributions of the model parameter values were derived based on the experimental data, by analyzing correlations between parameter values one could assess possible relationships between various physiological quantities and processes in the *in vivo* system in which the data was measured.

For example, we consider the correlations between parameters relevant to ERK deactivation obtained from fitting the phosphorylation data including signal decay. We find significant negative correlation (Pearson correlation coefficient = -0.7882, p-value <10^{-3}) between the concentration of ERK phosphatase (c60) and the *kcat* of its phosphatase (kd57). Such relationship can arise from the fact that in order to fit pERK signal decay, both e abundance and the rate of activity of the ERK phosphatase can be important, and possibly mutually compensatory. In addition, such information, could serve as a validation for our model by confirming various expected relationships between the model's parameters. Lastly, the correlation coefficient between the two pairs of parameters can be ʾul in quantifying the degree of relatedness between the parameters and determine if the ımeters are significantly or only marginally related, and in what way – positive or negative.
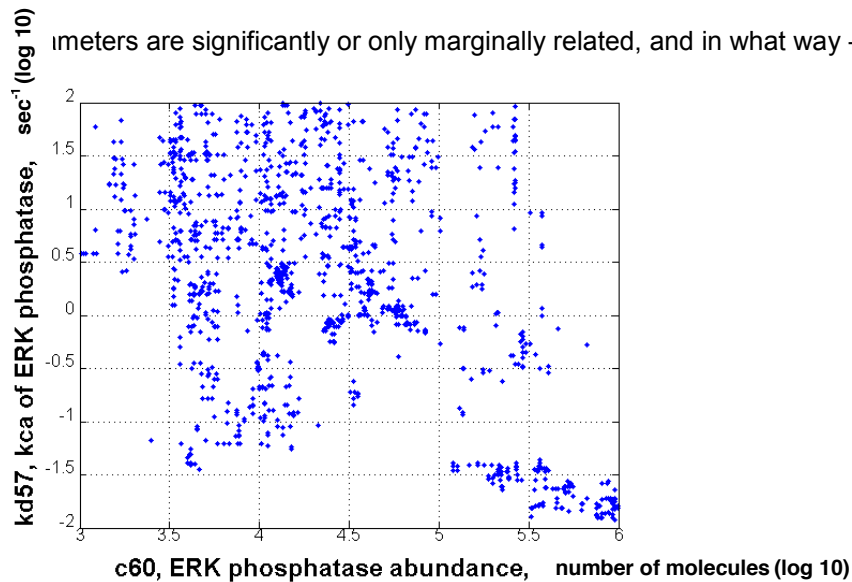
Figure 4.34: Negative correlation between abundance and the rate of activity of ERK phosphatase. Individual points represent coordinates of the sampled by MCMC chains in the space of the two parameters., ERK phosphatase abundance and its rate of activity.

Section 4.4: Long term responses: perfect adaptation

Section 4.4.1: Predicted distributions of long term phosphorylation responses

We fit our model to phosphorylation and total EGFR time-courses measured experimentally at several time-points within three hours of stimulation with constant doses of EGF (described previously in Section 4.2.1). We then predict time-courses of system responses over six hours of continued stimulation with EGF (Figure 4.35) and explore possibility of perfect adaptation in phosphorylation signals in our system (see Section 2.9 for background on perfect adaptation). To reiterate, perfect (or exact) adaptation is a condition in which signal responses at steady state are independent of the constant level of stimulus applied and are approximately same as the pre-stimulus level of signal in the system. Based on the experimental data of pEGFR, 6 hour time point might be a good approximation for system steady state behavior, as most of the pEGFR signaling appears to decay sufficiently several hours prior to that time. By analyzing dependence of steady state levels of pEGFR signals on the EGF dose applied we want to investigate if it is independent of the EGF doses (thus, implying perfect adaptation, by definition), and how this property might be transduced to the level of pERK.
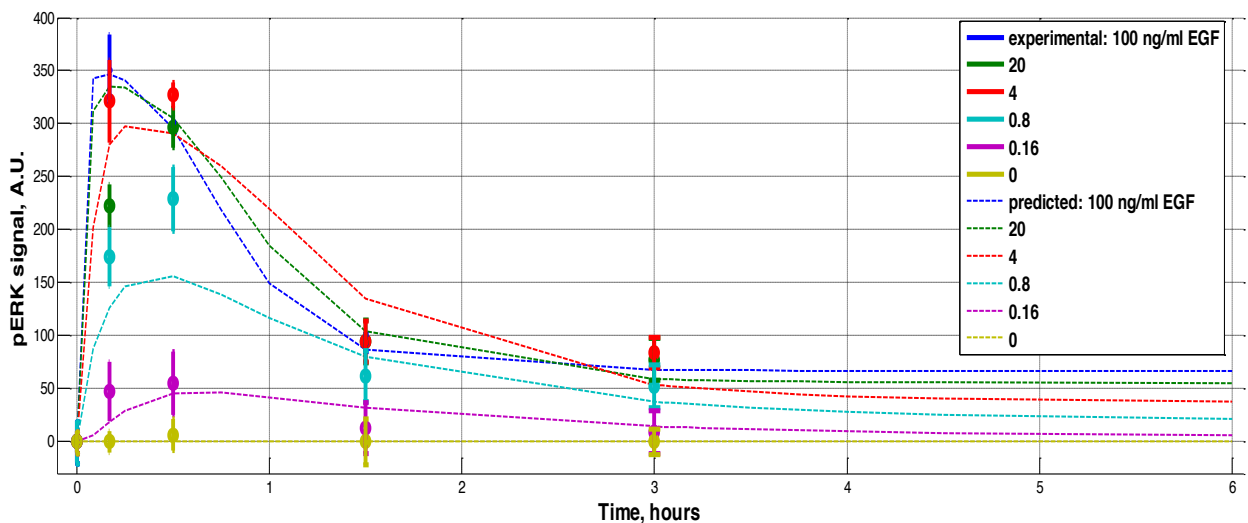
Figure 4.35: A fit to the three hour time-course experimental data predicts 6 hour response.Experimental data is shown in bold.

Since we fit our model only to several experimental time points measured between 0 and 3 hours of EGF stimulation, the predicted fits (individual time-courses of signaling responses of each

target at a given dose of EGF) can vary in their overall dynamics, while being similar at the points being fitted. For example, the predicted phosphorylation fits to high EGF can vary in times when their responses reach the peaks, or speed of the signal decay and behavior at 6 hours of stimulation.

We obtain a set of probabilistic distributions of phosphorylation responses at six hours of constant stimulation with each of the applied EGF doses (Figure 4.36, Figure 4.37).
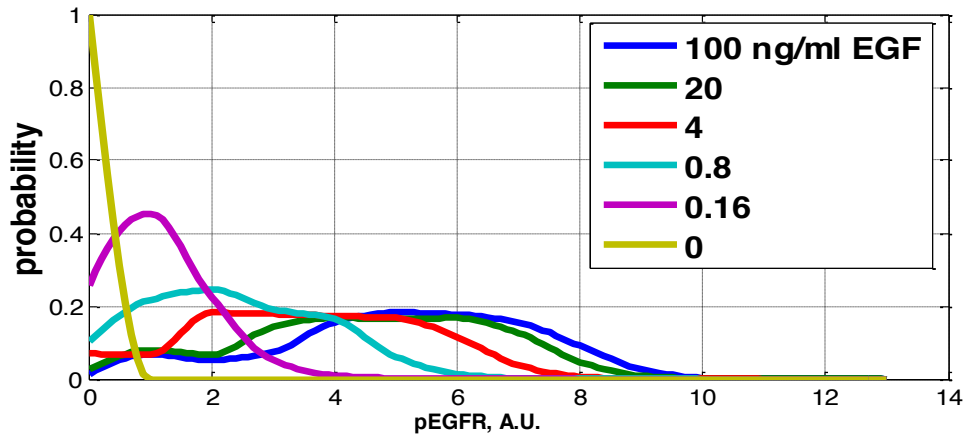
Figure 4.36: Predicted distribution of pEGFR responses after 6 hours of EGF stimulation.
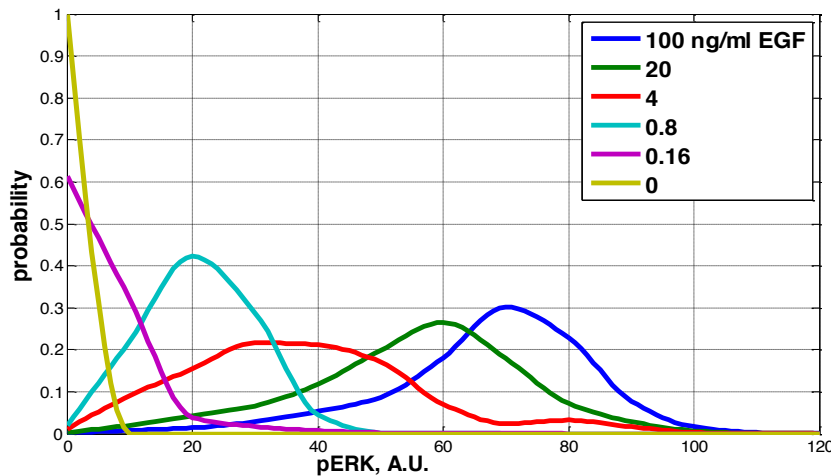
Figure 4.37: Predicted distribution of pERK responses after 6 hours of EGF stimulation.

The predicted distributions on both, pEGFR and pERK, are monotonically dependent on the EGF dose of stimulation, with higher EGF doses applied leading to higher phosphorylation remaining by 6 hours of treatment. Then, if such monotonicity is preserved for a given fit, one might expect that if phosphorylation response at highest EGF has dropped below certain threshold value, then, most likely responses at lower doses are also below that value. We then investigate EGF dose-dependent

134

dynamics of signal between 3 and 6 hours of stimulation: do signals further decay between 3 and 6 hours, or are they at steady state by 3 hours, and thus do not change between 3 and 6 hours. For this, we examine predicted distributions of phosphorylation signals at earlier time (3 hour) and check if the later time (6 hour) distributions have shifted to lower phosphorylation levels. The data shows that at EGF up to 20 ng/ml EGF, the distributions shift to lower levels by 6 hours, whereas, at 100ng/ml the distribution remains similar between 3 and 6 hours (Figure 4.38, Figure 4.39), therefore, suggesting that at higher EGF doses signal appears to reach steady state between 3 and 6 hours, whereas at lower EGF doses the signal continues to decay.
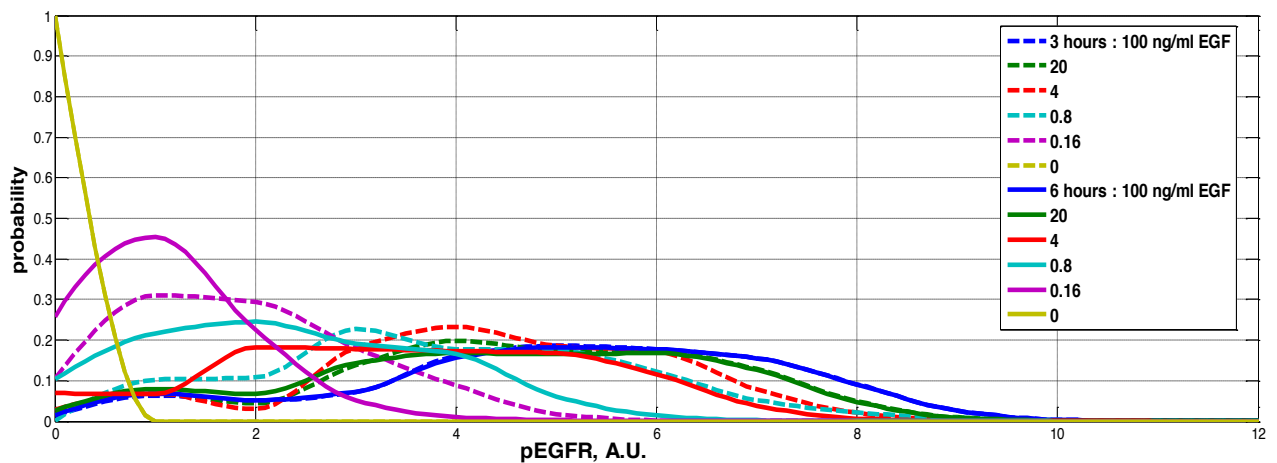
Figure 4.38: Comparing predicted distributions of pEGFR responses after 3 and 6 hours of EGF stimulation.
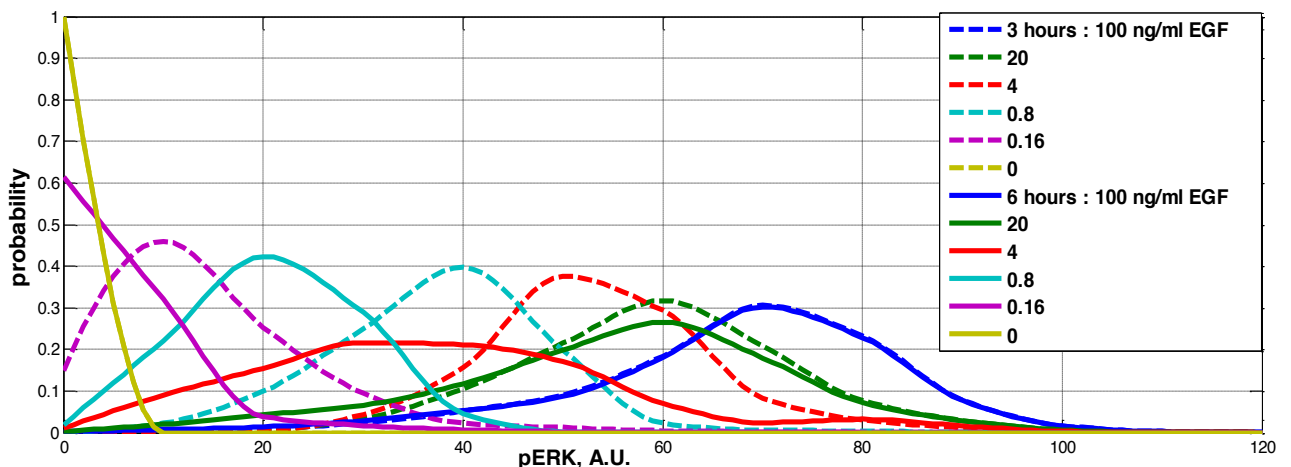
Figure 4.39: Comparing predicted distributions of pERK responses after 3 and at 6 hours of EGF stimulation.

To quantify degree of signal decay across many fits, we consider ratio of 6 hour response at high EGF (100ng/ml) to the peak of the high EGF response in the fitted experimental data (in this data set Max(pERK)=351+/-32 A.U., Max(pEGFR)=124+/-2 A.U.). Assuming the monotonic dependence of phosphorylation signals at 6 hours on the EGF doses, a low ratio of 6 hour response at high EGF to the signal peak would suggest that at most EGF doses signal returns close to its pre-stimulus level. The distributions show that by 6 hours of stimulation with highest dose of EGF, pEGFR responses on average have decayed down to 5% of pEGFR maximum, and for pERK the responses, on average, have decayed down to 20% of the pERK maximum (Figure 4.40). Therefore, pEGFR responses return closer to pre-stimulus level then do the pERK responses; such effect might be explained by MAPK cascade signal amplification from pEGFR to pERK. The experimental data shows that at various EGF doses most pEGFR signals adapt within the first three hours, and this is captured in our model predictions; pERK and pAKT adapt slower than pEGFR does, approaching closer to pre-stimulus levels only by 6 hours of stimulation.
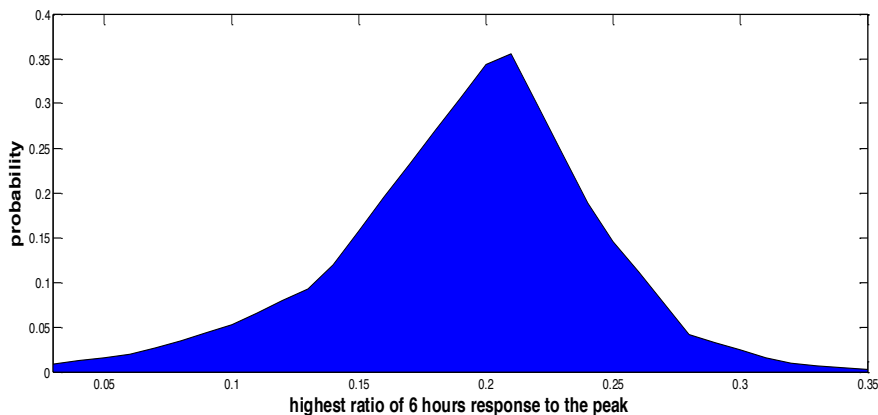
Figure 4.40: Predicted distribution of the ratio of the pERK 6 hour response to the pERK signal peak.

We then consider the subpopulation of fits in which pERK signal has a more pronounced decay down to 5% of the pERK maximum, which is predicted to happen in approximately 1.63% of all fits (Figure 4.40). Plotting average time-course of pERK responses of all predicted fits compared to the average of the subpopulation with 5% signal decay, shows that in the subpopulation with stronger signal decay, independently of the dose of EGF applied, most signals return to pre-stimulus level by 6 hours (Figure 4.41).
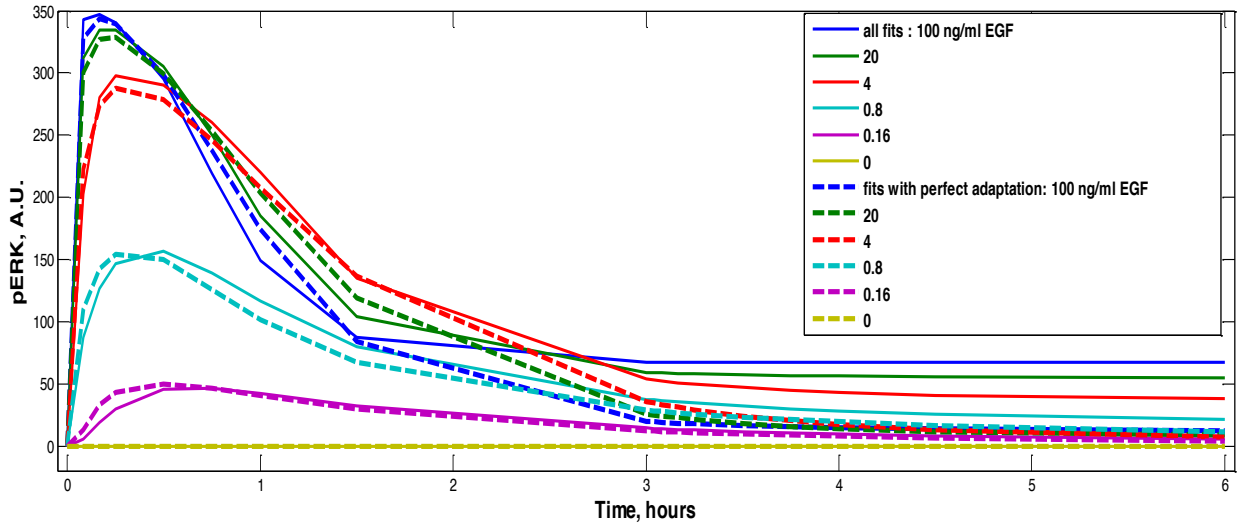
Figure 4.41: Average time-courses of pERK from the overall distribution of fits and the subpopulation of fits with a stronger signal decay by 6 hours of EGF stimulation.

We then plot average dose responses around signal peak (15minutes) as well as at long run (3 and 6 hours) of the two populations of fits (Figure 4.42). We find that the average dose responses at the peak signal agree in the two populations, and that the 3 and 6 hours dose response for both populations look significantly flatter than the original peak dose response, suggesting that at long term the responses do not show such noticeable dependence on the EGF doses, at least compared to dependence observed in the peak responses. Moreover, the average dose response in the subpopulation of fits with 5% degree of signal decay appears flatter than the average response of the overall distribution and have a more clearer trend for signal decay between 3 and 6 hours across all EGF doses. Hence, we conclude that the 5% signal decay fits have adaptation appearing more similar to perfect adaptation with signal response at steady state being independent of the constant level of stimulus applied and being similar to the pre-stimulus level of signal in the system.

137
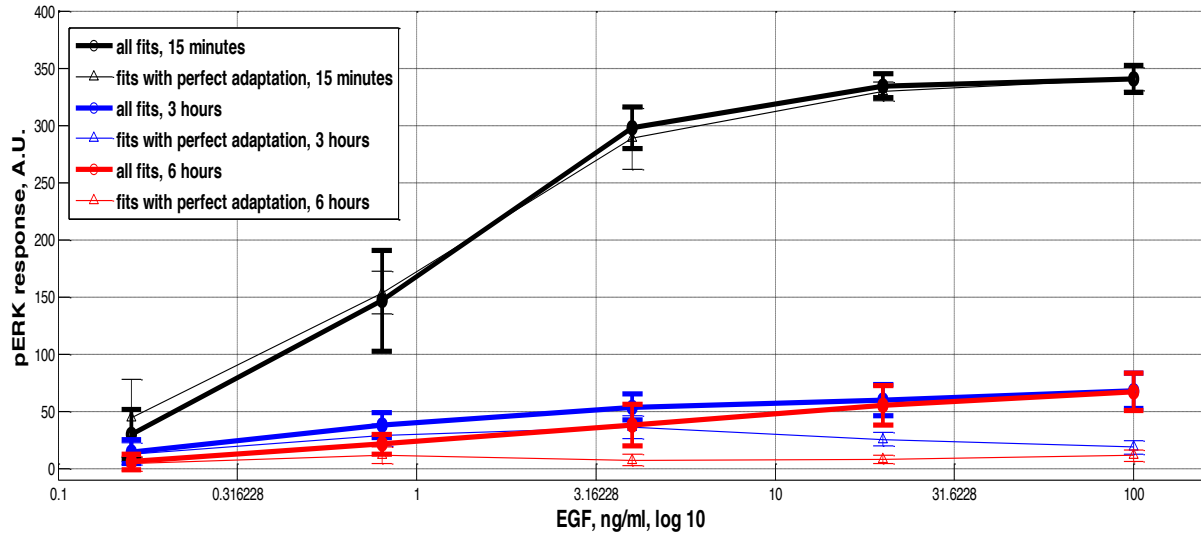
Figure 4.42: Average dose responses of the overall population of fits and the subpopulation of fits with perfect adaptation in pERK signal (<5% decay from the signal peak).

To find parameters affecting adaptation on pERK, we investigate shifts in parameter distributions between the two populations of fits by comparing averages of the parameter distributions from populations of all fits versus from population of the fits with the 5% signal decay. We find that the fits where pERK adapts better are predicted to have less EGFR receptors, with slower rate of EGFR production and higher rate of degradation of active receptors, and slower rate of activation of pERK by pEGFR, faster rate of dephosphorylation of ERK. This result suggests that for pERK signal adaptation, not only parameters directly regulating pERK activity are relevant, but also, parameters determining long-run equilibrium of number of EGFR receptors are important. In addition, rate of EGF binding to dimers and receptors dimerization rate are found to be different in the two subpopulations, suggesting that other processes, at the ligand-receptor interaction can also affect pERK long-term dynamics. The nature of the processes which have shifted thus points to the importance of pEGFR steady state levels to perfect adaptation on pERK, suggesting that a better perfect adaptation on pERK might be achieved if lower pEGFR steady state levels are maintained.

Table 5 : Parameters with most difference in the means of their predicted distributions in fits with stronger pEGFR adaptation (<5% from the signal peak).

Rank	parameter	mean of the entire distribution (log10)	mean of subset of fits with pERK decay at 6 hours below 5% of the maximum (log10)	difference in means (log10)
1	'k31' (EGF binding to EGFR dimers)	-8.91	-7.12	-1.79
2	'kd2' (undimerization of receptors)	0.17	-1.62	1.79
3	'k2'	-4.85	-5.90	1.06
4	'k77'(EGFR production)	0.51	-0.45	0.96
5	'k50' (pEGFR activates ERK)	-4.86	-3.92	-0.94
6	'kd91'	-3.98	-3.13	-0.85
7	'k60' (rate of pEGFR degradation)	-3.10	-2.38	-0.72
8	'c8'	4.32	3.61	0.71
9	'c531' (EGFR abundance)	4.72	4.25	0.48

The predicted distributions of pERK response suggest that although the predicted time-courses are similar at early signaling dynamics to which they are fitted to, they vary in their predicted response dynamics in the long run. As a metric of degree of adaptation in different fits we use a ratio of signal response at 6 hours to the peak of that responses across all doses in that fit. Since all predicted time-courses are fitted to the same experimental data point including the peak of experimentally measured responses, using such statistic provides allows comparison of property of perfect adaptation across different fits. In general, property of perfect adaptation implies two aspects: first, at steady state the responses of signal should be similar, and independent of the dose of stimulus applied, and second, the steady state responses should be similar to the pre-stimulus levels. We have previously observed an overall monotonicity in relationship of steady state signal responses and EGF doses. Therefore, the ratio of high EGF response at 6 hours to the peak of the response at high EGF is an upper bound on the steady state signal across multiple doses and it can suggest if the system has returned close to the pre-stimulus level: lower this ratio, means lower most responses across all EGF doses should be in the long run.

To investigate perfect adaptation by degree of independence of the long-term responses of EGF doses, we consider flatness of the dose response at 6 hours. We quantify the flatness by taking a coefficient of variation (CV) in the 6 hours responses in each fit. Such metric allows to account for possibly non-monotonicity in some regions along dose responses, as well as for the fact that different fits in long run can adapt to different signal levels. For example, Figure 4.43 and Figure 4.44 show two fits with different degrees of adaptation; fits with flatter dose response curve at 6 hours would have lower coefficient of variation metric. We then plot the distribution of the CV metric across all fits (Figure 4.45) and observe that most of the fits are predicted to adapt imperfectly, with CV on average being 0.7, however, some non-zero probability over very low CV values (0.04) suggests existence of fits with a more perfect adaptation, similar to example shown in Figure 4.44.
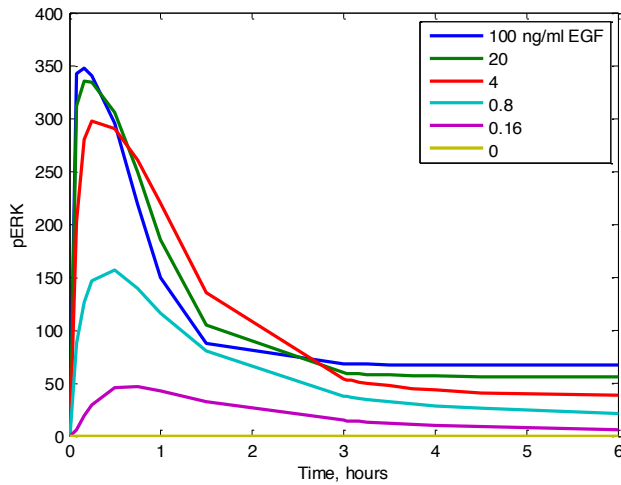
Figure 4.43: Example of a fit without perfect adaptation in pERK responses.
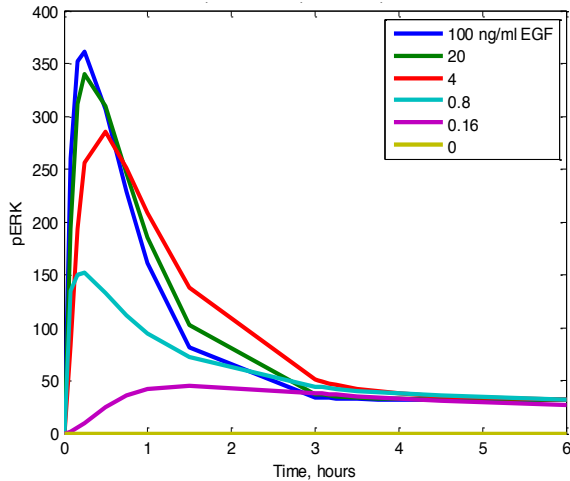
Figure 4.44: Example of a fit with perfect adaptation in pERK responses.
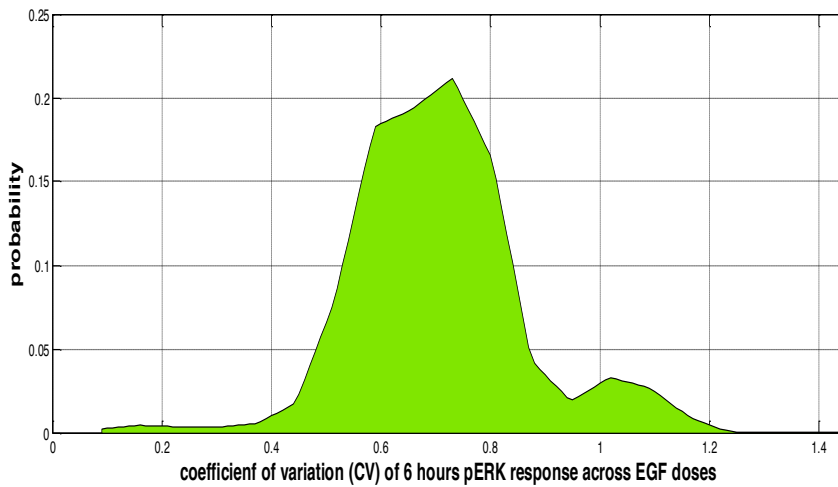
Figure 4.45: Predicted distribution of perfect adaptation on pERK across model fits.

We then look at correlation of the CV perfect adaptation metric and model parameters, thus trying to identify what parameters can affect flatness of the long-term dose response on pERK. We find that higher the ERK abundance, flatter the pERK dose response (Pearson correlation coefficient = -0.67, p-value $<10^{-4}$); whereas, faster ERK activation by pEGFR is correlated with steeper dose response on pERK (Pearson correlation coefficient = 0.41, p-value $<10^{-4}$). Interestingly, the parameters relevant to maintenance of steady state levels of EGFR were found to be correlated with flatness of pERK dose response. In particular, significant correlation between flatness of pERK dose response and rate of receptors production (Pearson correlation coefficient = -0.46, p-value $<10^{-4}$), basal receptors internalization (Pearson correlation coefficient = 0.42, p-value $<10^{-4}$) and basal

receptors degradation (Pearson correlation coefficient = 0.38, p-value <10^{-4}) suggest that faster receptors production and their slower turnover at steady state would produce flatter pERK dose responses, more similar to EGF dose-independent adaptation.

We found significant correlation between flatness of predicted dose responses on pEGFR at 6 hours and that of pERK (Pearson correlation coefficient=0.9, p-value < 10^{-3}), suggesting that most likely perfect adaptation on pERK is contingent on perfect adaptation on pEGFR. However, perfect adaptation on pEGFR barely passing a defined threshold of signal decay, for example, 5% of maximum signal peak, due to signal amplification in the kinase cascade, could result in no observed perfect adaptation on the downstream pERK.

Due to limitations of our resolution of our experimental assay, small differences in pEGFR usually might not be experimentally distinguishable, and thus appear similar in assay readouts. However, differences even in hundreds of activated pEGFR molecules through amplification in kinase cascade can produce significantly different responses at pERK. Hence, pEGFR responses at long run, even though experimentally appearing perfectly adapted, might differ and be amplified at pERK, and thus pERK might exceed its threshold of perfect adaptation.

Overall, the experimental data of long-term phosphorylation measurements of pEGFR and pAKT due to experimental errors does not allow to determine if the signals have adapted perfectly within 3 hours of constant EGF stimulation. The measurement error at 3 hours is on average 9 A.U. on pEGFR and 20 A.U. on pERK, and thus it is not feasible to distinguish how dissimilar are the responses at the 3 hours. As the distributions we predict rely on the fits to experimental data, given the magnitude of experimental error we cannot conclude if the difference between the averages of the predicted 6 hours distributions (Figure 4.36, Figure 4.37) is not significant to claim or refute perfect adaptation.

However, our model can help estimate degree of predicted adaptation and its various features, such as steady state levels of responses or EGF-independence of steady state values across EGF doses. In addition, our approach can help identify parameters and processes which can affect adaptation. Lastly, it suggests that pERK perfect adaptation might be contingent on the pEGFR

perfect adaptation, and in Section 4.7.1 we present analytical derivation of mechanisms of propagation of perfect adaptation from receptors to downstream signaling targets.

Section 4.4.2: On experimental validation of perfect adaptation

To experimentally validate the predicted perfect adaptation in ErbB signaling system, pEGFR, pERK and pAKT dynamic responses can be measured for several hours of stimulation with constant EGF signal. By comparing flatness of the long-term dose responses of each individual target we can then assess degree of perfect adaptation.

Experimental data, namely, phosphorylation time-courses up to three hours of EGF, suggest that in response to prolonged stimulations with constant EGF doses, pEGFR returns to its pre-stimulus level independently of the EGF dose by 1.5 hours of stimulation. Experimental validations also show that within the given experimental error in measurements, pEGFR appears to adapt perfectly across entire range of EGF doses we tested and that pAKT also is most likely adapting perfectly across this range of EGF doses, yet at slower rates. However, by three hours of EGF treatment pERK is observed to be adapted only within a 20% of its peak value and appears to stabilize at higher than the initial pre-stimulus steady state levels and the model suggests that pERK signal at high EGF is not likely to decay further beyond 3 hours.

Since pERK responses might take longer time to decay due to signal amplification in the kinase cascade, the experiments should consider monitoring pERK dose responses at longer times even after pEGFR perfect adaptation is detected. In practice, however, it might be infeasible to monitor pERK dose responses over extended hours, and hence, a trend suggesting a decreasing over time dose responses in pERK should be sufficient to indicate perfect adaptation on pERK. For example, pEGFR, pERK, pAKT can be measured within 6 hours of persistent EGF stimulation; even though pEGFR is observed to decay perfectly already by 1.5 hours. Interestingly, this observation also points out a feature that effect of EGF in a cell can be prolonged through extended duration of pERK and pAKT signaling, the downstream targets of pEGFR, even if pEGFR signal duration is shorter.

In the experiments measuring perfect adaptation, it is important to separate cell signaling responses dynamics from dynamics of changes in ligand concentrations Ligand depletion can alter

dynamic profile of the stimulus and contribute to signaling responses adaptation. Hence, one needs to maintain constant level of stimulus in order to establish existence of machinery for perfect adaptation in a signaling system. It is conceivable that at low absolute levels of ligand, signal responses to different ligand doses over time might return to pre-stimulus level due to complete ligand depleted from the media, independently of the initial dose of the applied ligand. Such scenario does not technically correspond to the definition of perfect adaptation, which entails stimulation with constant dose of ligand over time. Hence, stimulus should be kept constant to decouple stimulus dynamics from response dynamics and thus to better understand specifically functioning of cell machinery itself. Interestingly, the importance of separating stimulus dynamics from responses dynamics has also been brought up in recent work in olfactory receptors sensory system in drosophila (Martelli et al., 2013).

It will be interesting to explore what molecular factors in signaling cascades are responsible for buffering the upstream pEGFR signal and so able to preserve or distort the property of perfect adaptation on pERK. For example, among other factors, scaffolding proteins in the signaling cascade between pEGFR and pERK might alter the signaling response dynamics (Levchenko et al., 2000) and thus influence perfect adaptation. Two-step phosphorylation required for pERK activation might also contribute to distorting perfect adaptation between pEGFR and pERK. In addition, the distinction between pERK and pAKT perfect adaptation properties might be due to differences in signaling cascade components leading to either pERK to pAKT activation, for instance, pERK is activated through the ultrasensitive Raf-MAPK kinase cascade module, whereas pAKT is activated through a series of activating reactions involving PIP3 and PI3K. Applying varying inhibitor perturbations and measuring long term responses of the target should help elucidate signaling cascade mechanisms important for realizing perfect adaptation downstream of receptors level.

Overall, various biological functions can impose different evolutionary constraints on signaling through pEGFR, pERK, and pAKT can thus can lead to the observed distinct rates of perfect adaptation of these targets. For example, we observe that pEGFR signal adapts faster than its downstream targets, pERK and pAKT. It might be advantageous for cells to have a more rapid downregulation of signaling at the upstream, receptors level, as it would lead to a better control of the

signals propagated to the multiple downstream phosphorylation targets. Such scenario might be in line with the role of ErbB receptors in cell signaling - to detect extracellular stimulus and transmit the signal into the cell, leaving regulation of downstream targets up to other components of cell signaling.

Recently, a question has been raised if perfect adaptation in cell signaling responses is an intrinsic feature of individual single cells, or if it is an emerging property of cell populations (Marquez-Lago & Leier, 2011). Generally, differences in regulation of signal decay across individual cells can arise from single cell variability in gene expression, and thus, might be manifested in different timings of signal adaptation. However, our model predicts that signals perfectly adapt with high probability within several hours of stimulation, and hence, adaptation might be an essential feature of the ErbB signaling. A recent theoretical work has suggested that perfect adaptation is an emerging property of cell populations, and that individual cells do not necessarily adapt their signals perfectly (Marquez-Lago & Leier, 2011). In case we fit our model using population averages data, we find that, in agreement with the above-mentioned study, ensemble signaling behavior exhibits perfect adaptation. However, contrary to the mentioned study claim, our simulations predict that even though the distribution would predict a range of single cell behaviors including ones with saturation, and no perfect adaptation, cases where individual cell responses exhibit perfect adaptation are also predicted to exist.

Section 4.4.3: Model simulations suggest that phospho-ERK adapts perfectly to constant stimulation with EGF, but not constant levels of pEGFR

We set out to answer if the mechanism of pERK adaptation to EGF is contingent on pEGFR adaptation to EGF, or if it can be realized by the downstream signaling components independently of perfect adaptation on pEGFR.

To test this hypothesis, we fit the model to the three hour phosphorylation and total EGFR time-courses and simulate system's responses on pERK if pEGFR is sustained, and not decaying, in response to stimulation with constant dose of EGF (Figure 4.46). Such pEGFR response dynamics is biologically possible in case the ErbB receptors contain an over-activating mutation. We then plot the distribution of predicted ratios of pERK response at 6 hours to the peak of that response, a statistic to assess level of pERK signal decay and adaptation. We find that if pEGFR is kept constant, most

145

pERK responses do not decay and also remain constant at the levels, dependent on steady state pEGFR levels. Hence, the downstream of pEGFR part of a signaling cascade leading to pERK activation is not likely have its own mechanism of perfect adaptation and relies on perfect adaptation realized on ErbB receptors.
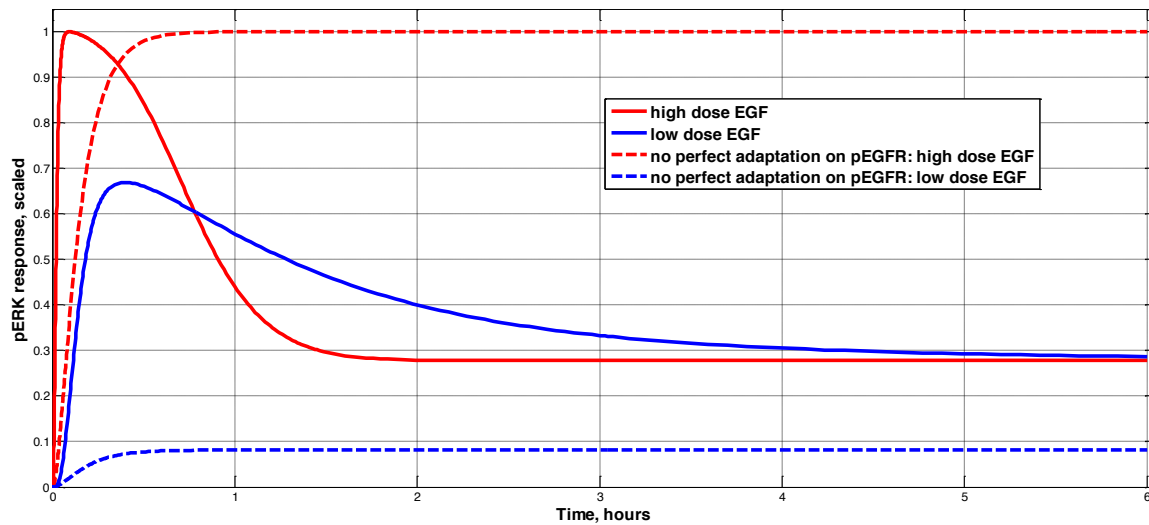
Figure 4.46: Model simulation of pERK responses in fits with and without perfect adaptation on pEGFR.Shown are the simulated pERK responses to stimulation with low (blue) and high (red) dose of EGF. pERK responses shown in dashed line are obtained by keeping pEGFR stimulation constant over time.

Recent experimental work on optogenetic stimulation of the isolated Ras/ERK module suggests that most likely the signaling cascade components which attenuate pERK responses are outside of the Ras/ERK pathways (Toettcher et al., 2013). This observation is consistent with our hypothesis that downstream signaling cascade does not possess its own mechanism of perfect adaptation, and that pEGFR signaling downregulation is a key factor enabling perfect adaptation on pERK.

As a next step it might be interesting to investigate what molecular mechanisms downstream of the receptors level can realize perfect adaptation on pERK. For example, can a negative feedback loop (NFL) from pERK to an upstream signaling cascade component help to implement perfect adaptation mechanism on pERK in response to sustained stimulations at pEGFR? Experimental evidence suggests that pERK can phosphorylate and inhibit activity of some of the upstream signaling cascade components, such as Sos, and Raf (Chen et al., 1996; Sturm et al., 2010), therefore existence of pERK negative feedback loop is a biologically realistic scenario.

146

Interestingly, the overall effect of a NFL on pERK perfect adaptation can be parameter-depend. For instance, the time of recovery of the NFL target from the inhibition by the NFL can be important in realizing perfect adaptation at pERK in the long run. In particular, reversing the effects of the negative feedback loop can be relatively fast, if it relies on, for example, dephosphorylation, or other post-translational modifications of comparatively short-time scale. In general, rapidly reversible protein phosphorylation might not be able to maintain cell memory of previous stimuli. As such, NFL acting through inhibiting phosphorylation might not be able to decay pERK signaling in response to sustained pEGFR stimulation: in the long run, an equilibrium between steady state levels of pEGFR and pERK activation can be established, with sustained activation of pEGFR leading to sustained activation of pERK, and the NFL mostly determining the balance between pEGFR and pERK, and not realizing the pERK decay. Yet, if recovery of the effects of NFL on its target is slow, the NFL can actually insure perfect adaptation on pERK to sustained stimulation on pEGFR. Such scenario might occur if the NFL acts through, for example, degradation of its target, and thus recovery of the NFL effects requires de novo synthesis of the target. Interestingly, the latter case pERK signaling component might be desensitized to changes in the pEGFR signaling until the activity of the NFL target is restored.

Section 4.4.4: Predicted distribution of the number of remaining EGFR receptors in the long run of EGF treatment

As described in Section 4.2, we fit our model to time-courses of fractions of initial EGFR abundances in response to stimulation with different doses of EGF. Although, an average total number of EGFR expressed by cells can vary from day to day due to variations in cell handling, we cannot experimentally measure exact number of EGFR present in the cells on every day we collect the data. However, we have experimental estimate a range of EGFR abundances in our cells taken on one day. Therefore, in our model simulations we do not fix the initial number of EGFR, but impose a constraint on a possible range of its values using our experimental estimates and, in addition, expanding the range to allow initial levels of tEGFR to vary within approximately 1.5 orders of magnitude, $10^{4.4}$ -$10^{5.9}$.

As a result, by fitting the model to the data we obtain a probabilistic distribution of initial number of EGFR in the MCF-10A cells, with mean of 198,000 receptors (Figure 4.47). We can also predict what absolute amount of receptors would be present in the system after several hours of stimulation with a given dose of EGF. However, given flexibility in initial abundances of tEGFR in the fits, it might be more informative to predict fraction of initial number of EGFR remains in the cells after several hours of stimulation with various constant doses of EGF.
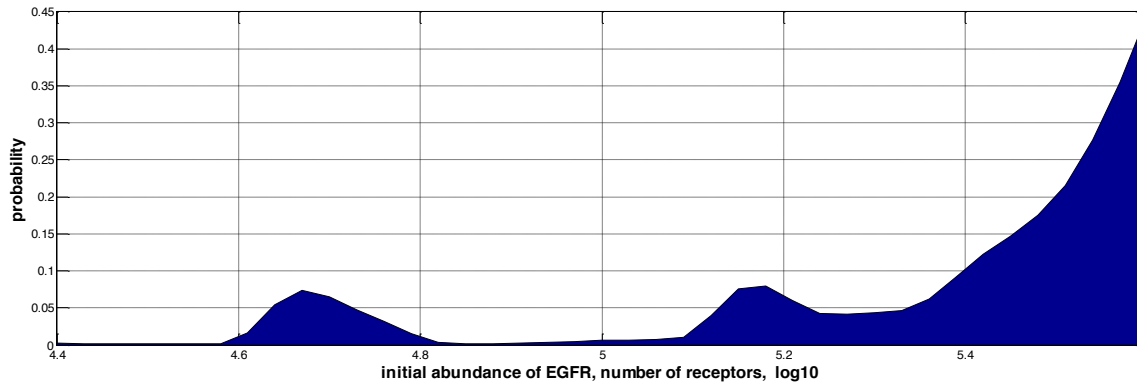
Figure 4.47: Predicted distribution of the initial abundances of EGFR.

For each fit we can then obtain a dose response of tEGFR after 3 hours of stimulation with a range of EGF doses, and plot predicted distribution of the dose response across different fits. The predicted dose response of total EGFR remaining after three hours of constant EGF stimulation shows negative correlation with EGF dose (Figure 4.48). This result agrees with our experimental measurements.
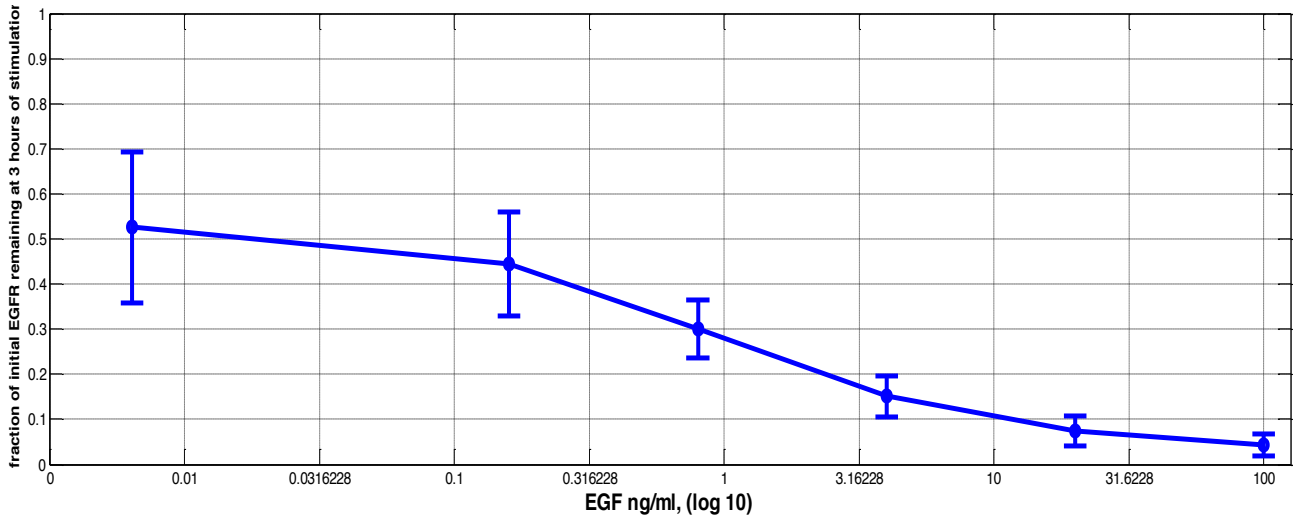
Figure 4.48: Predicted dose response of the fraction of EGFR remaining in the cell after 3 hours of stimulation with EGF.
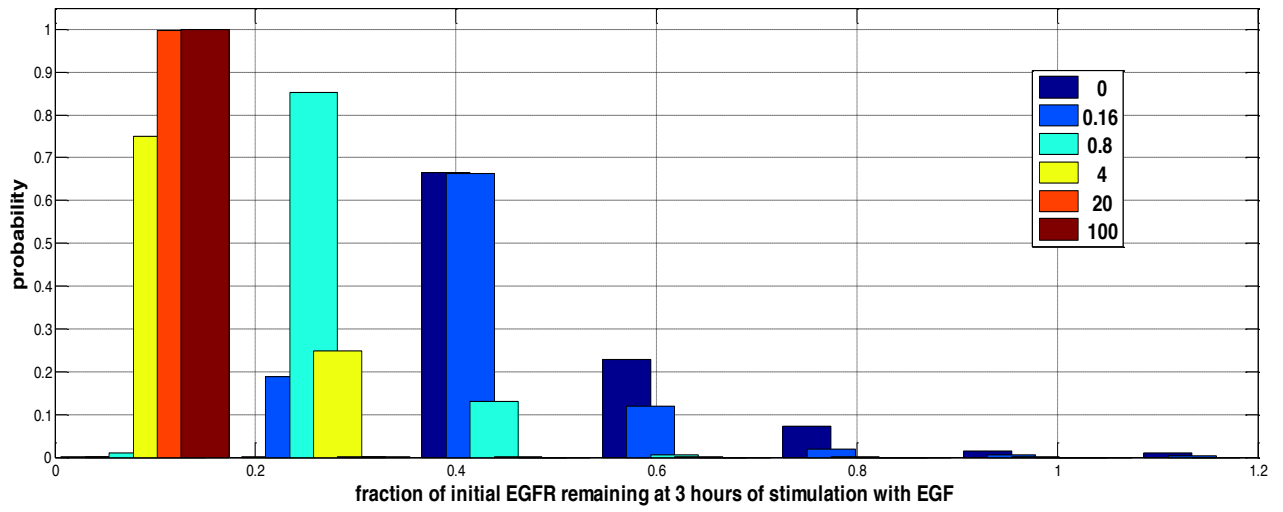
Figure 4.49: Predicted distribution of the EGFR abundance remaining at 3 hours of stimulation with EGF
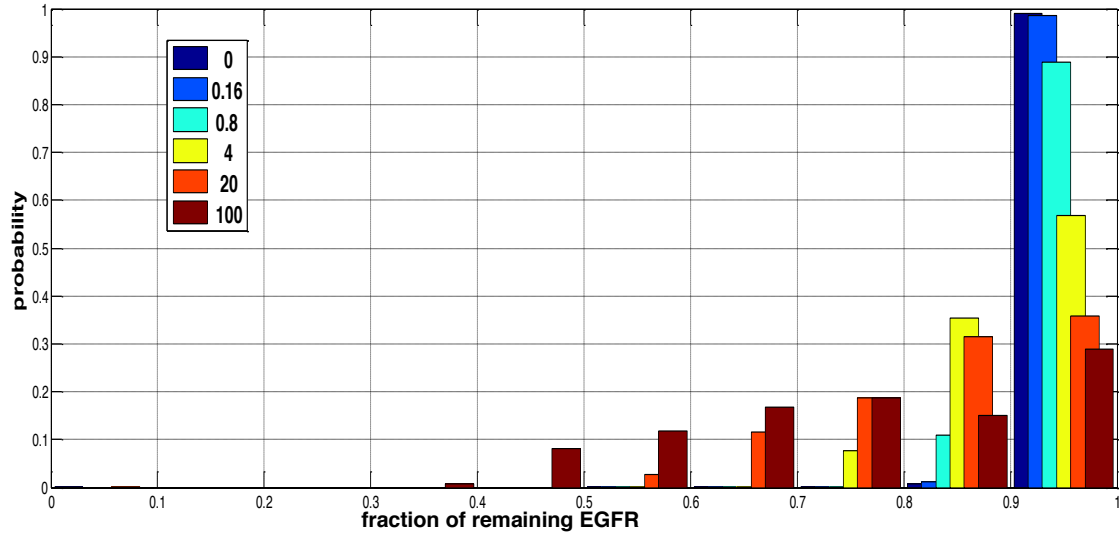
Figure 4.50: Predicted distribution of the fraction of EGFR receptors localized to the cell surface by three hours of stimulation with EGF.
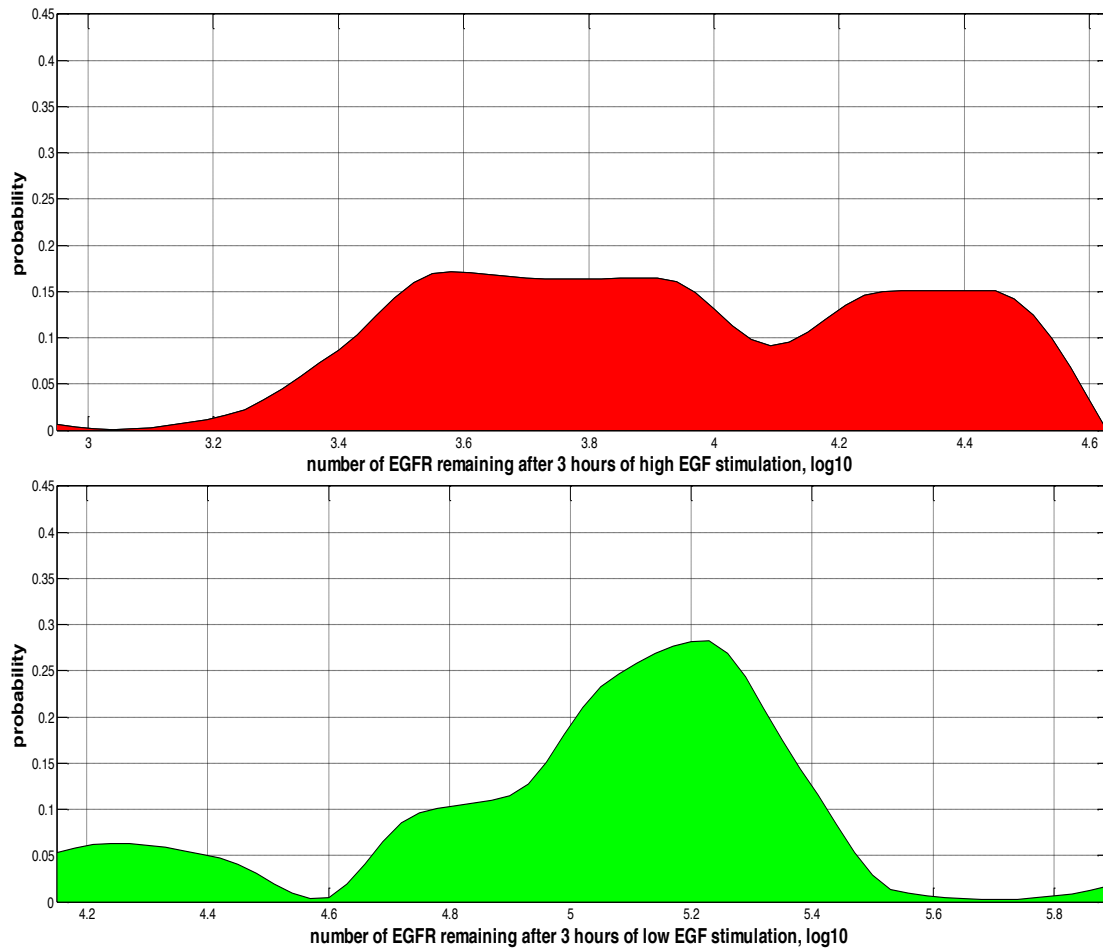
Figure 4.51: Predicted distributions of the number of EGFR receptors remaining after three hours of stimulation with high or low EGF doses.

150

Along with the total number of remaining receptors, we can predict various attributes of the remaining receptors population and quantify relative abundances of various types of receptors in each fit. For example, we can predict cellular localization of the receptors, specifically, how many of them are on the surface versus in cell interior in response to stimulation with various doses of EGF at extended times. We find that after stimulation with high EGF doses, most of the receptors are removed from the cell surface, although, on average, about 70% of remaining receptors remain on the surface, possibly as a result of a dynamic equilibrium by which residual receptors are being activated by the ligand, yet the activation is low enough to trigger receptors removal from the cell surface. Whereas, stimulations with various lower doses of EGF show that higher fraction of receptors remains on the surface (Figure 4.50). The threshold of EGF doses at which most of receptors are internalized occurs approximately between 0.8 - 4 ng/ml EGF, an estimate which agrees with previously reported observation that EGFR internalization is being threshold-controlled around that range of EGF doses (Sigismund et al., 2013).

In addition, our model allows to predict relative distribution of remaining receptors among different species, and so can answer, for example, what fraction of receptors in the long run is in monomers or in dimers, or how many receptors are phosphorylated, and how these fractions change depending on ligand dose or time of observation. Such results could be applied to drug design methods, as the insight on state of the cell surface receptors and EGF-occupancy in condition of chronic EGF presence could inform choice of EGFR inhibitors with various kinetic properties. Overall, abundance and state of receptors in the cells prior to ligand stimulation could be important as they can determine cell signaling responses to EGF (Macdonald & Pike, 2008).

Section 4.5: EGF depletion, Cell density, and EGFR expression

Section 4.5.1: Model predicts significant depletion of EGF from the media at low EGF doses and suggests alternative cell density to prevent EGF depletion

Ligand depletion can play an important role in cell signaling. For example, Epo receptors can sense a wide range of Epo doses due to their ability to internalize Epo ligand effectively from cell media; in addition, Epo depletion contributes to EpoR signaling attenuation (Becker et al., 2010). Studies indicate that ligand depletion can influence the mitogenic potency of EGF in cell culture. In

conditions of limited EGF, the extent of EGF depletion can depend on the absolute amount, and not the concentration, of EGF in the media; hence, the total volume of the media per cell can influence depletion in cell culture conditions (van de Poll et al., 2005). In turn, the total volume of the media per cell is determined by the total number of cells incubated in the well, or the cell density. Given that in our experimental setup we studied signaling responses to persistent EGF stimulations over extended times, we needed to establish if in our system EGF depletion presented a concern at lower EGF doses.

First, we fitted our model (see Section 4.2.1) to the phosphorylation and the total EGFR data obtained from wells containing on average 15000 cells. The model predicted that three hours after the addition of the lowest EGF dose (0.01 ng/ml EGF), as much as 70% of initial EGF can get depleted in certain fits (Figure 4.52).
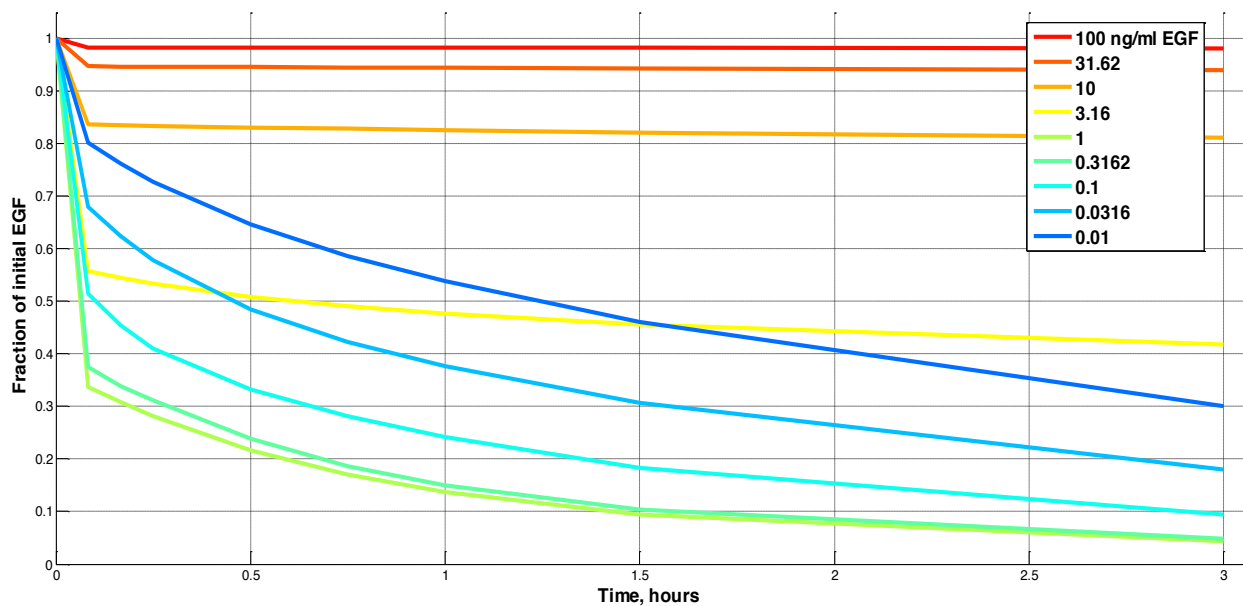
Figure 4.52: Predicted EGF depletion: time-courses of the fraction of initial EGF remaining in the media in a model fit.

We then validated this prediction experimentally with ELISA assays (Weemen & Schuurs, 1971). Specifically, we measured the change in EGF abundance in the media in which cells had been incubated with EGF for three hours. The EGF depletion was measured for four different doses of EGF. We found that in wells where ~15000 cells were plated initially, by the end of three hours of EGF stimulation as much as 48% of the initial EGF can be depleted (Figure 4.53).

152
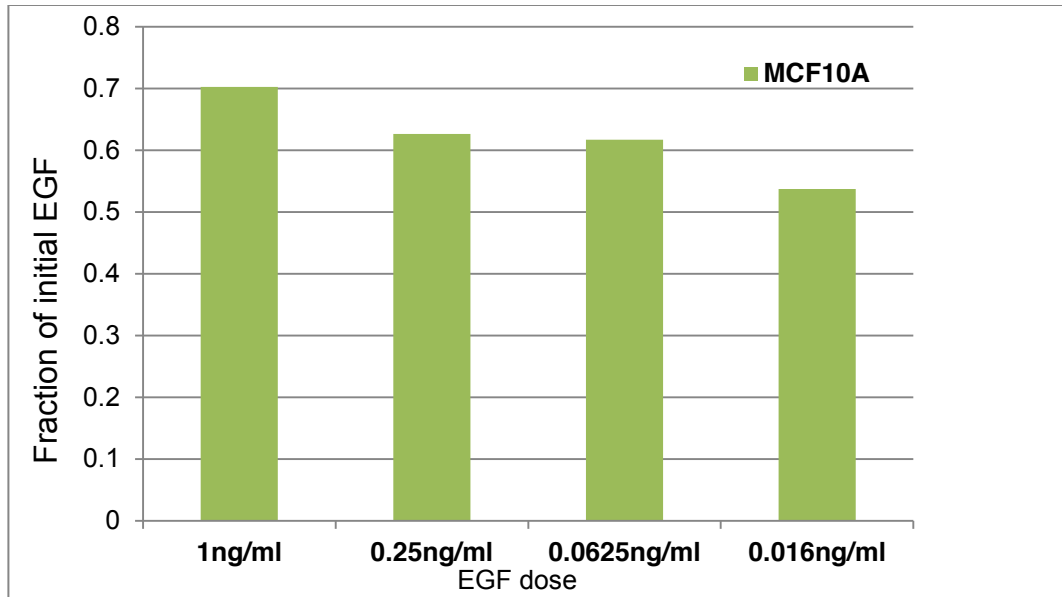
Figure 4.53: Depletion of EGF ligand measured experimentally with ELISA assays. Cells were plated at 15000 cells per well, and EGF depletion was measured three hours after incubation with a given EGF dose. Reported are the ratios of the amounts of EGF remaining after three hours to the initial amount of EGF.

We then used our model simulations to identify the range of cell densities which would prevent significant EGF depletion from the media. For this, we fitted our model to the phosphorylation and the total EGFR data as previously described, however, we also imposed a constraint that at low EGF doses no more than 10% of the initial EGF could be depleted after three hours of stimulation. In addition, we simulated the number of cells per well as a parameter in our model. Essentially, cell density parameter determines the absolute amount of EGF molecules that each individual cell is exposed to at a given EGF treatment.

Model simulations predicted that to prevent EGF depletion, each well should contain on average approximately 7,000 cells (median ~ 4,900) and no more than approximately 8,000 cells (Figure 4.54). We then adjusted our experimental protocol and decreased the amount of cells seeded in each well from ~ 10,000 to ~ 5,000. As controls, in parallel, we also seeded plates with ~10,000 and ~15,000 cells per well. We then reran all our main experiments at the ~ 5,000 cell density suggested by our model. Indeed, we found that the decreased cell density allowed to prevent significant EGF depletion and, thus, allowed to maintain EGF at nearly constant levels over three hours.
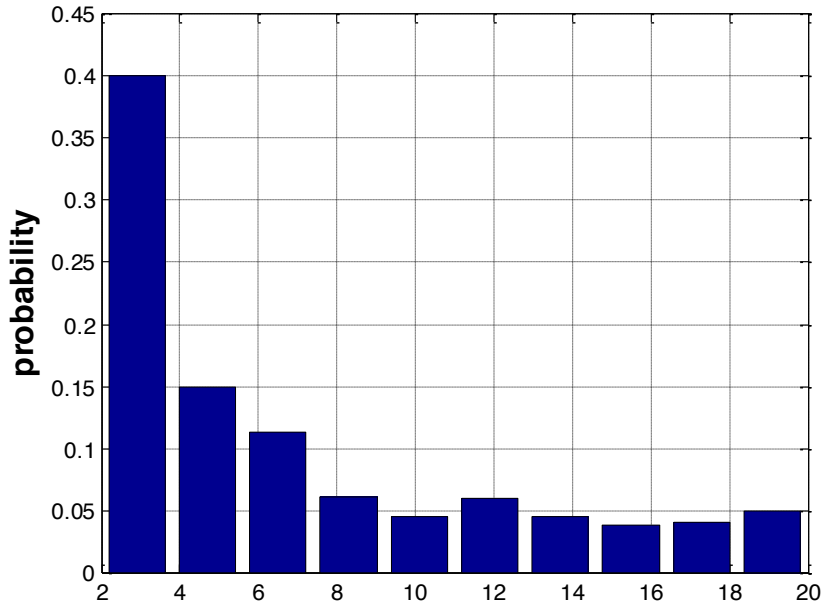
Figure 4.54: Predicted distribution of the number of thousands of cells per well with limited (<10%) EGF depletion by three hours of cell incubation with low dose EGF. Section 4.5.2: Cell density in culture affects cell signaling responses

We then wondered if the cell density could influence cell signaling responses to EGF, and if so, through which exactly mechanisms? How does cell-density dependent EGF depletion affect EGF signaling, if at all?

Cell-density dependent effects on signaling have been previously reported in cultured cell lines with EGFR signaling (Aoki et al., 2013). In a recent work, cells were seeded into the wells at various cell-densities ranging across three orders of magnitude and incubated with standard media for several hours (Aoki et al., 2013). The authors then investigated the relationship between cell density, cell proliferation rate and ERK activity. The proliferation rate was found to be most correlated with the ERK activity. The ERK activity was diminished at the lower cell densities and was the highest at the medium cell densities. In addition, the basal ERK phosphorylation was shown to be dependent on the cell density through the EGFR activation.

We investigated if in our experimental system cell density could affect cell signaling responses to EGF. Specifically, we hypothesized that the cells incubated at distinct cell densities prior to EGF addition could produce varying signaling responses to EGF. To test our hypothesis experimentally, we measured the phosphorylation and the total EGFR responses to the EGF

treatments in the cells plated at three different cell densities, namely, 5000, 10000 and 15000 cells per well. All three experiments were conducted on the same day to limit possible biological variations in the cells.

First, we looked at the dose responses of pEGFR, pERK, and pAKT 10 minutes after EGF addition in the cell populations incubated at the three different densities (Figure 4.55, Figure 4.56). In MCF-10A cells, some dose responses in the cells seeded at higher densities were to some extent shifted to the left (pERK), suggesting a higher sensitivity to the low EGF doses. In SKBR3 cells (Figure 4.56), the 10-minute dose responses at higher cell densities in addition had higher amplitudes of the responses. We also considered the time courses of the phosphorylation signals at various EGF doses and found that in the long run, the response dynamics appear similar at all three cell densities. Thus, our experiments suggested, that the cell density might affect not only the cell sensitivity to EGF, but also the amplitudes of the cell responses. This observation emphasizes the importance of maintaining similar cell densities in the experiments measuring EGFR signaling across repeats performed on different days, in order to increase the chances of observing a more consistent and reproducible experimental results.
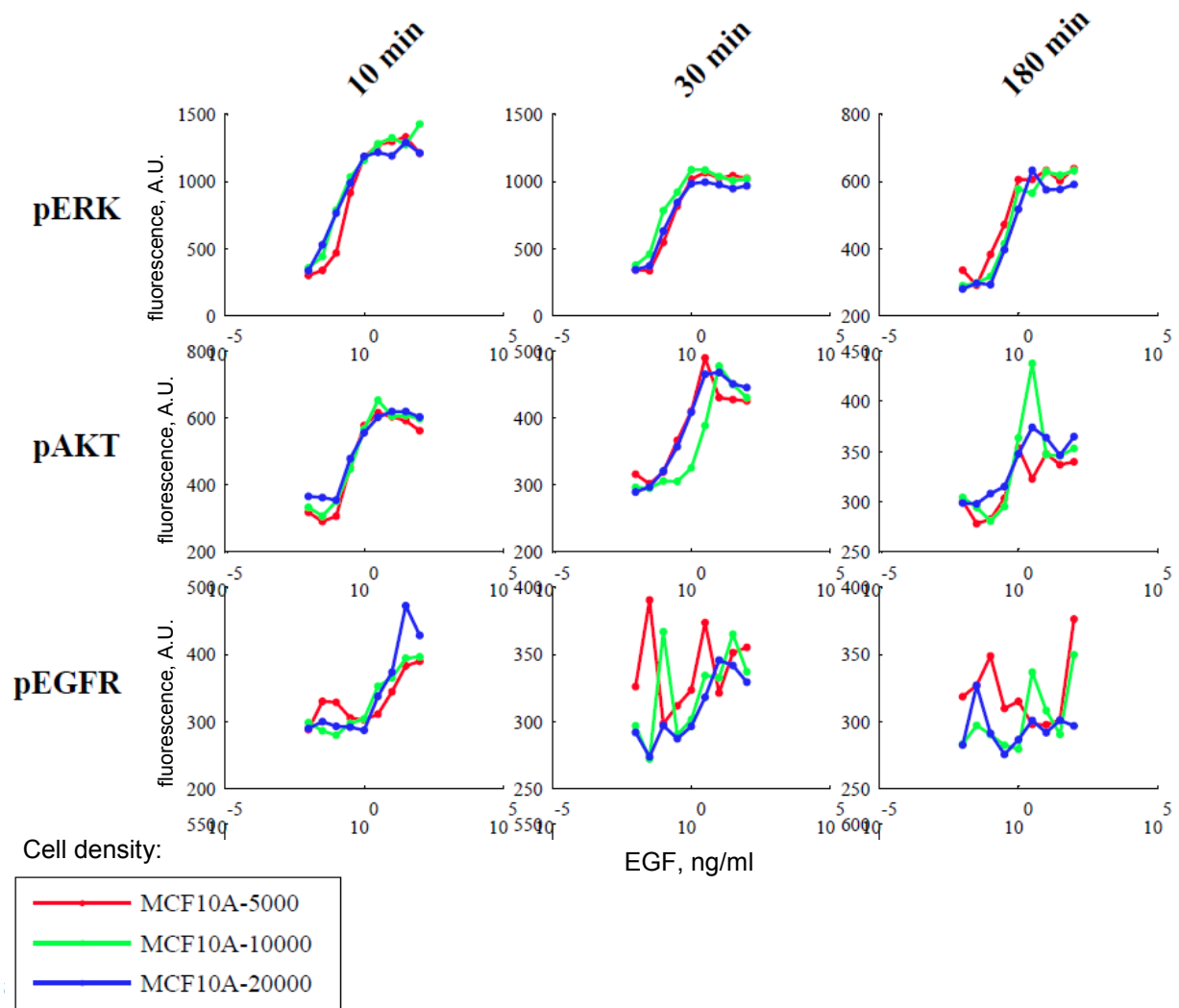
Figure 4.55: MCF-10A cells: show cell-density dependent variation in dose responses. MCF-10A cells were incubated at5,000,10,000 or 20,000 cells per well. On the X axis is shown the dose of EGF in each treatment. On the Y axis are shown the responses measured at 10 minutes after EGF addition.
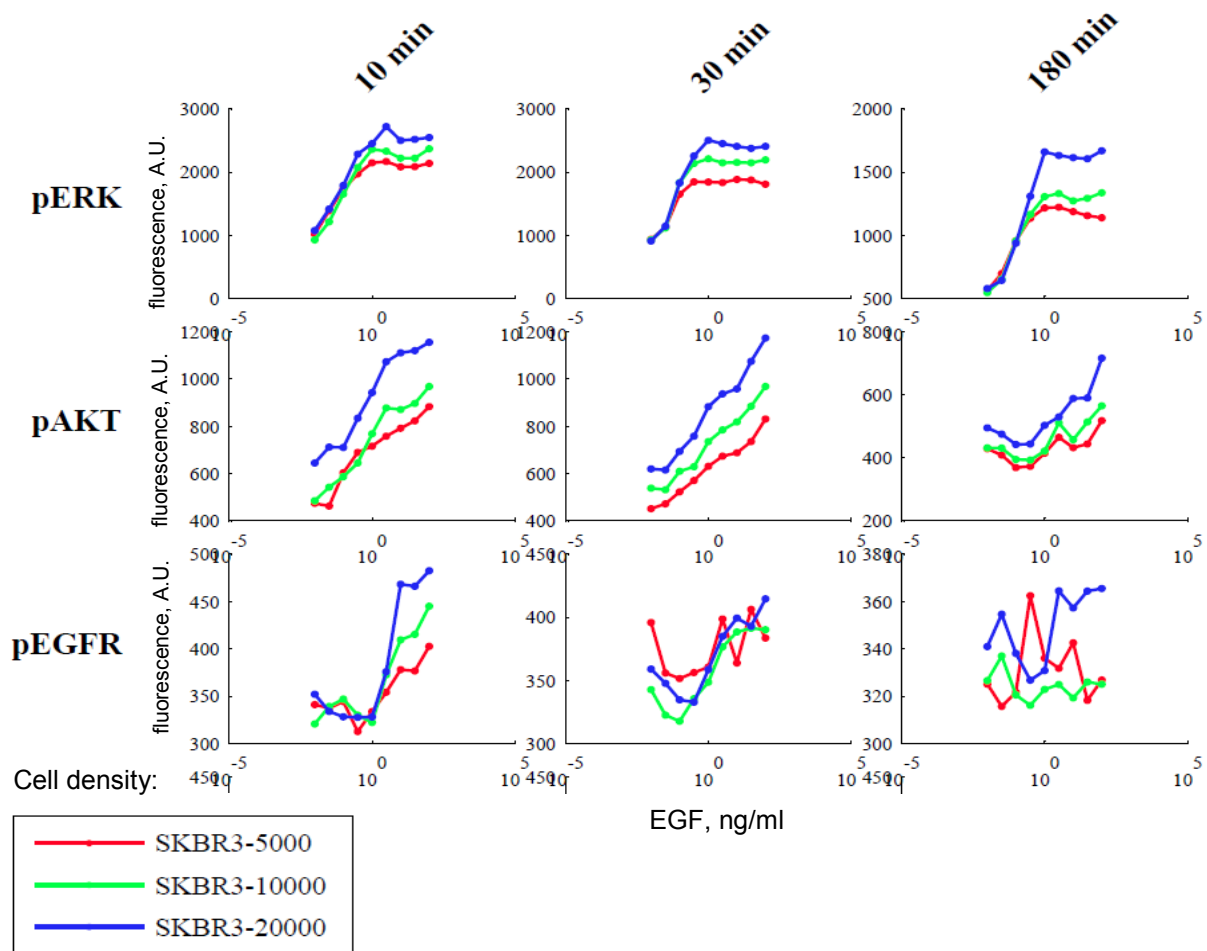
Figure 4.56: SKBR3 cells: show cell-density dependent variation in dose responses. SKBR3 cells were incubated at5,000,10,000 or 20,000 cells per well. On the X axis is shown the dose of EGF in each treatment. On the Y axis are shown the responses measured at 10 minutes after EGF addition.

We then fitted our model to the phosphorylation and the total EGFR data obtained from the experiments at ~10,000 cell density (as described in Section 4.5.1) and used the model predictions to uncover the differences in the parameters values of the fits predicted to have dissimilar rates of EGF depletion.

We compared the distributions of parameter values between the subset of fits predicted to have no significant EGF depletion (<10%) to the subset of fits predicted to have significant EGF depletion (>30%). We found that the EGFR abundance was among the top parameters most different between the two conditions. We found that the total EGFR abundance positively correlated with the

cell density: the higher the number of the cells in wells, more EGF receptors the cells are predicted to express (Figure 4.57).
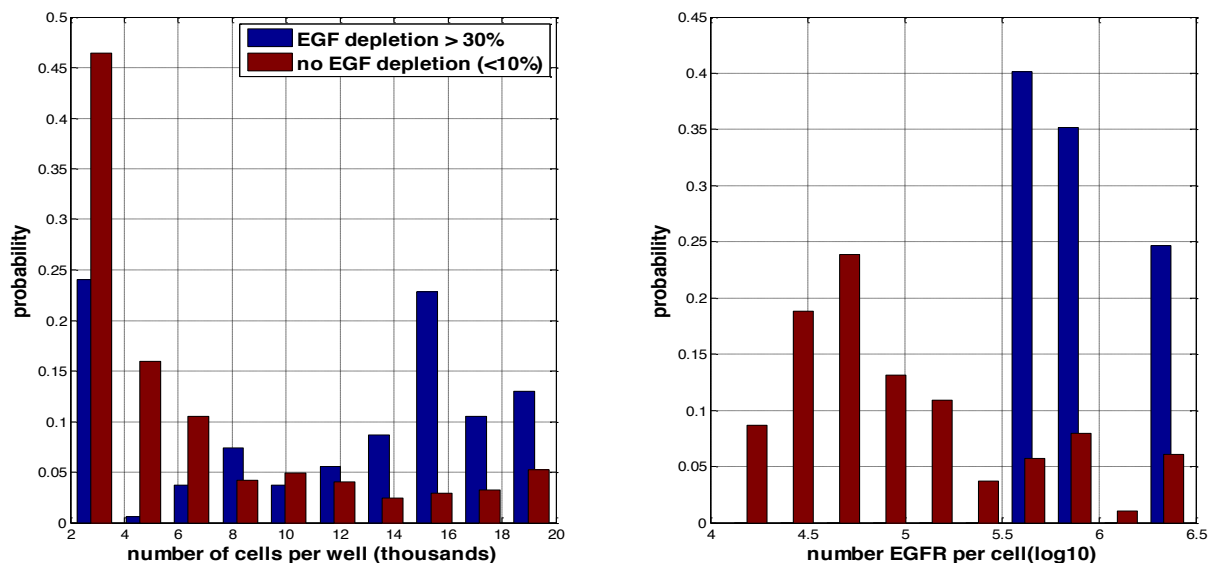
Figure 4.57: Predicted distributions of the number of cells per well and the number of EGFR per cell in subpopulations with or without significant EGF depletion by three hours of EGF treatment.

To validate the predicted relationship between the EGFR expression and the cell density, we analyzed the raw measurements of the total EGFR abundances at time=0 and no EGF addition collected from the cells incubated at different cell densities: 5,000, 10,000 or 15,000 cells per cell. Since all the measurements were conducted on the same day, fluorescence scaling and background offset were identical in all three conditions, and thus, we could directly compare the tEGFR readouts from these three conditions. We found that, indeed, at higher cell densities tEGFR appears to be higher (Figure 4.58). Hence, the differences in cell-density dependent responses to EGF could be partially explained by the different levels of EGFR expressed in the cells cultured at different cell densities. The importance of the levels of EGFR expression in cell signaling responses has been previously explored in literature and is reviewed in Section 2.6 of this thesis.
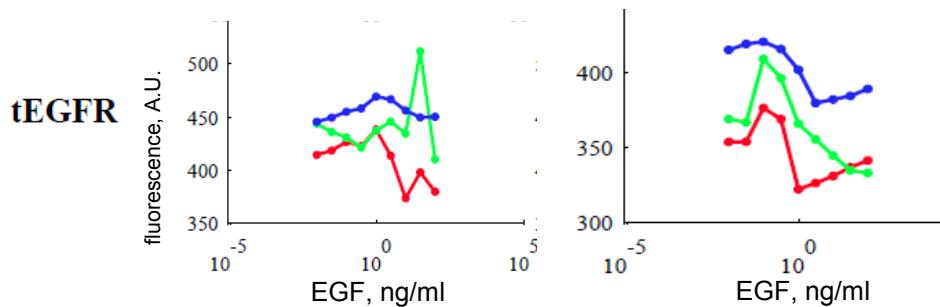
Figure 4.58: Experimental measurements of the total EGFR abundance at 10 minutes of EGF stimulation in cells plated at various cell densities: 5,000 (red), 10,000(green) or 20,000 (blue) cells per well. On the X axis are the EGF doses (ng/ml, log10). Figure on the left shows measurements for MCF-10A cells, figure on the right shows measurements for SKBR3 cells.

Section 4.5.3: The role of EGF depletion in the cell signaling through EGFR

Overall, the EGF depletion might have different effects on cell physiology, depending on the cellular context. For example, EGF depletion has been suggested to decrease the effects of the EGFR-mediated mitogenic responses (van de Poll et al., 2005). Moreover, studies indicate that it is the proportion of ligand to receptor, rather than just receptor abundance, that might influence mitogenic responses (van de Poll et al., 2005). Interestingly, in the cells expressing high levels of EGFR, such as A431 cells, EGF depletion has been shown to be relevant during the mitogenic stimulation, but not during the growth inhibition of these cells (van de Poll et al., 2005).

In this section we present a mechanistic explanation behind the observed positive correlation between the cell density and the EGFR abundance. Our results suggest that at limiting EGF concentrations, EGF depletion occurring in the course of cell pre-incubation can affect the EGFR expression in a cell-density dependent manner. Prior to the EGF treatment the cells are being incubated in a standard starvation media which might contain trace amounts of EGF. The cell density effectively determines the absolute number of EGF molecules each cell is exposed to given a certain level of EGF in the media. Given the same initial abundance of receptors, a higher number of cells in a well would imply a lower number of EGF molecules per cell. Hence, the residual EGF in the incubation media of plated cell cultures, over time, can be completely depleted by the cells, if the number of cells is sufficiently high. Eventually, ligand depletion from cell media would allow the cells to recover and accumulate recycled or *de novo* synthesized receptors on the cell surface, over time, leading to the higher receptors abundances.

159

However, if the number of cells in the well is low, the amount of EGF per cell might be sufficiently high thus avoiding sufficient ligand depletion. In such case, the EGF presence in the media could lead to receptor phosphorylation, followed by their internalization. Therefore, given the EGF levels in the media are maintained constant, the receptors will be continually activated and internalized. Unless EGF is removed from the media, EGFR abundance on the cell surface would not be able to recover. Hence, the cells incubated at low cell densities in the media with traces of EGF would maintain low levels of EGFR expression.

We then ask if ligand depletion is important to immediate-early phosphorylation responses to EGF. The model predicts that independently of the cell density, by 10 minutes of stimulation across all EGF doses, ligand is not significantly depleted (only in 0.04% of the fits ligand depletion exceeds 5%). In our model EGF is depleted through a rapid endocytosis of ligand-bound receptors, a process which relies on a relatively slow assembly of internalization vesicles. Therefore, by the 10 minutes of EGF treatment not many receptors have been internalized yet, and thus, EGF depletion is not significant at that time. Therefore, EGF depletion is not likely to play a role in directly shaping the signaling responses in the short run.

Here we suggest an explanation how cell density can affect signaling outcomes through EGF depletion and levels of EGFR expression: during incubation with limited amount of EGF in the media, higher cell density means less absolute number of EGF molecules per cell, which then allows EGF depletion from cell media, leading to higher levels of EGFR expression, which determines signaling outcomes. EGF depletion does not affect cell signaling in the short run, however, plays important role in regulating number of cell surface EGFR during incubation of cells in limited EGF media before the EGF treatment.

It is important to note that here we considered the effects of EGF depletion on the early stages of signaling responses, namely, the pre-incubation of the cells and the short-term (10-minute) phosphorylation responses. However, EGF depletion might have a significant effect on the cell signaling in the long run. Generally, in physiological context ligand can be depleted through receptors-mediated endocytosis, other type of membrane-mediated endocytosis, or through diffusion away from the cell surface. Ligand clearance from the cell surface could contribute to the attenuation

of cell signaling, thereby, effectively, influencing the dynamics of signaling responses. In addition, in the receptors sensory systems in which ligand-bound receptors are internalized faster than ligand-free receptors, ligand depletion could facilitate the recovery of receptors levels after initial stimulation with the ligand. The recovered receptor levels could then alter the cell sensitivity to the subsequent stimulations with the ligand. Therefore, ligand depletion potentially could have a significant effect on the long term cell signaling.

To summarize, we predict that in the case of limiting initial EGF levels in the cell incubation media, the cell density will have an effect on the EGFR expression: higher cell density would lead to higher EGFR expression in the cells. We also predict that in conditions when EGF is not limiting in the cell incubation media, the cell density should have no effect on EGFR expression in the incubated cells.

To test our hypothesis experimentally we propose the following experiment. On a given day, identical cells are being plated in four wells: two wells with low cell density, and two wells with high cell density. Two plates of different cell densities are then incubated with standard media, containing residual EGF, and two remaining plates are incubated in standard media with addition of a low dose EGF. After several hours of incubation, the total EGFR abundance and EGF levels are measured in each well. The measurements should be taken between three to six hours of incubation – the time frame long enough to allow for ample EGF depletion, yet not sufficient for most cells to undergo cell division, and thus significantly change the cell density. In addition, the total EGFR and EGF depletion in these four types of wells could be measured at several time points during the extended incubation to assess the dynamics of the relationship between ligand depletion and the total EGFR abundance. If the suggested hypothesis is true, we would expect to see a positive correlation between the average number of EGFR receptors on the cell surface and the incubation cell density. In addition, given the two batches of cells incubated at the same cell density, we would expect the cells incubated with the additional EGF to have a lower number of the cell surface receptors.

Section 4.5.4: Remarks on the cell-density dependent effects in cellular context

In Section 4.5.3 we proposed a relatively simple explanation of the cell-density dependent effects occurring at the pre-incubation stage on the signaling through EGFR and the role of EGF depletion in it. Testing the most parsimonious hypothesis first could yield a valuable insight about the biological mechanisms in the EGFR signaling system. However, the proposed explanation is limited due to its simplistic view on the behavior of cell populations in the media. In particular, cell-density can influence cells in many other ways, not just though EGF and its receptors: at high cell-density cell behavior can be affected by cell contact-inhibition (Swat, Dolado, Rojas, & Nebreda, 2009), autocrine, juxtacrine and paracrine signaling (Singh & Harris, 2005). Hence, multiple other cell-densities dependent factors can affect EGFR expression in the cells.

Importantly, autocrine secretion of EGF by the cells can influence levels of EGF in the wells, and thus, according to our hypothesis, can have effect on EGFR expression levels. However, it is not feasible in our experimental system to quantify the extent of autocrine EGF in the media. First, the ELISA assays measuring the EGF abundance in the media detect primarily the soluble EGF; whereas the autocrine-produced EGF might remain tethered to the cell surface for some time prior to being released into the media (Singh & Harris, 2005), and therefore, it could be missed by the ELISA assay. Second, the levels of the autocrine-produced EGF in our system can be very low and difficult to detect experimentally. Moreover, our experimental handling of the cells prior to the EGF treatment ensures that most of the EGF autocrine-produced during the pre-incubation is removed when the cells media is replaced with the fresh starvation media. Further, we expect that the three to six hours of EGF treatment is not sufficient for the cells to accumulate significant amounts of the autocrine EGF. The levels of the autocrine-produced EGF are expected to be much lower compared to the doses of EGF in the treatments, and hence, the autocrine signaling is not likely to compensate for the EGF ligand depletion or interfere with the EGF dose applied in the treatments.

Section 4.6: Predicted Weber's Law in pERK responses and its experimental validation

In this section we first present predictions of the extended model of responses to step- inputs in EGF doses. The model first is fitted to the 3 hours phosphorylation and total EGFR time-courses, and then responses to a 3-fold increase in EGF added at 3 hours are simulated. We compare peaks

of pERK responses to the same x3 fold increase steps across different initial doses of EGF and find approximately 1.5 orders of magnitude region of initial EGF doses in which pERK responses to same fold increase appear similar.

We then consider the corresponding total EGFR abundances at three hours of stimulation with a constant initial dose of EGF and observe that within a certain range of EGF doses, the total EGFR at steady state is negatively correlated with the dose of EGF in the initial treatment. This outcome leads us to hypothesize that the total EGFR can store memory of initial doses of EGF, and the amount of receptors available for ligand-binding at the time of the EGF step-increase could determine the maximum response to step input in EGF and enable sensing relative changes in EGF levels. We then present experimental validation of our prediction in Section 4.6.3. Since the results in this section were obtained by simulating a detailed ODE model of the ErbB signaling cascade which reproduces well the experimental data, but is complex and thus not well suited for intuitive understanding of underlying biological mechanisms, in Section 4.7 we develop series of simplified analytical models capturing ErbB signaling responses. Further, we explore properties of parameter space at which pERK responses exhibit Weber's Law behavior described in Section 4.6.1.

Section 4.6.1: Predicted responses of phospho-ERK signals to step inputs in EGF are able to discriminate different fold-changes in EGF

We fit the extended to model to phosphorylation signaling time-courses of pEGFR, pERK and total EGFR responses over up to three hours of treatment with constant EGF doses as described in Section 4.2.1 and at each sampled point in parameter space we simulate model predictions of system responses to a 3-fold step increase in EGF concentration added at 3 hours.

Shown in Figure 4.59, Figure 4.60, and Figure 4.65 are model responses simulated at a point in parameter space found by MCMC chains. Responses up to three hours are fitted to the experimental data, whereas, the 3-6 hours parts of the time-courses are predicted responses to a x3 fold step increase in the initial EGF dose. Shown in bold are the predicted time-courses in which peak responses in pERK at the 3 hour step increase appear approximately similar to each other. Although the initial EGF doses applied in these simulations differ across 1.5 orders of magnitude and produce

163

different pEGFR and pERK responses when added (see amplitude of the first peak in Figure 4.59, Figure 4.60), the responses to the same x3 fold increase in EGF dose produces responses with strikingly similar peaks. To verify that the observed behavior is not a result of signal saturation, we also apply step increases with a higher fold, x5 (Figure 4.61). We observe that at most doses (including ones responses to which are shown in bold in Figure 4.59, Figure 4.60) responses to a higher step increase have higher amplitudes then the ones to the x3 step increase, suggesting that the addition of more EGF molecules could still produce more EGF signal, implying that the system is not at saturation. Responses to conditions in which the system has been pre-treated with the highest EGF dose, 100ng/ml EGF, show only small signal increase in response to x3 fold step input in EGF, with a total of 300ng/ml EGF added (see Figure 4.60, dashed blue), suggesting that after stimulation with a high initial dose of EGF the system has possibly lost its sensitivity to the ligand.
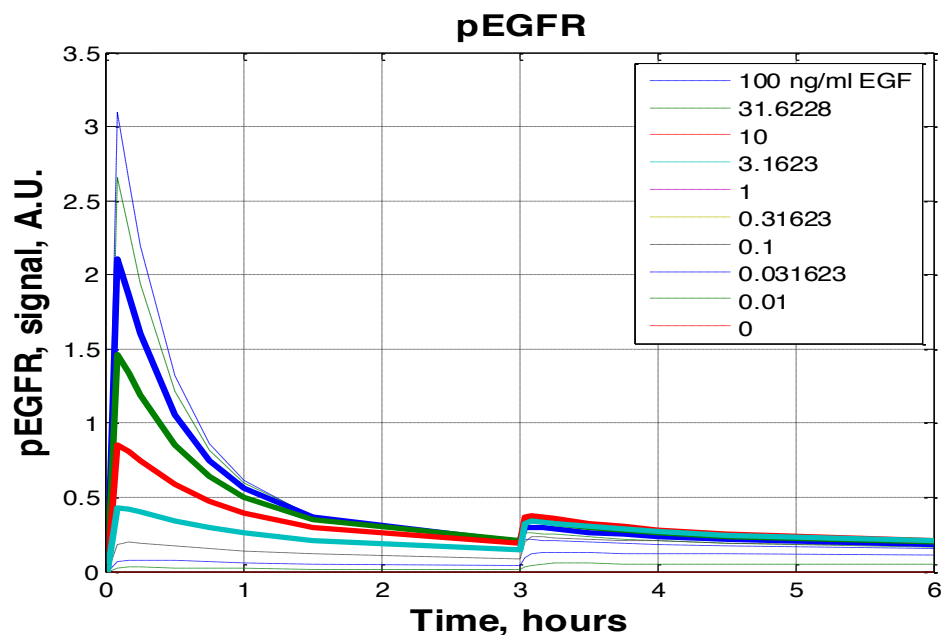
Figure 4.59: Predicted pEGFR time-courses in response to step inputs in EGF with different initial EGF doses. A x3 fold step input relative to the initial EGF dose (see legend) is applied at each simulated EGF dose at 3 hours
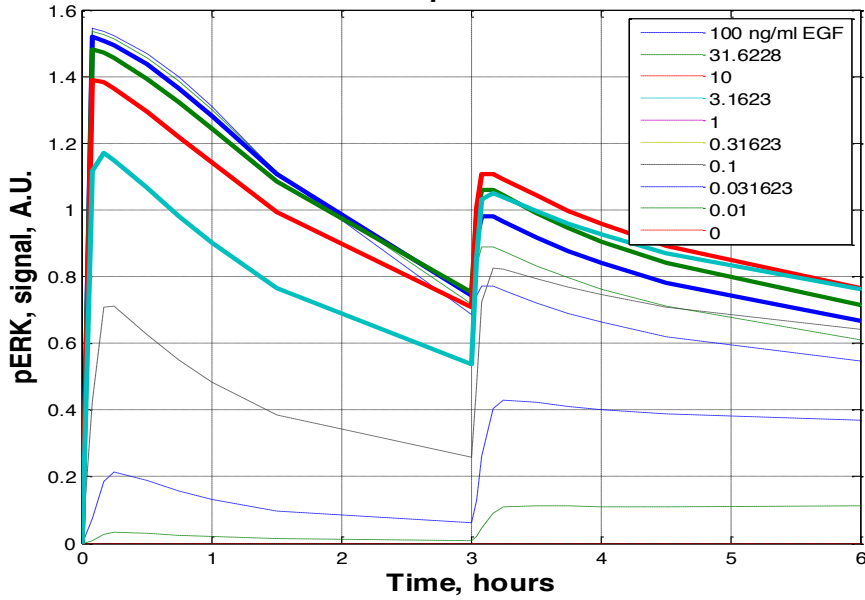
Figure 4.60: Predicted pERK response to x3-fold step inputs in EGF applied at 3 hours of stimulation.
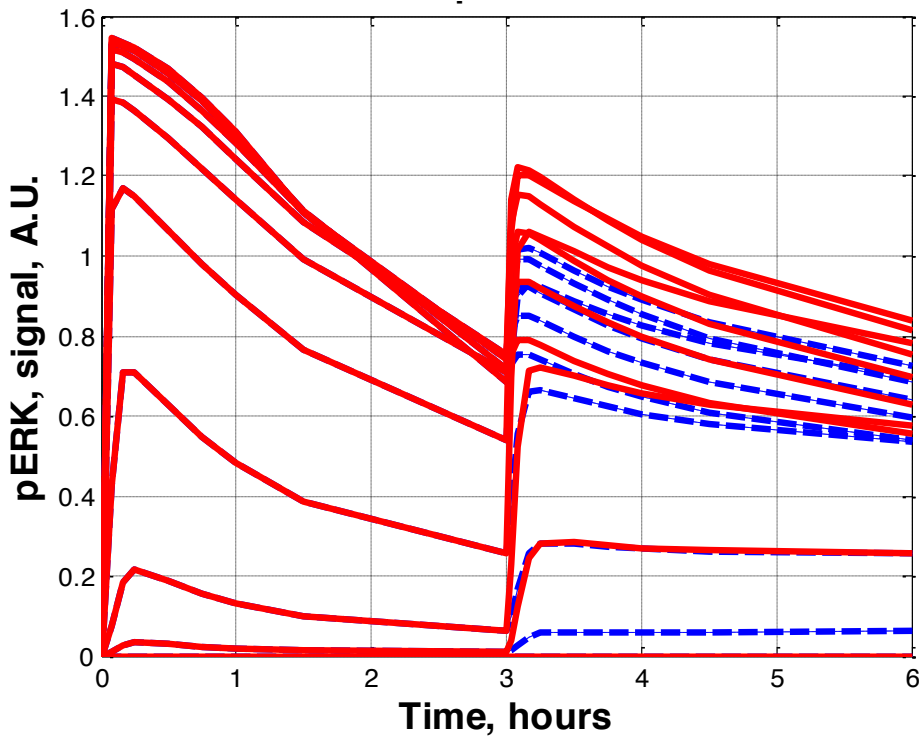
Figure 4.61: The comparison of pERK responses at x3 or x9 step inputs in EGF applied after three hours of EGF treatment .Responses to x3 step inputs are shown in blue, x9 - in red.

We then generate dose responses of maximal amplitude to x3 fold increases in EGF across all doses of EGF from the posterior parameter distributions obtained by fitting the model to the experimental measurements of up to three hours EGF stimulation (Figure 4.62). The predicted

distribution is relatively flat in the region of EGF doses approximately between the 0.15-10ng/ml range. In the range of high EGF doses (>10ng/ml) the predicted dose response appears to decay, possibly due to saturation effects (see below). However, we observe that the flatness of the dose-response in the 0.15-10ng/ml range is a feature predicted to be observed across the parameter space, and not just a characteristic of several selected fits.
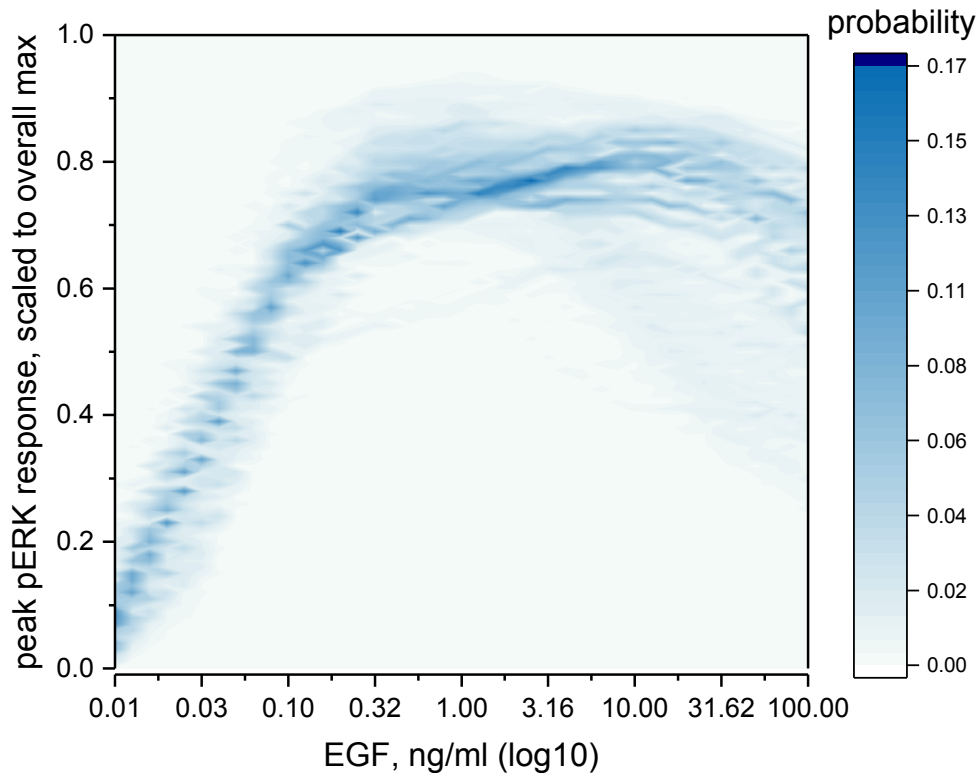
Figure 4.62: Predicted distribution of the maximal pERK responses to x3 fold step input in EGF dose applied three hours after stimulation with EGF. On the X axis are shown the doses of EGF in the initial treatment applied at time=0.

We then investigate if in the selected fit the system can distinguish between different fold increases in EGF levels. For this, we plot maximal system responses in that particular fit to different fold increases in EGF levels (Figure 4.63, Figure 4.64). The x1 control represents system responses if at three hours the EGF concentration was not changed. The dose-response of maximal responses to the initial treatment with EGF is plotted for reference (black). We find that, indeed, higher fold-increases produce responses with higher maximum values across the range of initial EGF doses ~0.1-3ng/ml. The dose-responses for each fold increase appear non-overlapping and separated from

each other, suggesting, that in response to treatment with different fold increases in EGF the system can produce distinct responses. However, such property holds only in a limited range of EGF doses, as at higher doses responses are not distinguished, whereas, at lower doses responses depend on the initial dose of EGF in the pre-treatment.
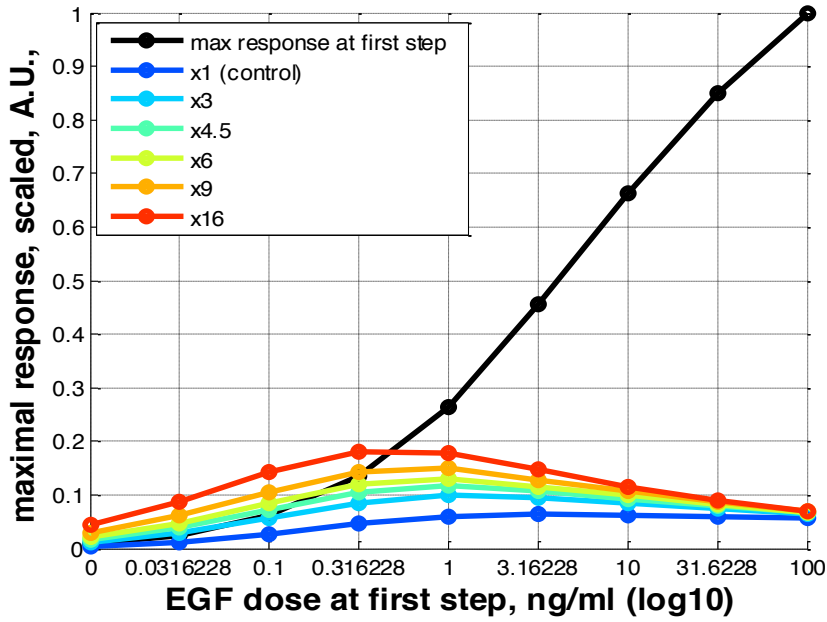
Figure 4.63: Predicted pEGFR dose-response to different fold-increases in EGF. On the X axis are shown the doses of EGF in the initial treatment applied at time=0.
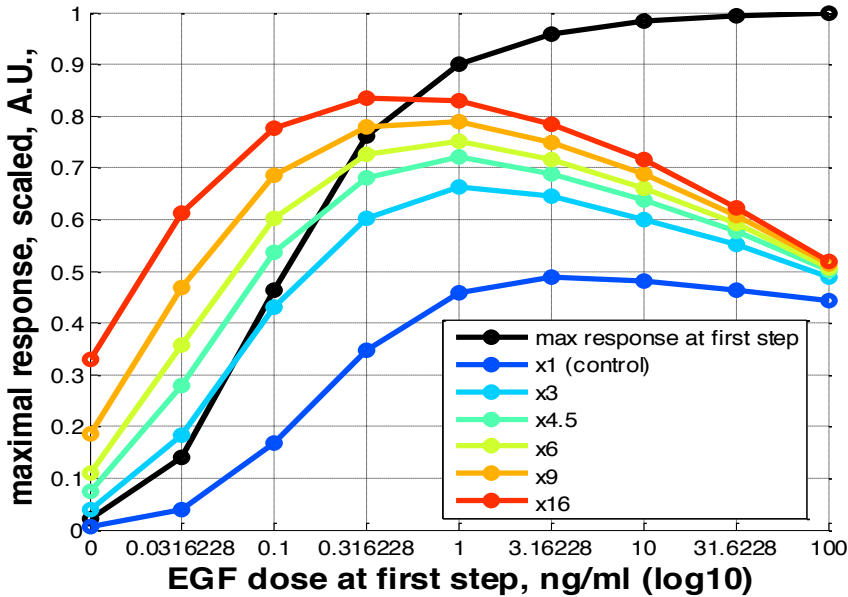
Figure 4.64: Simulated pERK dose-response to various fold change step increases in EGF. On the X axis are shown the doses of EGF in the initial treatment applied at time=0.

Peak pERK responses to initial treatment with EGF (black) exhibit saturation at high doses of EGF (>10ng/ml) (Figure 4.64), at the same time as the corresponding pEGFR responses are not yet saturated (Figure 4.63). Such pERK saturation is a result of ultrasensitivity, whereby the signal at pEGFR is amplified through the kinase cascade when transmitted to pERK. The responses to step increases in EGF, however, show a different type of saturation. We see that after three hours of stimulation with high dose EGF, most of the EGFR are predicted to be degraded (Figure 4.66), and therefore, pEGFR even at very high levels of additional EGF would produce responses with only small amplitudes (Figure 4.63, responses > 10ng/ml). Therefore, pEGFR response at step inputs in these cases is saturated. However, the corresponding pERK responses are still below the possible maximum of pERK response (maximum of the dose-response at first step, Figure 4.64), and given that the total ERK amount could not be degraded over the time of the signaling, we conclude that pERK responses appear to be saturated as a result of saturation on pEGFR due to decreased amount of receptors, and not as a consequence of phosphorylation of all possible ERK available.

The predicted flatness in dose-responses of pERK to EGF step-inputs occurs at EGF doses below the estimated constant of equilibrium of EGF unbinding from EGFR (~20ng/ml EGF (Chen et al., 2009)). Hence, to exhibit approximate Weber's Law behavior EGF binding to the EGFR receptors should be below its saturation.

Therefore, we find that the model fitted to experimentally observed phosphorylation and the total EGFR time-courses over three hours of treatment with different EGF doses predicts that in a certain range of EGF doses the system can exhibit Weber's Law in its responses and, hence, be able to sense relative changes in EGF doses.

Section 4.6.2: Total EGFR dose-responses to EGF shows negative correlation with EGF dose applied

We first computationally investigate what mechanisms in the model might be responsible for producing the observed Weber's Law in pERK responses to fold-changes in EGF levels. For this, we consider system output predicted by the model and representing variables other than phosphorylation responses. In order to produce phosphorylation responses proportional to fold-changes in the

incoming signal, the system should be able to store information about the previous concentration of EGF stimulation at the time the EGF step is added, and compare the new level of EGF to the previous one. Since the responses to EGF treatments are initiated at the cell surface when EGF is detected by the EGFR receptors available for activation on the cell surface, we consider total EGFR receptor abundance.

In particular, we plot the predicted time-courses of total EGFR receptors (Figure 4.65). We find that, in agreement with the corresponding experimental data, the predicted total EGFR abundances after three hours of stimulation with the constant dose EGF are decreasing with the increasing dose of the initial EGF stimulation. In addition, the total EGFR time-courses appear to be reaching steady state by 3 hours at some EGF doses. Shown in bold are total EGFR time-courses corresponding to the treatments with EGF in which the model predicts approximate Weber's Law in pERK responses (Figure 4.60).

We then plot a dose-response of the amount of remaining EGFR after three hours of stimulation with each corresponding EGF dose (Figure 4.66). The lower the number of receptors remaining, the more of the initial receptors were degraded in the course of signaling. The dose-response suggests that at lower EGF doses, the number of receptors does not change significantly after three hours of EGF stimulation. At the medium range of EGF the total EGFR negatively correlates with the dose of EGF in the initial treatment. We then hypothesize that such dependence of tEGFR on the EGF in the middle range of EGF doses enables the mechanism of Weber's Law in pERK responses. Intuitively, given the system is pre-adapted to stimulations with two different constant doses of EGF, u_o and u_1 ($u_1 > u_o$), the number of EGFR receptors remaining available will be higher in the system adapted to u_o. Therefore, to produce responses of similar maximal amplitudes, a smaller absolute amount of EGF molecules would have to be added to the system pre-adapted to signal u_o, compared to the system, pre-adapted to u_1. This conclusion assumes that the phosphorylation response, up to a constant, is proportional to the number of receptors available for ligand-binding and the dose of the applied EGF concentration.

We then look at the predicted distribution of dose-responses of total EGFR after three hours of stimulation with EGF generated by sampling the parameter space of the model fitted to the experimental measurements (Figure 4.67).

The predicted distribution of the dose-responses exhibits a similar trend as seen in Figure 4.66 with a flat region at lower doses and log-linear behavior in the middle range of EGF.

Note, that, the model-predicted degradation of EGFR differs from the one observed experimentally, although, the time-courses of EGFR abundances were included in model fitting (Figure 4.68). It is possible that the quality of total EGFR time-courses was compromised in these fits due to the fact that multiple phosphorylation time-courses were fitted in the model simultaneously, and thus, significantly constrained the model parameter space.

Although the model predicts that total EGFR after three hours of EGF treatment is negatively correlated with the EGF dose, a trend captured in the experimental measurements of tEGFR fitted in the model (Figure 4.3), the model predicts somewhat different then experimentally observed rates of degradation at EGF doses above 10ng/ml EGF. Such discrepancy is possibly due to the normalization and scaling of total EGFR experimental measurements: in the absence of proper control to account for fluorescence background offset in total EGFR measurements, we might have over or underestimated the initial abundance of total EGFR in the system (counting fluorescence emitted by background as fluorescence due to detected EGFR), and thus, underestimated the amount of tEGFR degraded. We address the problem of estimating the total number of EGFR remaining in the cells and able to bind ligand in Section 4.7.6.

An important difference between model predictions and experimental measurements of the total EGFR is that at higher doses experimental data suggests that tEGFR degradation is saturated, whereas the model predicts a more gradual receptors decay dose-response at higher EGF doses. In EGFR signaling time-courses of total EGFR are related to time-courses of pEGFR, as a higher number of phosphorylated receptors means more of them will be degraded in the course of stimulation. Our experimental data (Figure 4.3) measures 10 minute responses of pEGFR and does not imply that pEGFR responses are saturated. However, the underlying biological responses might have been saturated instantly after EGF treatment, but were missed to be detected as saturated

when measured 10 minutes after the EGF addition. This observation may explain why the total EGFR responses after three hours of treatment appear to be saturated whereas; the pEGFR dose-responses at 10 minutes did not seem saturated in the experimental data.

The fits found by the model correctly capture that the pEGFR responses may not necessarily be saturated at higher EGF doses, however, do not fit correctly the total EGFR saturation at higher EGF doses (Figure 4.68). Overall, the discrepancy in model predictions of total EGFR and their experimental validation can be resolved with the improved normalization and scaling procedure of tEGFR experimental data, and we believe that that does not disagree with key findings of this work.

Since model fit to total EGFR varies from experimental measurements, the range of EGF doses in which the model predicts Weber's Law might also vary from the one in the underlying biological system. Therefore, in the experimental validation of predicted Weber's Law we test range of EGF doses somewhat broader than the one predicted by the model (Section 4.6.4).
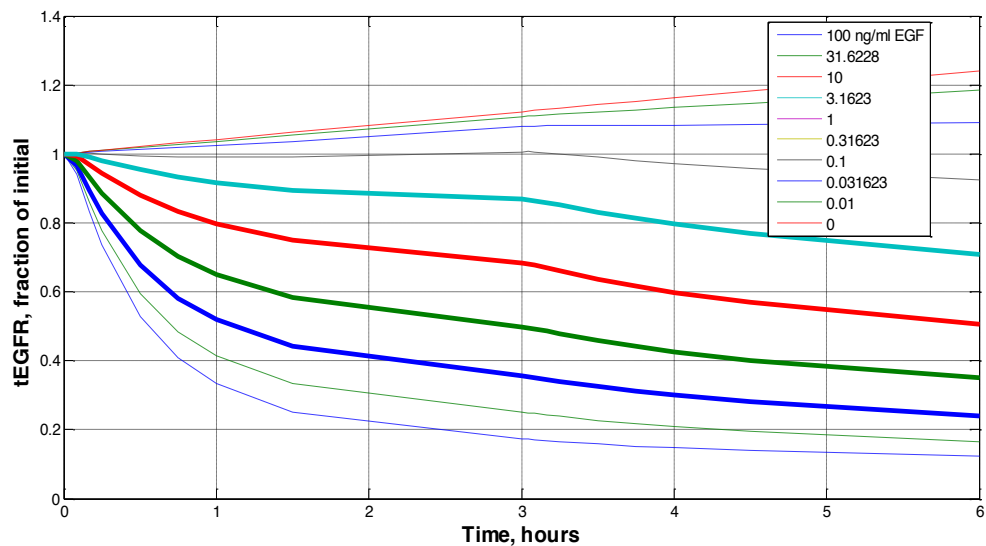
Figure 4.65: Predicted total EGFR time-courses in simulations with x3 fold step inputs in EGF doses added at three hours of stimulation with initial EGF.
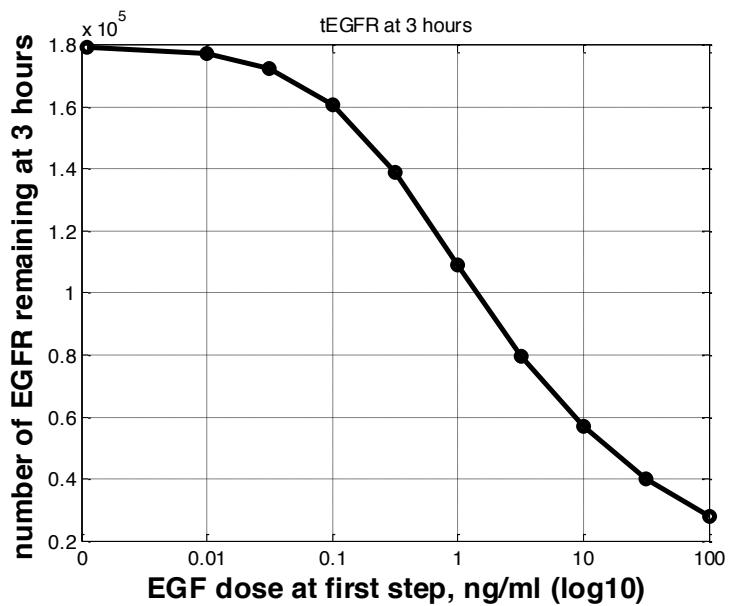
Figure 4.66: Predicted dose-response of tEGFR three hours of stimulation with constant dose EGF.
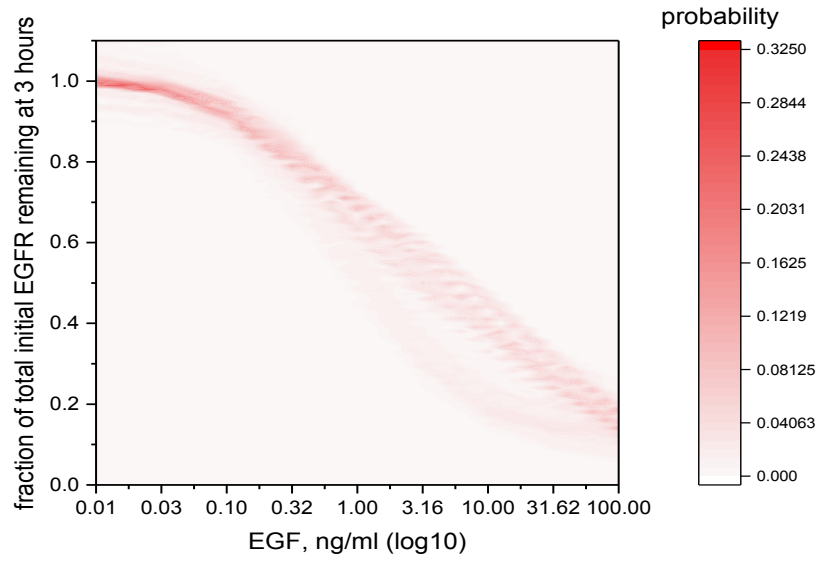
Figure 4.67: Predicted distribution of the dose-response of total EGFR abundances measured at three hours of stimulation with constant dose EGF.
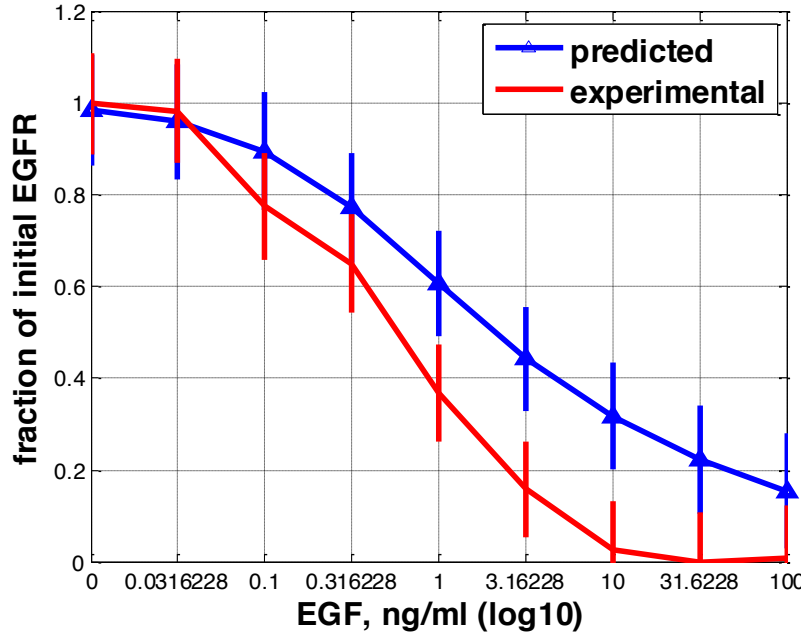
Figure 4.68: Dose-responses of total EGFR abundance at three hours of EGF stimulation: predicted vs.
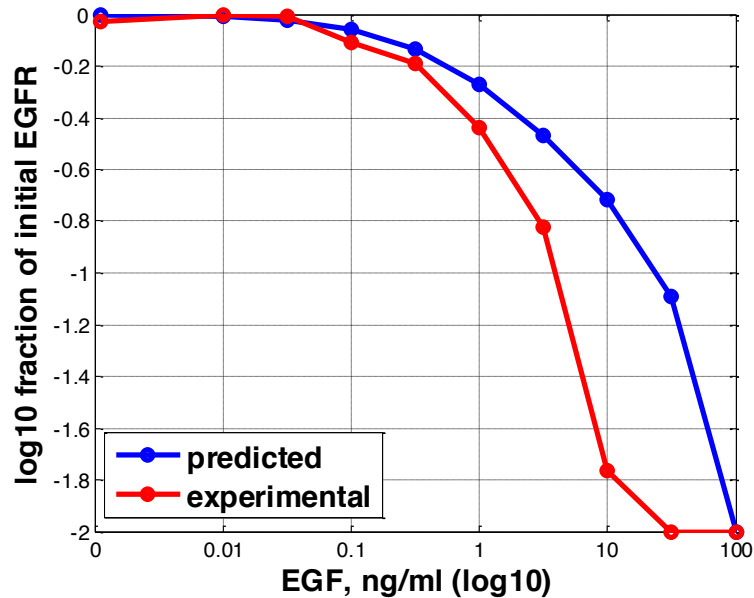
experimentally measured.

Figure 4.69: Dose-responses of the total EGFR remaining in the cells three hours after stimulation with constant dose EGF.

Experimental measurements of total EGFR time-courses (Figure 4.3) suggest that by three hours of stimulation with constant dose EGF most of the time-courses appear to adapt to their steady state levels. We then consider the experimentally measured dose response of amount of total EGFR receptors remaining in the system after three hours of stimulation (Figure 4.68, red). We find that at EGF doses between approximately ~0.03-10ng/ml EGF total EGFR remaining decreases with

increasing dose of EGF. Interestingly, at low EGF doses number of receptors does not appear to vary with different EGF doses; and at high EGF stimulations total number of EGFR remains similar after treatments with EGF on the order of 10-100ng/ml EGF. It is plausible that at EGF concentrations above 10ng/ml pEGFR signal reaches saturation (which might not necessarily be captured correctly by the pEGFR dose-response measured at 10 minutes) and that the maximal amount of EGFR receptors capable of participating in signaling is being removed. Therefore, at saturating EGF doses total EGFR does not reflect memory of the ligand concentration in the EGF treatment; at lower EGF doses the ligand doses are eliciting significant responses in pEGFR signal, and therefore do not elicit substantial degradation of receptors; however, for EGF doses in the range of ~0.03-10ng/ml, EGF doses appear to be remembered in number of receptors remaining undegraded, in inversely proportional to the EGF dose relationship. We plot the log-log plot of number of remaining EGFR at three hours of stimulation with EGF (Figure 4.69) and find that within the range of 0.03-10ng/ml the predicted and experimentally measured dose responses show negative slope of approximately ~-0.5..-0.75, suggesting that at steady state tEGFR dependence on EGF functionally might appear like Hill function in EGF.

We then hypothesize that the negative proportionality between EGF doses in the applied initial treatments of EGF and the number of EGFR remaining after the cells pre-adapted to the treatment with constant dose of EGF, could allow cells to sense relative changes in consequent stimulations with EGF and generate responses proportional to fold-changes, and not absolute changes, in EGF doses. In the next section we present experimental validation of the predicted Weber's Law behaviors in pERK responses and in Section 4.7 we present analytical derivation explaining mechanisms of Weber's Law realized through number of EGFR receptors remaining in the cells in long run of stimulation with constant doses of EGF.

Section 4.6.3: Experimental validation of Weber's Law on pERK responses to fold-changes in EGF

To experimentally validate predicted Weber's Law on pERK, we first have to measure peaks of the pERK responses when system pre-adapted to a certain EGF dose, is stimulated with a certain fold-increase in EGF dose, across a range of initial EGF doses. Next, we need to assess if total

EGFR in response to stimulation with range of different EGF doses at the time when EGF step is applied, exhibits a negative correlation with the initial ligand concentration, which could then be used to explain

Phospho-EGFR signal at low amount of stimulation even though can produce potent downstream signal on pERK is unfeasible to be reliably detected experimentally. Since pERK acts as a linear amplifier of pEGFR signal we can rely pERK levels for validation of Weber's Law prediction on level of pEGFR. Observing confirmation of Weber's Law at pERK level and the accompanying it negative relationship between total EGFR and initial dose of EGF applied could then imply presence of Weber's Law on pEGFR responses.

Based on total EGFR measurements collected at 3 hours of stimulation with range of EGF doses (Figure 4.3, Figure 4.68) we observe that tEGFR levels negative correlation with the initial doses of EGF applied change more dramatically between 0.03-10ng/ml EGF. Interestingly, this range of EGF doses is believed to be physiological in human serum and certain tissues (Rouger et al., 2014). Since, however, 10ng/ml EGF is on the upper limit of physiological EGF ranges and the fold-increases applied at 10ng/ml EGF are beyond physiological levels of EGF, we limit the selected range of EGF doses to ~0.03-3ng/ml to validate Weber's Law experimentally. We measure pERK responses to step inputs in EGF with initial EGF doses ranging between 0.03 and 3ng/ml. Note, that due to possible variation in number of cells per well between experiments conducted on different days, levels of EGFR expression on the cell surface might vary on different days, and this could lead to shifts in their dose responses to EGF. Therefore, we expand the range of EGF validation from 0.016 to 4ng/ml to ensure we cover the Weber's Law region and possibly detect the limits where it ends.

We treat the system with EGF dose shown on the X axis for three hours, and at 3 hours dose in the media is increased by a certain fold (x3, x6, x9, x16 of the initial EGF dose). We then measure pERK responses 10 minutes after the step increase in EGF (Figure 4.70). For comparison, we plot the 10 minutes pERK dose response at the initial treatment with EGF. We find that within ~0.03-0.3 ng/ml EGF range the dose responses to same fold-changes appear flat, whereas dose responses to different fold-changes (which started with the same dose of initial EGF treatment) appear distinct and

not overlapping in that region of EGF, suggesting that system responses to step increases in EGF depend on the ratio of EGF increase and not on its absolute level.

Note that the maximum pERK response detected in the dose response to initial EGF stimulation appears to be below the maximum response detected at the step increases, which seems controversial as in this experimental data set pERK responses appear to reach saturation in pERK in their initial EGF pre-treatments at ~ 3ng/ml, and over time, system's sensitivity is expected to decay, not increase. We believe that in the responses in which 10 minute measurements pERK were obtained, the maxima were possibly achieved prior to 10 minutes and thus were missed by the experimental measurements. Whereas, at lower levels of stimulations (e.g. Figure 4.3, pERK response at 0.03 ng/ml EGF) pERK responses decay from their peak at a slower rate, and thus, the pERK measurements at second step of EGF addition most likely do capture the peak responses.
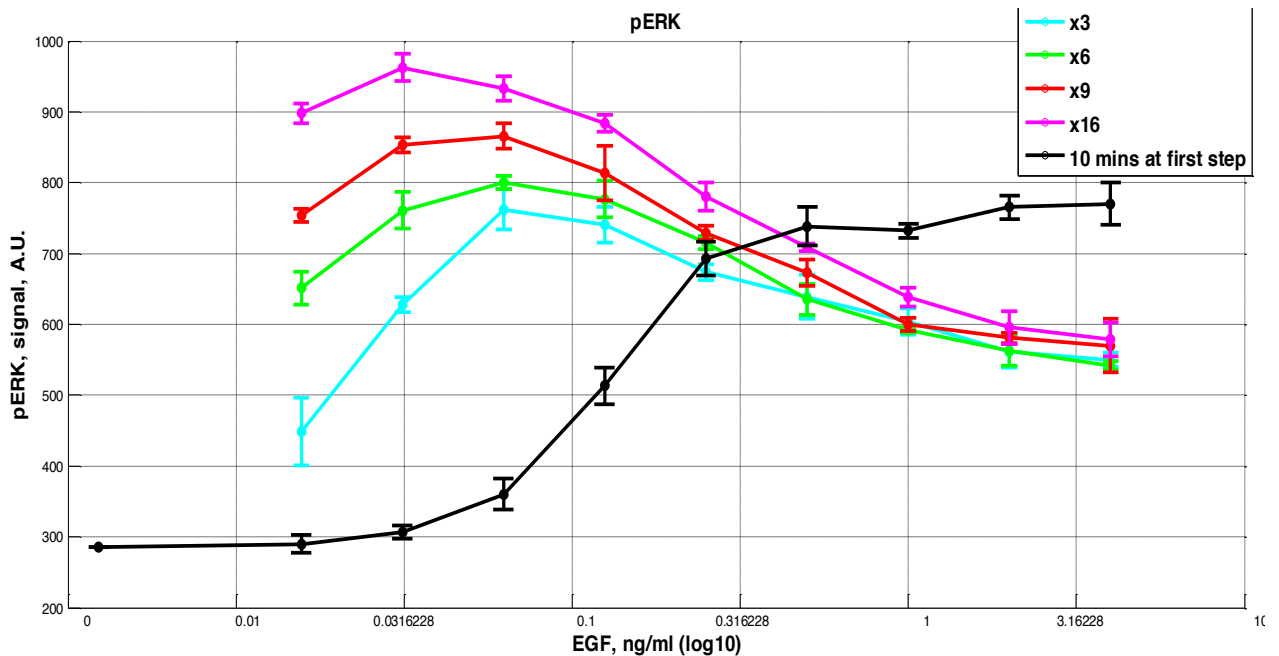
Figure 4.70: Experimental validation of Weber's Law at pERK responses: pERK responses measured 10 minutes after corresponding step inputs in EGF were applied three hours after stimulation with constant dose EGF.

The experimental data suggests (Figure 4.70) that although the initial EGF doses in the 0.03-0.3 ng/ml range vary almost by one order of magnitude and produce significantly different initial responses in pERK (relative difference ~700A.U.), when treated with same fold-changes in EGF doses at 3 hours, the system produces very similar responses, which differ only within ~ 100 A.U.

177

Three-hour time interval between initial EGF treatment and the fold-increase steps in EGF was chosen for the following reasoning. We aimed to minimize the time interval between the initial treatment and the step-increase to limit possible gene expression which might occur in response to initial stimulation with EGF. However, prior to step increases in EGF the system needed to be more or less adapted to the initial EGF stimulation. Based on the experimental data, pEGFR signal adapts to its pre-stimulus level within 1.5hrs of stimulation at highest doses of EGF. In addition, time of adaptation appears to be inversely dependent on the EGF dose: higher dose of EGF elicits phosphorylation response of higher amplitude, and decays faster. Such effect can be explained by rapid receptors internalization and degradation at higher EGF doses. However, at level of pERK, signal decays at a slower rate compared to the pEGFR rate of decay. Such difference in observed rate of signal decay can be partially due to signal amplification in MAPK kinase cascade, and due to differences in phosphatases acting on phospho-ERK and on phospho-EGFR. We observed that experimentally measured pERK by 1.5 hours of stimulation with highest EGF doses drops to 30-60% of its peak level, whereas by 3hrs it reaches anywhere between 10-40% of the peak level. Hence, as duration for treatments, we chose 3 hour time interval between the initial EGF treatment and time of addition of EGF fold increases. Based on our model simulations (results not shown) of comparing 1.5 hours interval to 3 hours interval, the 3 hour interval provided more time for the system variable of total EGFR to reach nearly steady state across most EGF doses; in addition, by 3 hours of stimulation with EGF, pERK responses decayed more significantly then by the 1.5 hours, and thus, responses to step-increases in EGF at 3 hours were more distinguished from the pERK levels before the step-increases in EGF.

We tried multiple distinct fold-increases in EGF doses to make sure the behavior we observe is not due to saturation of EGF ligand binding to available receptors. Since up to ~ 0.3 ng/ml EGF (Figure 4.70) higher fold-increases produce higher responses, the system is below saturation, as it is able to distinguish between different fold increases. In addition, we observe that at pre-treatments with EGF doses above 0.3ng/ml the system might be unable to discriminate between different folds due to saturation in EGF ligand binding to EGFR receptors.

178

Section 4.7: Analytical derivation of perfect adaptation and approximate Weber's Law in a simplified kinetic model

Fitted to the experimental data, our detailed dynamic ODE model (see Section 4.4) suggests a possibility of perfect adaptation on phospho-EGFR signal and approximate Weber's Law on phospho-EGFR and phospho-ERK. However, the main model is too unwieldy to explain the underlying causes of these phenomena. In order to understand what biological mechanisms are responsible for producing perfect adaptation and Weber's Law on phospho-EGFR, we developed a smaller kinetic model using simplifying assumptions on EGFR biology. In particular, the simplified analytical model focuses on the EGFR receptor layer and encompasses only the most essential interactions: ligand binding to the receptors, basal synthesis of receptors, their activation, dephosphorylation by phosphatases and degradation. The simplified analytical model can be considered as a reduction of the detailed model onto a parameter values subspace limited by the simplifying assumptions. As we demonstrate, the simplified model does allow a more intuitive understanding of the interplay between complex dynamic processes responsible for Weber's Law and perfect adaptation in ErbB signaling pathway.

We derive the models based on a generally applicable mass action kinetics formalism (Chen et al., 2010).

In Section 4.7.1, we describe the simplified model and present an analytical explanation of perfect adaptation on pEGFR. This is done by deriving analytical expressions for phospho-EGFR and total EGFR at steady state in response to stimulation with certain constant level of EGF. We thus show that perfect adaptation on pEGFR is a feature inherent to the topology of the simplified model and is independent of the parameters of the simplified model. Then, in Section 4.7.2, we examine the functional dependence of the steady state expression of tEGFR on the dose of applied EGF stimulation. Next, we sample probabilistically the parameter space of the simplified model, imposing Weber's Law in the model response over a certain physiological range of EGF stimuli. Using the resulting distributions of parameters, we show that at parameter values where the system exhibits approximate Weber's Law, tEGFR depends on the EGF dose through function inversely proportional to the EGF dose. Hence, the ability to remember EGF doses might be the feature enabling

computation of ratios of EGF doses at step inputs and thus leading to responses dependent on fold-changes, rather than on the absolute changes in EGF levels.

It is important to note that we did not fit the simplified model to the experimental data that we use to fit our detailed ODE model. Instead, we imposed certain requirements on the simplified model to capture general features of the signaling responses, such as, achieving perfect adaptation within a given time frame, and others. The simplified model can produce fits with the Weber's Law behavior, and thus can be used to provide a more intuitive way to understand mechanisms behind perfect adaptation and Weber's Law in the EGFR signaling.

Section 4.7.1: Derivation of perfect adaptation of phospho-EGFR and phospho-ERK signals

In this section we present a simplified analytical model to derive perfect adaptation on pEGFR and pERK signals. The simplified model assumes that EGF binding to a receptor is rapid and leads to its dimerization with an unoccupied receptor, followed by activation of the dimer. For simplicity, we assume that the abundance of dimers fully occupied by EGF is negligible in the system. Such assumption relies on the suggested decreased affinity of the second EGF molecule to a dimer with one EGF molecule (Macdonald & Pike, 2008). We omit the basal degradation of inactive receptors from the simplified model, as it is significantly slower than the degradation of active receptors. Given there is no basal degradation of inactive receptors in this model, its main limitation is that at steady state total number of receptors in the absence of ligand cannot be constant unless the basal rate of receptor production is very low.

The following are the variables in the simplified model:

- R is the total number of inactive EGFR receptors present in the cell.
- u is the concentration of EGF which is being maintained in the course of stimulation.
- B is the number of EGF-bound inactive receptors.
- D is the number of one-EGF-bound inactive dimer.
- P is the number of active (phosphorylated) receptor dimers.
- T is the total number of receptors.

We can then write down mass action equations of rate of change of each variable through its corresponding interactions:
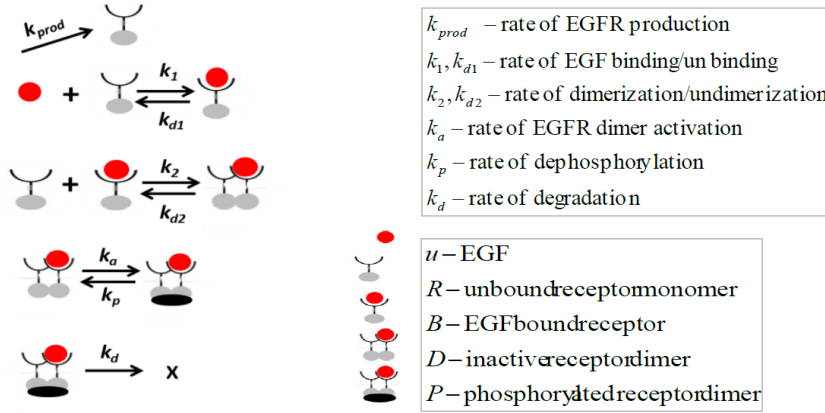
k_{prod} – rate of EGFR production
k_1, k_{d1} – rate of EGF binding/un binding
k_2, k_{d2} – rate of dimerization/undimerization
k_a – rate of EGFR dimer activation
k_p – rate of dephosphorylation
k_d – rate of degradation

u – EGF
R – unbound receptor monomer
B – EGF bound receptor
D – inactive receptor dimer
P – phosphorylated receptor dimer

$$
\begin{cases}
\dfrac{dR}{dt} = k_{prod} - k_1 u R + k_{d_1} B - k_2 B R + k_{d2} D \\[2mm]
\dfrac{dB}{dt} = k_1 u R - k_{d_1} B - k_2 B R + k_{d2} D \\[2mm]
\dfrac{dD}{dt} = k_2 B R - k_{d2} D - k_a D + k_p P \\[2mm]
\dfrac{dP}{dt} = k_a D - k_p P - k_d P
\end{cases}
\tag{1}
$$

Figure 4.71: A simplified analytical model of EGFR signaling for the derivation of perfect adaptation on pEGFR responses.

The total number of EGFR present in the system is the sum of all monomers and dimers remaining non-degraded and can be expressed as

$$
T = R + B + 2D + 2P \tag{2}
$$

We then write down an expression for dynamics of T by summing up expressions of dynamics for each of its component to obtain:

$$
\frac{dT}{dt} = k_{prod} - 2k_{degr} \cdot P \tag{3}
$$

181

This suggests that at steady state, when total number of receptors remains constant over time, equilibrium between constitutive receptor production and degradation of active receptors is reached. The same expression (3) suggests that the number of active receptors at a steady state is

$$P = \frac{k_{prod}}{2k_{degr}} \qquad (4)$$

and is independent of ligand concentration u, which, by definition, implies that P adapts perfectly.

Note that the model's conclusions regarding perfect adaptation of signal P do not change with a more elaborate implementation of receptors dimerization which would include formation of double-occupied with EGF dimers. Expression (3) also shows that the total number of receptors T, is controlled by an integral of P over time. Such mechanism in engineering is known to as integral feedback control (Alon et al., 1999). Hence, in the long run, integral of phospho-EGFR signal over time determines change in the total number of EGFR receptors remaining in the system.

We then derive perfect adaptation on pERK based on perfect adaptation on pEGFR and assumed linearity of pERK responses on EGF doses. We can write down the equation of dynamics of pERK via its activation through upstream kinase Y_i and deactivation by ERK phosphatase as follows:

$$\frac{d(pERK)}{dt} = k_{a_i} \cdot Y_i - k_p \cdot pERK \qquad (5)$$

where k_{a_i} is kcat of ERK phosphorylation and k_p is kcat rate of ERK phosphatase. Given the linearity in signaling cascade between pEGFR and downstream kinases leading to activation of pERK, steady state equation for levels of activated kinase Y_i can be expressed as $Y_i = c \cdot P$, where c is a constant and P is activation on pEGFR. In this model EGF is acting on pERK only through pEGFR and, hence, is not affecting directly any intermediate components in the cascade.

We obtain that at steady state $pERK = c \cdot P$ Therefore, given exact adaptation on pEGFR and linearity of the downstream signaling cascade we obtain that perfect adaptation on pERK follows. Similar reasoning applies to show perfect adaptation on pAKT.

Section 4.7.2: Steady state analysis of the EGFR signaling system and responses to step input in EGF: monomer-activated EGFR signaling model

In Sections 4.7.2 and 4.7.4 we consider simplified analytical models of EGFR signaling and conduct analysis of steady state behavior of the EGFR system in response to chronic stimulation with constant dose of EGF. The existence of steady state in mammalian cells in response to chronic stimulation with EGF has been previously reported and validated experimentally (Wiley & Cunningham, 1981).

In this section we consider analytical model of EGFR signaling with the assumption that ligand-driven dimerization of receptors happens instantaneously, and therefore, the unphosphorylated dimers could be neglected. This model assumes that phosphorylated receptor cies P are monomers, and the same notations as in Section 4.7.1 are used:
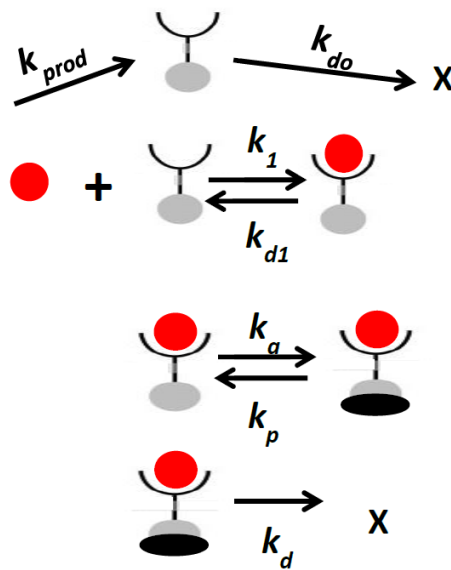
$$\begin{cases} \dfrac{dR}{dt} = k_{prod} - k_1 uR + k_{d_1}B - k_{do}R \\[2mm] \dfrac{dB}{dt} = k_1 uR - k_{d_1}B - k_a B + k_p P \\[2mm] \dfrac{dP}{dt} = k_a B - k_p P - k_d P \end{cases} \qquad (6)$$

We then apply a simplifying assumption that ligand binding to receptors is very rapid, therefore, the reaction of ligand binding and unbinding from the receptor can be assumed to be at quasi-equillibrium, and hence the amount of bound receptors can be expressed as follows:

$$B = \frac{k_1}{k_{d1}} uR \tag{7}$$

The simplified model can then be rewritten through dynamics of activated receptor P, ligand-free receptor R, and total number of receptors T as follows:

$$\begin{cases} \dfrac{dP}{dt} = k_a \dfrac{k_1}{k_{d1}} uR - k_p P - k_d P \\[2mm] \dfrac{dT}{dt} = k_{prod} - k_d P - k_{do} R \end{cases} \tag{8}$$

We then solve the system at steady state and express variables through R:

$$R = \frac{k_{prod}(k_p + k_d)}{(k_p + k_d)k_{do} + k_a \dfrac{k_1}{k_{d1}} u k_d}$$

$$B = \frac{k_1}{k_{d1}} uR \tag{9}$$

$$P = \frac{k_{prod}}{k_d} - \frac{k_{do}}{k_d} R$$

$$T = R + B + P = R + \frac{k_1}{k_{d1}} uR + \frac{k_{prod}}{k_d} - \frac{k_{do}}{k_d} R = \frac{k_{prod}}{k_d} + R\left(1 - \frac{k_{do}}{k_d} + \frac{k_1}{k_{d1}} u\right)$$

Based on the established fact that for the EGFR receptors the internalization and degradation of inactive receptors is significantly slower than the internalization and degradation of EGF-bound receptors (Sigismund et al., 2012), we have:

$$1 \ll \frac{k_d}{k_{do}} \tag{10}$$

Note that since in the simplified model receptors internalization, the step preceding receptors degradation is known to be significantly faster for EGF-bound receptors (Becker et al., 2010), k_d and k_{do} represent effective constants of receptors downregulation, and therefore can differ by more than the corresponding internalization rates do.

We then consider expression for T and apply simplification (10) to obtain:

$$T = \frac{k_{prod}}{k_d} + R\left(1 - \frac{k_1}{k_{d1}}u\right) \tag{11}$$

In the limit when ligand concentration u is low, most of the receptors will be in the inactive, ligand-free form R, and thus the total number of receptors could be expressed as

$$T = \frac{k_{prod}}{k_d} + R, \qquad u \sim 0 \tag{12}$$

In case u is very large, all receptors will be ligand bound or phosphorylated, and therefore R will be close to 0. Therefore, we obtain

$$T = \frac{k_{prod}}{k_d}, \qquad as \ \ u \to \infty \tag{13}$$

We then consider the case when u is not within these two limiting conditions. The steady state solution for R is given in (9) and we rearrange rate constants as follows:

$$R = \frac{k_{prod}(k_p + k_d)}{(k_p + k_d)k_{do} + k_a \frac{k_1}{k_{d1}}u k_d} =$$

$$= \frac{k_{prod}(k_p + k_d)/k_d}{(k_p + k_d)k_{do}/k_d + k_a \frac{k_1}{k_{d1}}u} =$$

$$= \frac{k_{prod}(1 + k_d/k_p)/k_d}{(1 + k_d/k_p)k_{do}/k_d + k_a \frac{k_1}{k_{d1}}u/k_p}$$

Since for pEGFR rate of phosphatase activity has been shown to be significantly faster than the degradation (Kleiman et al., 2011), we can assume $1 \gg k_d/k_p$ and therefore the expression for R can be simplified to:

$$R \sim \frac{k_{prod}/k_d}{k_{do}/k_d + k_a \frac{k_1}{k_{d1}}u/k_p}$$

In addition, dephosphorylation of EGFR by phosphatase was shown to be faster than the activation of the receptor, and we can consider such range of EGF doses, u, in which the following holds:

$$1 \ll \frac{k_p}{k_a \frac{k_1}{k_{d1}}u} \ll \frac{k_d}{k_{do}} \tag{14}$$

We apply the assumption (10) to derive a simplified expression for R and substitute it into expression (11) for T:

$$R = \frac{k_{prod}/k_d}{k_{do}/k_d + k_a \frac{k_1}{k_{d1}} u/k_p} \sim \frac{k_{prod}/k_d}{k_a \frac{k_1}{k_{d1}} u/k_p} = \frac{k_{prod}/k_d}{k_a \frac{k_1}{k_{d1}}/k_p} \frac{1}{u} \tag{15}$$

$$T = \frac{k_{prod}}{k_d} + R\left(1 - \frac{k_1}{k_{d1}} u\right) = \frac{k_{prod}}{k_d} - \frac{k_{prod} k_p}{k_d k_a} + \frac{k_{prod}/k_d}{k_a \frac{k_1}{k_{d1}}/k_p} \frac{1}{u}$$

Applying the left-most inequality in (14) we obtain:

$$T = \frac{k_{prod}}{k_d} + \frac{k_{prod}/k_d}{k_a \frac{k_1}{k_{d1}}/k_p} \frac{1}{u} \tag{16}$$

Therefore in the considered region of u the total number of remaining receptors, T, will be inversely proportional to ligand dose u up to a constant. We then consider a scenario in which the system has reached steady state to stimulation with constant dose of EGF u_o and the EGF dose is increased in a step input to u_1 where $u_1 > u_o$. We derive expression for the maximum of the response P_{max}, representing total maximal phosphorylation of the receptors the system could produce in response to the step input $u_o \rightarrow u_1$. The dynamics of phosphorylated receptors to the step input in EGF is given by the first equation of (8):

$$\frac{dP}{dt} = k_a \frac{k_1}{k_{d1}} u_1 R - k_p P - k_d P \tag{17}$$

We consider phosphorylation response right after the system adapted at steady state to u_o is treated with input step increase $u_o \rightarrow u_1$. At the time the signal reaches the maximum, P_{max}, derivative (17) is 0 and can be written as:

$$0 = k_a \frac{k_1}{k_{d1}} u_1 R - k_p P_{max} - k_d P_{max} \tag{18}$$

Since at early times of signaling response, degradation of active receptors is insignificant and most of signal downregulation happens through dephosphorylation by strong phosphatases, we can simplify equation (18) by omitting the $k_d P_{max}$ term:

$$k_a \frac{k_1}{k_{d1}} u_1 R - k_p P_{max} = 0 \tag{19}$$

$$P_{max} = k_a \frac{k_1}{k_p k_{d1}} u_1 R \tag{20}$$

Since we assume that while signal response reaches its peak P_{max}, no significant degradation of receptors occurred and that the total number of receptors is as at steady state in response to u_o we have: $P_o + R_o = P_{max} + R_{max}$. Therefore, when the response reaches P_{max} the number of ligand unbound receptors R_{max}, assuming P_o was negligible at steady state, is $R_{max} = R_o - P_{max}$, and applying assumptions from (14) we get:

$$P_{max} = k_a \frac{k_1}{k_p k_{d1}} u_1 R_o - P_{max} * k_a \frac{k_1}{k_p k_{d1}} u_1 \tag{21}$$

$$P_{max} = k_a \frac{k_1}{k_p k_{d1}} u_1 R / ((1 + k_a \frac{k_1}{k_p k_{d1}} u_1)) \sim k_a \frac{k_1}{k_p k_{d1}} u_1 R_o \tag{22}$$

R_o is given by expression (15):

$$R_o = \frac{k_{prod}/k_d}{k_a \frac{k_1}{k_{d1}}/k_p} \frac{1}{u_o}$$

Substituting it into (22) shows that

$$P_{max} = k_a \frac{k_1}{k_p k_{d1}} u_1 \frac{\frac{k_{prod}}{k_d}}{k_a \frac{k_1}{k_{d1}}} \frac{1}{u_o} = \frac{k_{prod}}{k_d} \frac{\boldsymbol{u_1}}{\boldsymbol{u_o}} \tag{23}$$

We showed that maximum of the signal in response to step input $u_o \to u_1$ up to a constant is proportional to the fold-change in the ligand doses of this step increase, $\frac{u_1}{u_o}$, and therefore implies that phosphorylated responses exhibit Weber's Law in ligand concentrations.

Given Equality (23), we can then compare maximum signal responses if the system is treated with two different step inputs, $u_o \to u_1$ and $u_o \to u_2$ where, $u_1 < u_2$:

$$P_{max}(u_2) - P_{max}(u_1) = \frac{k_{prod}}{k_d}\frac{\boldsymbol{u_2}}{\boldsymbol{u_o}} - \frac{k_{prod}}{k_d}\frac{\boldsymbol{u_1}}{\boldsymbol{u_o}}$$

$$= \frac{k_{prod}}{k_d}\left(\frac{\boldsymbol{u_2}}{\boldsymbol{u_o}} - \frac{\boldsymbol{u_1}}{\boldsymbol{u_o}}\right)$$

(24)

Therefore, the difference between the amplitudes of the system responses to two step inputs in EGF is proportional to the difference in fold-changes in the two inputs, given that before the step inputs the system was pre-adapted to steady state at basal stimulation with ligand concentration u_o. Interestingly, expression (24) also suggests that higher the rate of basal EGFR production, or, slower the rate of degradation of active receptors, higher the resolution between different folds that the system can sense.

It is important to emphasize that the current derivation relies on the assumptions (10) and (14), from which we can derive limits on ligand concentrations at which we would expect Weber's Law to hold:

$$1 \ll \frac{k_p}{k_a \frac{k_1}{k_{d1}} u}$$

$$u \ll \frac{k_p}{k_a \frac{k_1}{k_{d1}}} = \frac{k_p}{k_a} K_{eq}$$

Therefore, given the ligand concentration is sufficiently low, and below K_{eq}, where K_{eq} is equilibrium rate constant of EGF unbinding from receptors, conditions (14) are satisfied. In other words, Weber's Law will hold in the range of EGF doses at which there is no saturation of EGF ligand binding to the receptors, and the phosphorylation response P will increase proportionally to the increase in ligand dose u.

For EGFR signaling at EGF doses below K_{eq} equilibrium unbinding constant, most of signaling occurs through EGFR dimers with only one EGF bound molecule, supporting plausibility of signaling being linear in EGF dose of stimulation in that regime.

Note that the simplified model considered in this section omits internalization of active and inactive receptors. However, the overall results regarding Weber's Law in the system responses would not change if internalization were included. The steady state expressions for the number of

ligand-free and phosphorylated receptors in the model with internalization have been previously derived (Zi & Klipp, 2007).

The simplified model shows that information about EGF dose of current stimulation is stored in the number of EGFR receptors remaining at steady state. According to the derived formula for the number of receptors at steady state, the higher the initial EGF treatment, lower the amount of receptors remaining. Overall, cell response to EGF treatment depends on the actual EGF dose applied and on the amount of available EGFR receptors present at the time of step input. If initial dose of EGF was not very high, fewer receptors get activated and removed from the cell; whereas at initial treatment with high EGF, more receptors are removed from the cell. Intuitively, to elicit a certain level of response to a step input in EGF, in cells adapted to low initial EGF, one would have to apply relatively less EGF, than if the cells were adapted to high initial EGF and thus had less receptors. This reasoning helps to understand how system in states adapted to stimulation with two different doses of constant EGF, could have very different number of receptors at steady state, however, if treated with very different absolute amounts of EGF, yet, same fold changes, could produce similar responses.

In the fits where Weber's Law is observed in analytical model, total EGFR stores memory of previous EGF doses, which could enable the system to compute a ratio of the step increase in EGF concentration, and thus generate a fold-change dependent response in a certain range of EGF doses. Therefore, the memory of EGF doses retained in total number of receptors might be instrumental in realizing the Weber's Law in EGF in the EGFR signaling pathway.

Section 4.7.3: Experimental validation of predictions of the monomer-activated EGFR signaling model

In this section we show predictions and experimental validations of the simplified analytical model of EGFR signaling derived in Section 4.7.2. We extend the analytical model presented in Section 4.7.2 of linear-activated EGFR signaling to include internalization and recycling of active and inactive receptors. Given that the rate of recycling of receptors is negligible, the results obtained from this model do not differ significantly from the model presented in Section 4.7.2 and are consistent with similar derivations presented in literature (Zi & Klipp, 2007).

To generate predictions of the analytical model we substitute parameter values derived from literature (Table 6) into the analytical solutions of the steady state values for total EGFR in response to stimulation with constant dose EGF across a range of EGF doses. The rate of EGFR production is taken such that the number of receptors at 0 EGF stimulation at steady state agrees with our experimental estimate of EGFR abundance in MCF-10A cells and is on the order of $3x10^5$ - $6x10^5$ receptors per cell. The parameter k_1 for the rate of EGF binding to receptors represents a compounded value of the product of rate of binding, phosphorylation of receptors and EGF unbinding; reaction of EGF binding and unbinding is assumed to be at quasi equilibrium. We plot the predicted normalized dose response of total EGFR (Figure 4.72) and compare it to the experimentally measured total EGFR abundance in MCF-10A cells treated with EGF dose for three hours. The resulting model predictions at the literature-derived parameter values agree with the experimentally measured dose response of total EGFR.

Table 6: Literature-derived parameter values of the analytical model.

Parameter	Description	Value	Reference
k_{prod}	rate of EGFR production	1.1220 rec* sec^{-1}	(fitted)
k_1	effective rate of EGF binding to EGFR and activation of receptor	19.0546 nM^{-1} sec^{-1}	(fitted)
k_i	rate of EGFR internalization	0.00033 sec^{-1}	(Zi & Klipp, 2007)
k_{ri}	rate of EGFR recycling	0.0033 sec^{-1}	(Zi & Klipp, 2007)
k_{do}	rate of EGFR degradation	1.6834e-05 sec^{-1}	(Zi & Klipp, 2007)
k_a	rate of EGF-EGFR internalization	0.0025 sec^{-1}	(Wiley et al., 1991)
k_{ra}	rate of EGF-EGFR recycling	0.0033 sec^{-1}	(Zi & Klipp, 2007)
k_d	rate of EGF-EGFR degradation	1.7783e-04 sec^{-1}	(Zi & Klipp, 2007)
k_p	rate of EGF-EGFR dephosphorylation by phosphatases	0.01-0.06 sec^{-1}	(Kleiman et al., 2011)
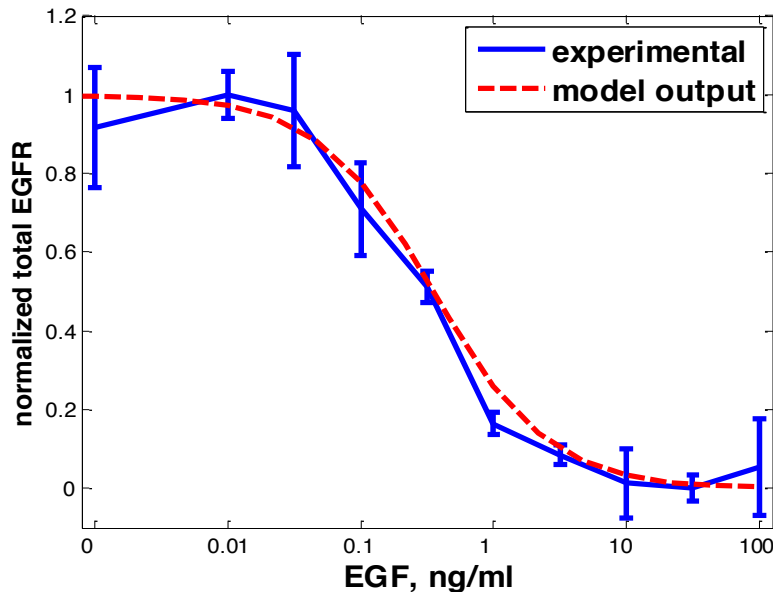
Figure 4.72: Experimentally measured dose response of the total EGFR abundance three hours of EGF stimulation in MCF-10A cells and the prediction of the analytical model. Plotted on the Y axis is the normalized amount of total EGFR obtained by subtraction of total EGFR measured at saturating EGF doses (100ng/ml) and normalization of the obtained curve by its maximum.

We then use the simplified analytical model with the substituted literature-derived parameter values to numerically compute transient dynamic responses to step-inputs in EGF applied when system is at steady state to chronic stimulation with EGF. Since the analytical model predicts phosphorylation responses of EGFR, we use a transfer function to map the corresponding pEGFR response to pERK responses. pERK responses are feasible to validate experimentally compared to the pEGFR responses, and are possibly more relevant directly to the cell fate decisions. The pEGFR-pERK transfer function is obtained by fitting a Hill function from the dose responses of pEGFR and pERK measured in MCF-10A cells (Figure 4.4).

We numerically simulate pEGFR responses to step inputs in EGF with various fold increases (x1.5, x2, x3, x6) in EGF doses, compute the maximal values of the transient pEGFR responses and convert them into the corresponding pERK responses using the transfer function. We plot the dose response of pERK to the initial EGF treatment applied to system at steady state with 0 EGF (green) (Figure 4.73) and the dose responses of maximal pERK to a given fold-change increase in EGF. For comparison we also show the predicted and experimentally measured total EGFR dose response. Model predicts that in the physiological range of EGF doses (approximately 0.03-3ng/ml EGF),

maximal dose responses of pERK to step inputs with the same fold change appear flat across at least an order of magnitude of initial EGF doses. In addition, the model predicts the dose responses to different fold-changes to be distinct from each other in this range.
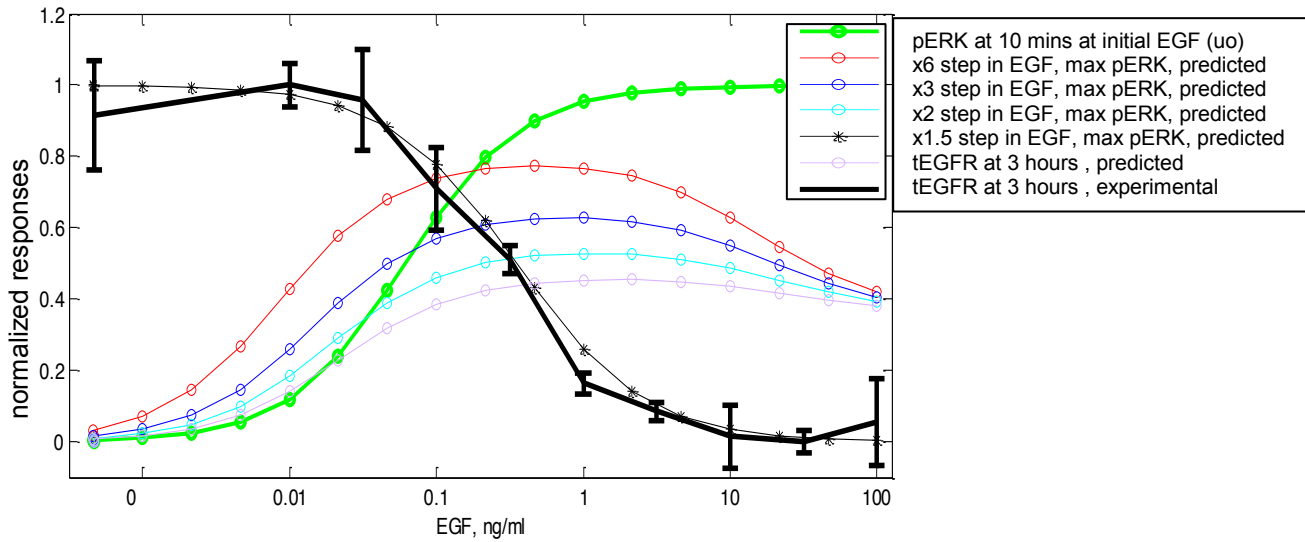
Figure 4.73: Predicted dose responses of the maximal pERK responses to step-inputs in EGF at a given step increase in EGF.

We then experimentally validate these predictions of the analytical model and measure pERK responses in MCF-10A cells treated with EGF for three hours and at three hours treated with a step input in EGF with the same fold-change across all the doses. Figure 4.74 shows experimentally measured time-courses of pERK responses in MCF-10A cells, which were incubated at 0 EGF and at time=0 were treated with physiological levels of EGF(0.16 or 0.8 ng/ml). At time=3 hours in each case EGF concentration was increased 5-fold. The data demonstrates that despite significant difference in transient initial responses in pERK to the two different concentrations of EGF, given the fold-change in the step input is the same, the peaks of the produced responses are strikingly similar. Such result further supports presence of Weber's Law in EGFR signaling system in the physiological range of EGF.
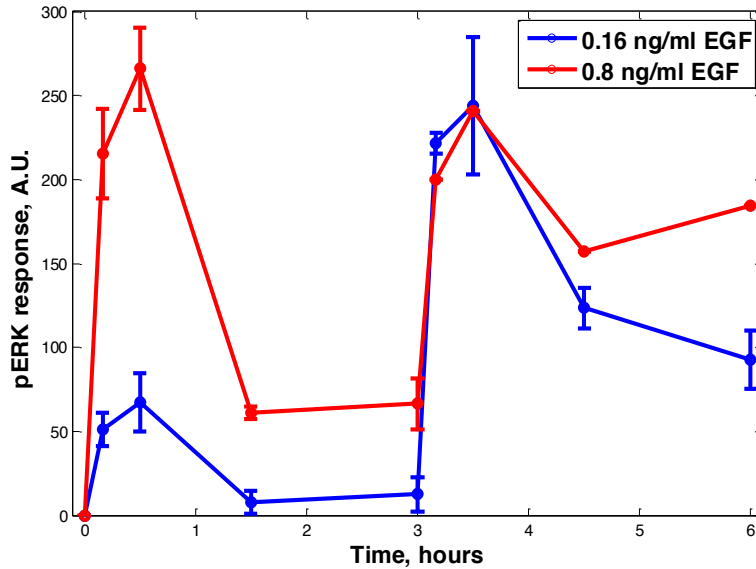
Figure 4.74 Experimentally measured pERK responses in MCF-10A cells treated with a x5 fold step input in EGF applied three hours after initial EGF addtion. At t=0 cells were treated with 0.16 ng/ml (blue) or 0.8 ng/ml EGF (red) and at t=3 hours EGF concentration in each case has been increased x5 fold. Most of the points represent the averages and the standard errors of duplicate experiments; 0.8ng/ml EGF treatments after three hours were measured in one replicate.

We further validate the predicted Weber's Law in the EGFR system and measure pERK responses to different step inputs in EGF. In particular, we measure pERK responses at 10 minutes after 3-fold and 6-fold step increase in EGF applied at 3 hours of stimulation with initial dose of EGF from the physiological range. Given our previous experimental observations, pERK response at 10 minutes is close to the maximal response of pERK to the step input. Figure 4.75 shows numerically computed 10 minute dose response and peaks of the pERK responses to 3-fold and 6-fold step increase in EGF with initial EGF stimulations within a physiological range and the corresponding experimentally measured pERK dose responses at 10 minutes. Model simulations (Figure 4.75) suggest that at higher EGF doses, transient responses might reach their peak faster, therefore, the measured 10 minutes dose response might differ from the maximal response. Still, the fact that model's predictions of the 10 minutes pERK dose response agree reasonably well with the experimentally measured values, suggests that the maximal dose responses of pERK to different fold increases will be distinct and flat in the physiological range of EGF. The fact that the predicted dose responses at 6-fold increase are distinct from the 3-fold increase eliminates possibility of flatness of the pERK response

at 3-fold increase to be caused by the saturation of pERK responses, as higher fold-increase in EGF can still produce a higher pERK response.
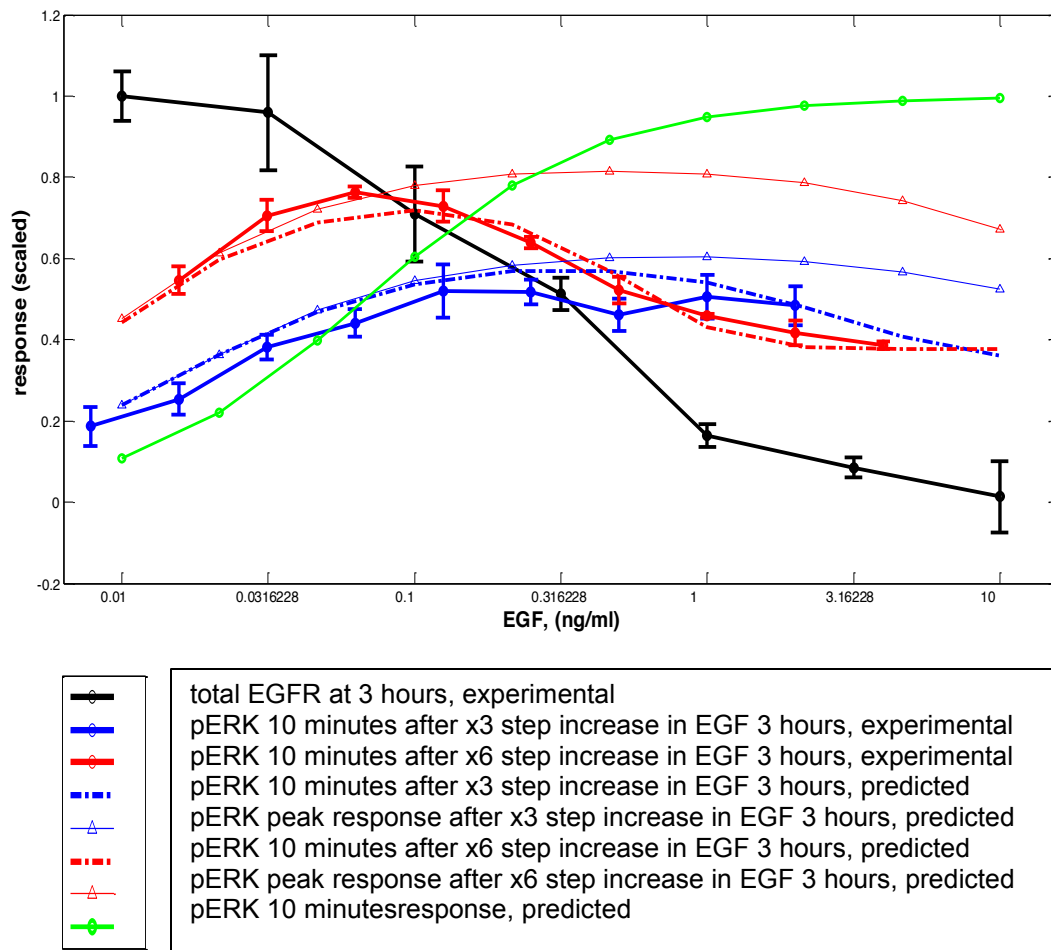
Figure 4.75: Experimental validation of Weber's Law in pERK responses.

Section 4.7.4: Probabilistic parameter scan of simplified analytical model

We explore parameter space of the presented in Section 4.7.2 simplified model implementing monomer-activation of EGFR receptors. We run MCMC simulations as described in Section 3.2.1, with the goal of finding parameter dependencies which guarantee Weber's Law in the system's responses to fold-increases in EGF. We do not fit the simplified model to the full experimental data; instead, we fit the analytical model to the total EGFR dose response (Figure 4.72) and compute the coefficient of variation in maximal peak of system's responses to step inputs with the same fold changes in EGF and initial levels of EGF from the physiological range (0.1-3ng/ml). More specifically, we simulate a scenario in which system pre-adapted to stimulation with a constant level of EGF, u_0, at

time=0 is treated with a step input in EGF $\alpha \times u_0$ (where $\alpha > 1, u_0 > 0$). We require that the system's responses decay from their maximal peak. We simulate system's responses to three steps with the same fold increase, $\alpha = 5$, at different initial EGF doses in the physiological range 0.1-3ng/ml. In all fits we require that the model's phosphorylation responses to each of the step inputs are non-trivial and rise at least 5% above its initial steady state level. Objective function used to perform MCMC simulations computes the distance between the peaks of the responses to the three step inputs and is described in Section 3.2.3.The more similar the peaks are, the lower the distance between them and the higher the value of the objective function. The MCMC chains are preferentially sampling model's parameter space with higher objective function. It is important to note that we do not require that the timing of the peaks of the responses at the EGF steps inputs of the simplified model are the same, we only require that their values are similar

We derive parameter distributions in which Weber's Law holds for the simplified model. Priors on the parameter values were derived based on literature estimates (see Table 6), however we allow parameters to vary within 1-2 orders of magnitude from these estimates. We then plot the ratios of parameter values $\frac{k_p}{k_a \frac{k_1}{k_{d1}} u}$ and $\frac{k_d}{k_{do}}$ in the sampled points and highlight the points with better Weber's Law score in the peaks of responses (Figure 4.76). The results suggest that in the fits with more similarity in the peaks of their Weber's Law responses (shown in magenta, Figure 4.76), the ratio of degradation of active receptors to the rate of degradation of inactive receptors ($\frac{k_d}{k_{do}}$) has to be high and exceed the ratio $\frac{k_p}{k_a \frac{k_1}{k_{d1}} u}$. In addition, in the fits with best Weber's Law score, the latter ratio has to exceed 1. Therefore, the model simulations demonstrate that the Weber's Law conditions, derived analytically, are indeed reflected in the relationships in model's parameter space.

We find that Weber's Law holds in less than 1% of the fits and that this number is significantly smaller compared to the number of fits in which nearly-perfect adaptation holds. This result is consistent with similar studies which reported that approximate Weber's Law holds only in about 0.15% of fits with perfect adaptation (Skataric & Sontag, 2012) (note that in their study the authors

used a less constrained requirements for Weber's Law and imposed it across a narrow range of initial ligand doses).

Overall, we find that even in the parameter space of the simplified analytical model, fraction of fits with Weber's Law is very low. Therefore, working with a more detailed ODE model of ErbB signaling with a larger parameter space would not necessarily allow us to find parameter values which would fit well experimental data and demonstrate Weber's Law in phosphorylation responses. Hence, a strategy of using smaller, analytical model with literature-derived parameters is more suitable to explore Weber's Law in the EGFR signaling system.
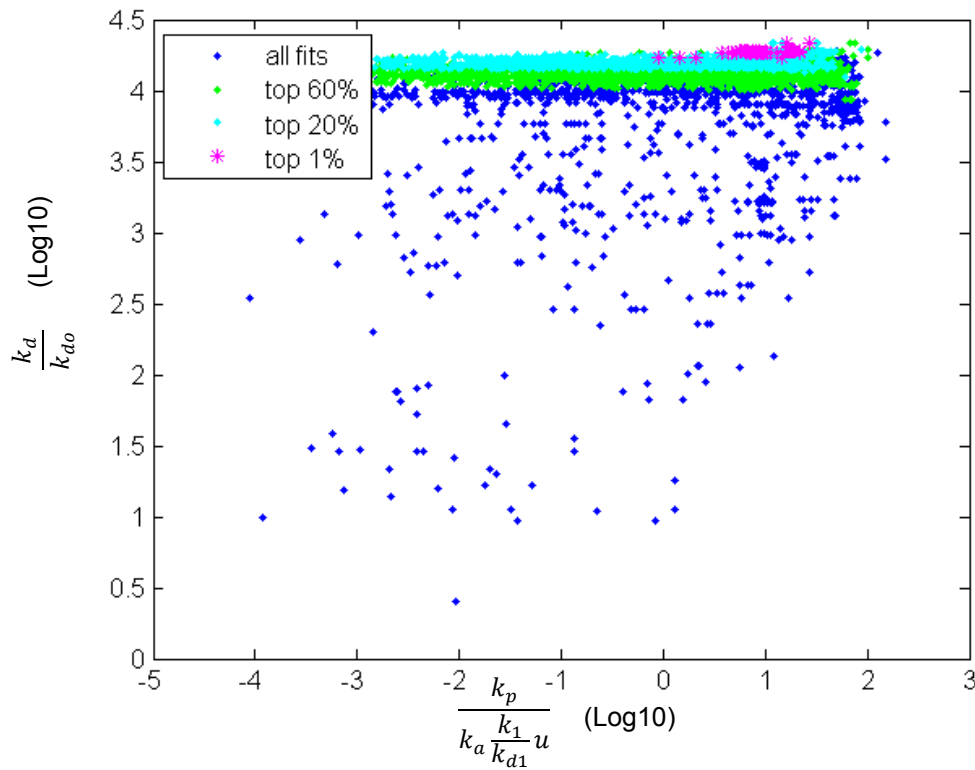
Figure 4.76: Relationship between parameters in the fits with Weber's Law. Plotted are the ratios of phosphatase activity to the term representing effective activation of receptors (X axis) and ratio of degradation of active receptors to inactive (Y axis). Points shown in blue are all simulated fits. Highlighted in different colors are the fits ranked by their Weber's Law score in responses to step inputs in EGF with initial dose being in a physiological range of EGF.

Section 4.7.5: Regimes of the total number of EGFR receptors at steady state in chronic stimulations with EGF

In Section 4.7.2 we analytically derived expressions of total EGFR number of receptors at steady state in response to treatment with a constant dose of EGF. Experimentally it can be often

unfeasible to precisely quantify abundances of receptors available for ligand binding on the cell surface, and thus establish the number of receptors actively participating in cell signaling. Moreover, quantifying amounts of the available receptors is obscured partially due to problem of subtracting background fluorescence offset from measurements of total EGFR abundances. Our theoretical analysis allows to estimate abundances of receptors available for ligand binding in cells adapted to treatments with chronic EGF doses. Moreover, our analysis allows to derive analytical expressions of abundances of active receptors through various cell parameters, such as, receptors production or basal degradation. We then consider the resulting regimes in dependencies of total EGFR available for ligand binding and the doses of the applied EGF stimulation (Figure 4.77).

If EGF concentrations were very low, we derive that most of the receptors should be ligand-free and available for ligand binding and their total number is given as

$$T = \frac{k_{prod}}{k_d} + R, \qquad u \sim 0 \tag{25}$$

In the case when u is very large, all receptors will be ligand bound or phosphorylated, and therefore R will be ~0. Then

$$T = \frac{k_{prod}}{k_d}, \qquad as \ \ u \to \infty \tag{26}$$

We showed that In the middle range of EGF doses number of receptors remaining in the cell will always be a decreasing function of the EGF dose, suggesting that the higher the EGF dose was applied, the lower the number of receptors remains at steady state:

$$T = \frac{k_{prod}}{k_d} + \frac{k_{prod}/k_d}{k_a \frac{k_1}{k_{d1}}/k_p} \frac{1}{u} \tag{27}$$

Mechanistically, this is explained by the fact that at higher initial EGF treatment more receptors can get activated and be removed from the cell.

Overall, the dependence of total EGFR steady state levels on EGF doses described by (25), (26), and (27) agrees well with the experimentally measured tEGFR dose-response curve at 3 hours of stimulation with EGF measured in MCF-10A cells (Figure 4.68). By 3 hour time most of phospho-

EGFR time-courses have decayed to 0 and the system can be assumed to be approximately at steady state.

The presented here analytical derivation can then be used to estimate, for example, number of receptors actively participating in cell signaling by subtracting the amount of total EGFR measured at chronic stimulations with saturating doses of EGF from the measured steady state total EGFR abundances at no EGF stimulation. Such method allows to quantify fluorescence corresponding to the receptors actively participating in signaling, while eliminating the necessity of establishing precise amount of fluorescence corresponding to the background offset, which might be challenging experimentally (see Section 4.1.3).
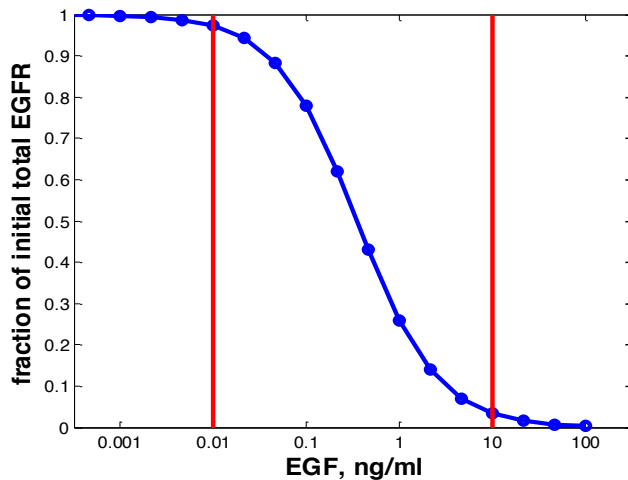
Figure 4.77: Three regimes in the total EGFR remaining at steady state to chronic stimulations with EGF.

Section 4.7.6: Steady state analysis of the EGFR signaling system and responses to step input in EGF: dimer-activated EGFR signaling model

In this section we derive a simplified analytical model of EGFR signaling which implements dimerization of receptors to demonstrate mechanism and conditions under which system responses would exhibit Weber's Law. We do this through first deriving steady state expressions of ligand-bound and ligand-free receptor species and considering maximal responses of system which has been pre-adapted to stimulation with a constant stimulus and is then treated with a step-input in EGF. The goal is to quantify amounts of total receptors remaining when system has reached steady state to stimulation with constant dose of ligand and investigate under what parameter constraints and ligand doses the system's response to step input applied to pre-adapted system is proportional to fold-

change in that step-input. We apply simplifications to the model expressions based on known biochemical properties of ErbB receptors and their downregulation processes and limit EGF ranges to physiologically relevant regions.

Other works have previously derived steady state expression for total unbound EGFR and activated receptors in simplified analytical models (Zi & Klipp, 2007). We specifically use the steady state expressions for amount of EGF-free receptors to derive estimates of peak responses to step increases in EGF concentrations applied when system has pre-adapted to non-zero EGF stimulation.

The model encompasses the following processes: basal receptors production k_{prod} and degradation of inactive receptors, reversible ligand binding of EGF (u) to receptors (k_1, k_{d1}), receptors dimerization (k_2, k_{d2}), activation (k_a) and deactivation through phosphatase (k_p) and degradation of active receptors (k_d). We assume in the model there are only dimers with one EGF bound, an approximation, which should hold if EGF concentration is limited.

k_{prod} – rate of EGFR production
k_1, k_{d1} – rate of EGF binding/unbinding to unoccupied EGFR monomers
k_2, k_{d2} – rate of dimerization/undimerization
k_a – rate of EGFR dimer phosphoyrlation
k_p – rate of dephosphorylation of receptors by phosphatase
k_d – rate of degradation of active receptors
k_{d0} – rate of degradation of inactive receptors

u – EGF
R – unbound EGFR receptor monomer
B – EGF bound receptor
D – inactive receptor dimer
P – phosphorylated receptor dimer
T – total amount of receptors

$$\begin{cases} \dfrac{dR}{dt} = k_{prod} - k_1 uR + k_{d_1} B - k_2 BR + k_{d2} D - k_{do} R \\[2mm] \dfrac{dB}{dt} = k_1 uR - k_{d_1} B - k_2 BR + k_{d2} D \\[2mm] \dfrac{dD}{dt} = k_2 BR - k_{d2} D - k_a D + k_p P \\[2mm] \dfrac{dP}{dt} = k_a D - k_p P - k_d P \end{cases} \tag{28}$$
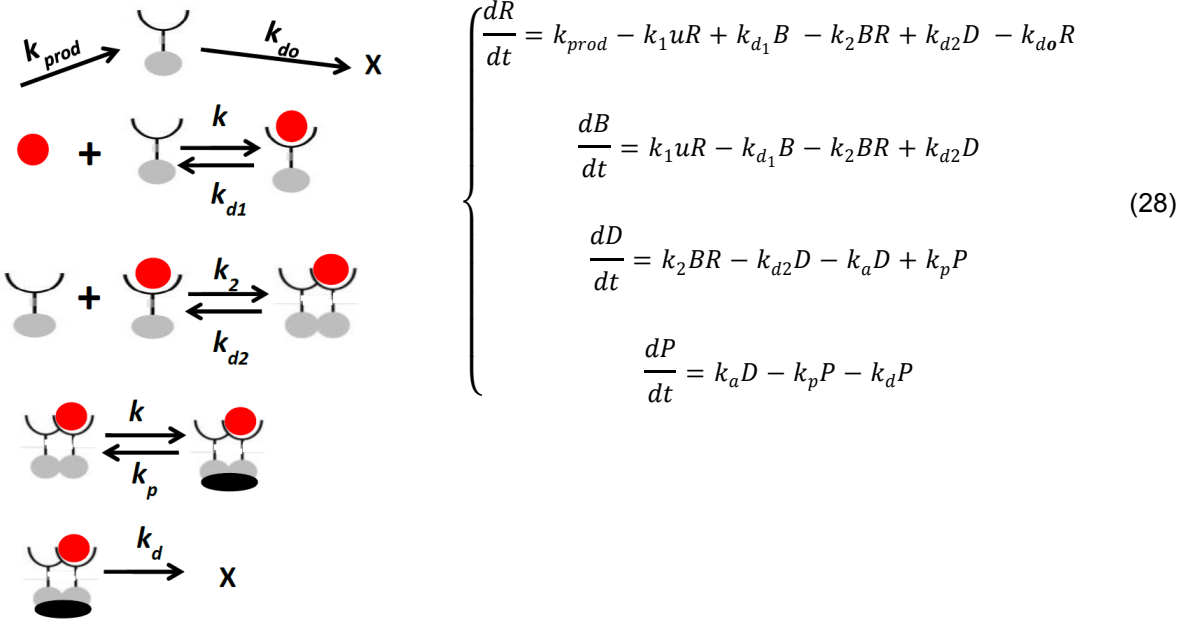
The total number of receptors can be written as follows:

$$T = R + B + 2*D + 2*P$$

and its dynamics is described by the following expression:

$$\frac{dT}{dt} = k_{prod} - 2*k_d P - k_{do} R \tag{29}$$

Note, that if signal activation is weak, at steady state most of receptors will remain unphosphorylated and therefore $P \sim 0$. In the case of strong activation in the pathway, at steady state all receptors will be phosphorylated and $R \sim 0$ and, hence, $P = \frac{k_{prod}}{2*k_d}$. However, this condition might not in fact be perfect adaptation, but merely a result of signal saturation at higher EGF doses. In all other cases, given that the rate of degradation of activated receptors is significantly higher than the rate of degradation of inactive receptors $k_d \gg k_{do}$, neglecting the term $k_{do} R$ in (29) and leading to $P = \frac{k_{prod}}{2*k_d}$. This suggests that steady state value of P will be independent of ligand concentration, which by definition is perfect adaptation of signal P.

We then apply series of simplifying assumptions, namely, that EGF binding to monomers is fast and thus can be assumed to be at quasi-equilibrium:

$$k_1 uR = k_{d_1} B \tag{30}$$

Next, we assume that ligand binding to EGFR leads to rapid dimerization and activation of receptors:

$$k_2BR - k_{d2}D - k_aD = 0 \tag{31}$$

From (30)and (31) we can express B and D as:

$$B = \frac{k_1}{k_{d1}}uR$$

$$D = \frac{k_2}{k_{d2}+k_a}BR = \frac{k_2}{k_{d2}+k_a}\frac{k_1}{k_{d1}}uR^2$$

The system can then be simplified as follows:

$$\begin{cases} \dfrac{dR}{dt} = k_{prod} - k_a\dfrac{k_2}{k_{d2}+k_a}\dfrac{k_1}{k_{d1}}uR^2 - k_{do}R \\[2mm] \dfrac{dP}{dt} = k_a\dfrac{k_2}{k_{d2}+k_a}\dfrac{k_1}{k_{d1}}uR^2 - k_pP - k_dP \end{cases} \tag{32}$$

If we assume that at certain range of u at steady state degradation of inactive receptor is negligible compared to its consumption through the activation step, $u \gg \frac{k_{d2}+k_a}{k_ak_2}\frac{k_{d1}}{k_1}\frac{k_{do}}{R}$, we can drop $k_{do}R$ term in (32) and solve for R at steady state:

$$R = \sqrt{\frac{k_{prod}}{k_a\dfrac{k_2}{k_{d2}+k_a}\dfrac{k_1}{k_{d1}}u}} \tag{33}$$

Substituting (33) and solving the steady state equation for P we have:

$$P = \frac{k_a\frac{k_2}{k_{d2}+k_a}\frac{k_1}{k_{d1}}uR^2}{k_p+k_d} = \frac{k_a\frac{k_2}{k_{d2}+k_a}\frac{k_1}{k_{d1}}u}{k_p+k_d}\frac{k_{prod}}{k_a\frac{k_2}{k_{d2}+k_a}\frac{k_1}{k_{d1}}u} = \frac{k_{prod}}{k_p+k_d}$$

Steady state solutions for this system then are as follows:

$$\begin{cases} R = \sqrt{\dfrac{k_{prod}}{k_a\dfrac{k_2}{k_{d2}+k_a}\dfrac{k_1}{k_{d1}}u}} \\[4mm] P = \dfrac{k_{prod}}{k_p+k_d} \end{cases}$$

Therefore, when the system has reached steady state to stimulation with constant dose of EGF, u_o, the model can show $R \sim \frac{1}{\sqrt{u_o}}$ dependence of the amount of unbound receptor on EGF concentration, and perfect adaptation on the signal P. We then derive expression for maximal signal P_{max} when the system adapted to steady state to stimulus u_o is being stimulated with step increase in input $u_o \to u_1$.

Based on equation of dynamics for P from (32) and assumption that before the signal reaches

maximum $Pmax$, degradation of active receptors is insignificant, compared to their dephosphorylation by phosphatases, thus preserving total number of receptors before the signal reaches the maximum. Then we can assume that $R \sim R_o$ and rewrite P_{max} as follows:

$$P_{max} = k_a \frac{k_2}{k_p(k_{d2} + k_{a)}} \frac{k_1}{k_{d1}} u_1 R^2$$

Substituting expression (33) of R_o at steady state we obtain:

$$P_{max} = k_a \frac{k_2}{k_{d2} + k_a} \frac{k_1}{k_{d1}} u_1 R^2 \sim k_a \frac{k_2}{k_{d2} + k_a} \frac{k_1}{k_{d1}} u_1 R_o^{~2} = \frac{u_1}{u_o} \frac{k_{prod}}{k_p}$$

Hence, we showed here that the maximum response of the phosphorylated signal up to a constant is proportional to the fold increase in the step input $\frac{u_1}{u_o}$, where $u_o \gg \frac{k_{d2}+k_a}{k_a k_2} \frac{k_{d1}}{k_1} \frac{k_{do}^2}{k_{prod}}$.

We presented two analytical models of ErbB signaling able of exhibiting Weber's Law responses under certain conditions. The model presented in this Section 4.7.2 assumes simplifications in receptors dimerization and activation steps and suggests that tEGFR memory can be stored as $\sim\frac{1}{u}$ in the region of EGF doses with Weber's Law. The model presented in this section implements receptors dimerization step and suggests that the system responses can exhibit Weber's Law in certain region of EGF doses in which the memory of the EGF dose, u, will be stored in the number of EGFR receptors remaining in a functional form $\sim\sqrt{\frac{1}{u}}$. Our experimental data suggests that functional dependence of tEGFR at steady state on the EGF dose can be a mixture of the two possible proposed relationships (Figure 4.69).

Section 4.7.7: Extrapolation of Weber's Law on pEGFR to pERK and pAKT

Our detailed kinetic model (Section 4.6.1) shows that at certain parameter assignments and range of initial EGF doses pERK can exhibit Weber's Law. We extend our simple analytical model to include pERK activation. We implement the signaling from pEGFR to pERK through one intermediary kinase which is activated by pEGFR and performs two phosphorylation steps needed to activate pERK. The extended analytical model can reproduce qualitative features of signals on both pEGFR and pERK reasonably well. Probabilistic simulations of our extended analytical model with imposed

constraint of Weber's Law on pERK can find regions of parameter space in which pERK shows Weber's Law. Moreover, the simulations suggest that the Weber's Law on pERK responses depends on Weber's Law at pEGFR.

In the extended analytical model pERK is separated from pEGFR only by one intermediate kinase and reactions of phosphorylation and dephosphorylation. The intermediate component thus has no molecular mechanisms to implement memory of EGF doses on its own, and yet, pERK can exhibit Weber's Law even in such simplified network topology. Therefore, it is most likely that the Weber's Law on pERK is a result of linear amplification/transmission of the Weber's Law present on pEGFR. The simplified analytical model thus demonstrates one parsimonious topology to realize Weber's Law on pERK and suggests that Weber's Law at the receptors level can be propagated throughout the signaling kinase cascades.

Indeed, dose responses of pEGFR and pERK from our experimental data suggest that pERK acts as linear amplifier of signal on pEGFR: pERK dose response is shifted and somewhat scaled compared to pEGFR dose response (Figure 4.4). It is possible that if in the course of the signaling no significant changes occur in the intermediate signaling component (such as irreversible modification or degradation), the dose response relationship between pEGFR and pERK activation could remain more or less preserved over time. Hence, if pEGFR responses to step inputs with different initial levels of EGF but same fold-changes have similar amplitudes, it will result in responses with similar amplitudes on the pERK as well (except for the regions of pERK signal saturation).

Recent studies indicate that the signaling components downstream of Sos do not retain memory of previous stimulations (Toettcher et al., 2013). Without molecular mechanism to remember past doses of stimulus, the downstream kinase cascade would not be able to compare doses of incoming stimuli to doses in the past stimulations, and thus, would not be able to realize responses with Weber's Law. This observation further supports the hypothesis that in the EGFR signaling system, Weber's Law to EGF stimulation is most likely implemented at the pEGFR level, the receptors level, and propagates linearly to the downstream targets of pEGFR.

Chapter Five: Discussion and Future Directions

In this chapter we present discussion on the results of this work (see Chapter Four) as well as the future directions of this research. Previously, Weber's Law in mammalian signaling cascades has been reported at the level of transcriptional networks (Goentoro & Kirschner, 2009; Lee et al., 2014). To the best of our knowledge, Weber's Law has not been explored in mammalian phosphorylation signaling cascades. One of the key contributions of our work is the discovery of the EGFR receptors-based mechanism enabling Weber's Law in phosphorylation responses sensing relative changes in levels of extracellular EGF. Most importantly, we suggest that the discovered Weber's Law mechanism can be ubiquitous to other receptor sensory systems. Overall, the proposed receptors-based mechanism of Weber's Law in signaling pathways may be a result of convergent evolution in mechanisms of cell signaling. Regarding the topology of signaling networks exhibiting Weber's Law, we believe that Weber's Law in mammalian signaling networks is parameter, not topology dependent. Therefore, signaling network reduction into a smallest unique network motif responsible for Weber's Law in a given system may not necessarily be possible for mammalian signaling cascades.

Section 5.1: Theoretical insight on properties of network topologies capable of Weber's Law in mammalian signaling networks

Recent works on biological networks were informed by physics outlook and set out to find the simplest topologies capable of various specific system properties, such as Weber's Law or perfect adaptation. For example, such approaches have helped to elucidate mechanism of perfect adaptation in bacteria chemotaxis (Alon et al., 1999) and explore Weber's Law in transcriptional networks in mammalian cells (Goentoro & Kirschner, 2009; Goentoro et al., 2009; Lee et al., 2014). In general, in the course of evolution, bacterial genomes have experienced selection pressure to optimize cellular functions, while also to have smaller genomes. Thus, bacterial genomes may have simpler molecular networks which could have been advantageous for their species survival in light of large population sizes (Sella & Hirsh, 2005). However, for mammalian organisms, evolutionary pressure to reduce

topologies of signaling networks could have been not that pronounced, given there were no significant selection pressure on the genome size, possibly due to genome recombination and much smaller population sizes in mammals. Hence, we believe that mammalian phosphorylation signaling cascades might not necessarily have simplest unique underlying network motifs which would enable them to realize Weber's Law. In fact, we think that driven by evolution of signaling components which might thrive to optimize the function, but not necessarily care for the simplicity of the resulting network, their architecture may have more complicated network modules responsible for Weber's Law.

This work raises an important question concerning the topologies of mammalian phosphorylation signaling networks exhibiting Weber's Law. More specifically, our work suggests that mammalian kinetic signaling networks may exhibit Weber's Law in a parameter-dependent manner.

For example, the simplified analytical model presented in Section 4.7 is just one possible incarnation of a simplification of a more detailed kinetic ODE model we derived (Section 3.1.2). This simplified analytical model relies on several specific assumptions which effectively impose constraints on model's parameter space. In principle, we could have derived several alternative simplified models, each with its own set of simplifying assumptions constraining the model to a specific region in the parameter space of the larger model. Hence, the simplification of a more detailed mechanistic model might produce several smaller topologies with Weber's Law. Each such simplified network topology could have its own mechanism of Weber's Law. It might be possible to distinguish between the true and the incorrect schematics only through experimental work. Thus, Weber's Law in the kinetic model we derived (see Section 4.7) is contingent on parameter assignments, and holds only at a certain range of ligand doses. Therefore, the model topology alone does not guarantee Weber's Law, implying that there can exist parameter assignments at which the model would not exhibit Weber's Law at any range of EGF doses.

In the future, properties of mammalian signaling networks exhibiting Weber's Law could be explored further. It will be interesting to see a theoretical methodology to check the possibility of reducing a given complex signaling network with approximate Weber's Law property into a simplest unique network topology exhibiting exact Weber's Law.

In addition, it is conceivable that the same complex signaling network could realize Weber's Law through different molecular mechanisms, specific to the parameter space and regions of ligand doses. For example, in addition to the EGFR receptors-based mechanism of Weber's Law, pERK responses might realize its own mechanism of Weber's Law through a negative feedback loop, for instance, through activating its own negative regulators (such as DUSP phosphatases). The various mechanisms of Weber's Law might be apparent at the same or different ranges of extracellular ligand, therefore providing cell signaling pathways with a more elaborate control on their information processing.

Lastly, different types of mechanisms of Weber's Law can exist at different levels in signaling cascades and even complement each other. For example, the receptors-mediated mechanism of Weber's Law can realize pERK responses dependent on relative changes in levels of extracellular stimuli, whereas a transcription-network based mechanism (such as described in Goentoro et al., 2009) can detect relative changes in pERK levels and generate transcription activation responses based on the relative changes in pERK.

Section 5.1.1: Previous theoretical studies on topologies of biological networks with Weber's Law

Recent theoretical studies explored properties of topologies of biological networks capable of fold-change detection (FCD) (Shoval et al., 2010). By definition, a system exhibits fold-change detection, if dynamic responses of the system to step inputs with the same fold change, yet different absolute levels of the stimulus, are identical. The FCD property entails both, perfect adaptation and Weber's Law on the responses, and, in addition, requires that their timing and dynamics be the same. Overall, perfect adaptation constraints the long-term steady state system behavior; whereas, Weber's Law imposes limitations on the maximal amplitudes of the responses. FCD concerns systems responses on both the long-term and the immediate time scales, and in fact, applies to the entire time-course of the responses. Hence, fold-change detection imposes very stringent constraints on the underlying network topology.

It is important to note that the authors in Shoval et al., 2010 used a limited interpretation of Weber's Law, evaluating Weber's Law only at the amplitudes of the immediate responses. However,

the most general definition of Weber's Law does not impose constraints on the timing of the maximal amplitudes of the responses.

The authors derive theoretical conditions necessary and sufficient for a system to show FCD and present several examples of simplified network topologies capable of achieving FCD. In addition, they present variable-substitution methods for transforming one type of topology into another, and so, to be able to better understand relationship between different classes of topologies.

The authors describe three simplest network topologies capable of FCD: incoherent feed forward loop, nonlinear integral feedback, and logarithmic input topology with linear integral feedback, and a more general class of network topologies with log-linear input systems (Figure 5.1).

In brief, the incoherent feed-forward loop (IFFL) topology entails simultaneous stimulus-induced activation of an output target and of a repressor of that output target. Recent work explored properties of the parameter space of the IFFL topology which would allow fold-change detection (Goentoro et al., 2009). In particular, the authors find that the IFFL is capable of FCD only when the activator is in its linear regime and the suppressor acts in saturation. Several experimental studies suggested that an IFFL at the transcriptional level could allow fold-change detection in mammalian cells (Goentoro & Kirschner, 2009; Lee et al., 2014).

The nonlinear integral feedback topology described in (Shoval et al., 2010) essentially is a modification of a well-known integral feedback topology, in which, a deviation of an output variable from its intended steady state value is integrated over a certain time and is fed back to control this output variable. By non-linearizing the feedback expression and using a ratio-based activating controller one obtains a FCD-capable topology referred to as a nonlinear integral feedback. Interestingly, with a variable transformation in the nonlinear integral feedback topology one can derive a third simplified network topology presented in Shoval et al - a logarithmic input with linear integral feedback topology.
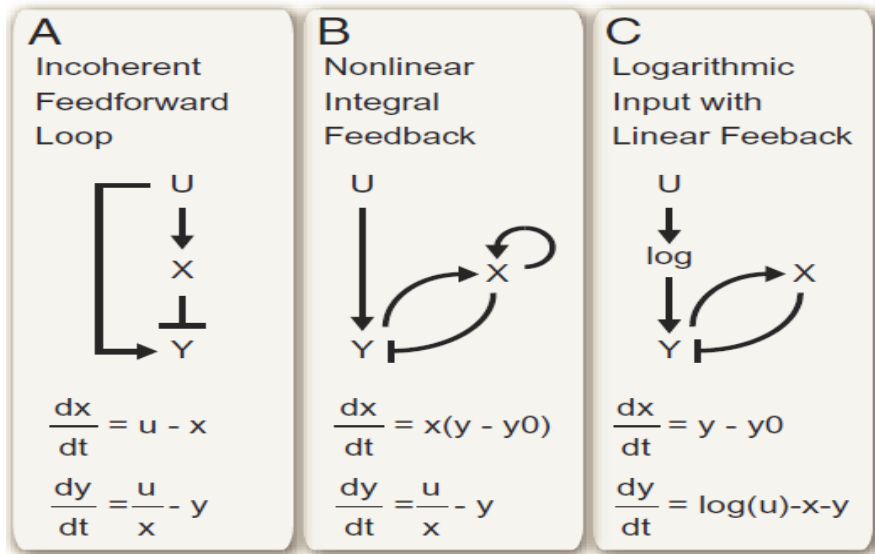
Figure 5.1: Suggested mechanisms of fold-change detection: incoherent feed-forward loop, nonlinear integral feedback, linear integral feedback with log input. Reproduced from (Shoval et al., 2010)

In addition, Shoval et al., 2010 present experimental strategies to validate the presence of fold-change detection in biological systems. They suggest investigating the input-output relationship of the system, checking if the biological system has one of the enabling network topologies; and lastly, testing if the system can perform scale-invariant searches. However, whereas perfect adaptation and Weber's Law are easier to validate individually, verifying fold-change detection may require comparing not just the peaks, but entire dynamical profiles of system responses, which may be experimentally unfeasible in certain systems. This leads to a broader, philosophical question whether the biological systems may even be required to exhibit exact fold-change detection at all. Often operating in environments with molecular noise, biological systems perhaps may have developed mechanisms which rely only on certain features of responses, rather than on the entire shape and the amplitude of the signaling response.

Notably, most of the simplest network topologies capable of FCD presented in the study by Shoval et al., 2010 concern transcriptional regulation networks, which operate via different mechanisms and on a different time scale then the phosphorylation signaling networks. Overall, Shoval et al., 2010 main goal was to find the simplest network topologies with FCD. This approach may be possible for simple organisms such as bacteria, which under evolutionary constraints to have

smaller genomes, could have evolved to have smaller signaling networks. However, this assumption may not be relevant in the case of complex mammalian signaling networks.

A recent study addressed a possibility of Weber's Law specifically in enzymatic signaling networks (Skataric & Sontag, 2012). The theoretical work explored all possible three-node enzymatic topologies with adaptation which also exhibit Weber's Law. Necessary and sufficient conditions for a topology to be capable of Weber's Law are derived (Skataric & Sontag, 2012). In addition, the authors make a point that most enzymatic networks would possess only approximate Weber's Law, and that under a certain simplifying transformation, larger kinetic networks with approximate Weber's Law can be transformed into smaller topologies with exact Weber's Law. However, we find several shortcomings of the study. First, the study relies on a not quiete valid *in vivo* Michaelis-Menten formalism to encode enzymatic reactions. Next, to assess Weber's Law responses the authors consider very similar inputs (i.e. within one order of magnitude) and only low fold increases (within 20%), thus leaving a possibility that the observed approximate Weber's Law is a result of comparing systems responses on very similar inputs and trivial step perturbations. Most importantly, the transformation applied to obtain simplified network topologies is very constraining to model's parameter space. In particular, the authors impose that the model's output variable reaches steady state much faster than the other internal variable. Such constrain would hold only in a limited parameter subspace. Hence, the described above studies of network topologies capable of FCD (Shoval et al., 2010; Skataric & Sontag, 2012) demonstrate that even the simplest enzymatic networks with Weber's Law are not, in fact topology-, but parameter-dependent.

Section 5.2: Receptor-mediated mechanism of Weber's Law in sensing EGF levels at pERK responses may be advantageous to cell information processing

Dynamics and amplitude of phospho-ERK signal are known to play an important role in cell fate decisions, such as cell growth, division, and proliferation (see Section 2.8.4). Ligand stimulation of cells might produce different signaling outcomes at phospho-ERK level, depending on the cell type and context of gene expression in the cells (Ebisuya et al., 2005). Currently, the relationship between pERK signalling dynamics and resulting cell phenotypes has not been clearly understood and is

actively investigated (Albeck et al., 2013). Many studies focused on understanding the relationship of steady-state pERK levels to growth factor stimulation and cell responses (Goldbeterl & Koshland, 1984; Blüthgen, 2006). However, it is likely that both, the steady state as well as the dynamics of pERK responses affect cell fate decisions (Sasagawa et al., 2005, Selimkhanov, Taylor, Yao, & Pilko, 2014).

Hence, given that pERK dynamics matters to cell fate decisions, it is biologically plausible that to produce a coordinated cell response to ligand stimulation neighbouring cells in a tissue may need to generate similar responses at their pERK levels, thus justifying the need to exhibit Weber's Law at pERK.

Our experimental data as well as our model suggest that pERK can exhibit Weber's Law in sensing EGF doses. We find that Weber's Law at pERK responses is most likely realized through Weber's Law at pEGFR responses, and that possibly pERK does not possess its own mechanism of Weber's Law. pERK is known to be activated by a series of kinase cascades via two-step phosphorylation. Another downstream target of pEGFR important to cell fate, pAKT, is known to be activated via the PIP3 activation. Unlike ERK, AKT requires only one phosphorylation to be fully activated. It will be interesting to explore if despite such distinctions in activation, pAKT is also capable of Weber's Law. Such observation could then allow us to hypothesize that other targets of pEGFR, not explored in this work might also exhibit Weber's Law, given that the pathway leading to the target is linear and there are no irreversible inhibitory loops or degradation of intermediary components. In case some intermediary components are degraded in the course of ligand stimulation, the downstream signaling cascade may become insensitive to the repeated stimuli, unless the pools of the functioning intermediate component are restored.

Having a single unified control over the information transmitted down into the cell interior can allow cells to better regulate the interaction between the cell environment and the cell signaling responses. In addition, such unified control may help to achieve coherence in activation and regulation of downstream components between themselves: in the case when each intracellular signaling component implemented its own regulation mechanism, different components could respond differently to the same change in extracellular stimuli levels. It may be advantageous to cells

first to closely regulate signalling at the upstream receptor level, as the effects of such regulation would then be propagated to the downstream targets. Therefore, relying on intricate regulation of the number of functional cell surface receptors could provide cells with a more effective way to adjust and dynamically maintain their sensitivity to extracellular stimuli. Given that the number of functional cell surface receptors can be regulated through multiple mechanisms, such as, for example, receptors endocytosis, or inactivating post-translational modifications, receptors synthesis and others, the cells would thus have versatility in its mechanisms to control and fine-tune the number of cell surface receptors and to ensure cells sensitivity to consequent stimuli.

The MAPK signaling cascade, which leads to pERK activation, is one of the most ubiquitous and well-studied signaling cascades in mammals (Chen et al., 2009). MAPK signaling can be activated through a variety of cell surface receptors, such as, for example, the ErbB receptors, and the G-protein coupled receptors (Goldsmith & Dhanasekaran, 2007). Therefore, it is plausible that the mechanism of Weber's Law in pERK responses presented here could be generalizable to signaling through many other receptors sensory systems. Relying on Weber's Law in receptors responses to relative changes in levels of extracellular stimuli and on the linear propagation of the receptors activation to their downstream signaling targets, such schematics represents a versatile tactic of the cells to be able to respond to relative changes in levels of extracellular stimuli.

Section 5.3: Comparing parameter distributions between normal and cancer cell lines

We obtained experimental measurements of EGF signaling in three different mammalian cell lines: MCF-10A – an immortalized normal mammary epithelial tissue, MCF-7 and SKBR3 two breast-cancer derived tissues. Dose-responses of pERK measured over 1.5 hours after EGF addition show that in the three cell lines pERK decay has different dynamics: after reaching the peak response around 10 minutes, in MCF-7 pERK decays rapidly, in MCF-10A the decay is slower, whereas in SKBR3 after transient peak the signal remains unchanged between 30 minutes and 1.5 hours (Figure 5.3). We then also consider dose-responses of other targets in these cell lines (Figure 5.7). The data for all three cell lines was collected on the same day. The fluorescence measurements most likely have different fluorescence background offset (see Section 4.1.3) in these three cell lines due to

differences in the off-target binding specific to each cell line. However, we observed that the same primary antibodies are used to measure each of the three targets and the "no primary" control measurements were similar between the three cell lines. Therefore, we can compare the fluorescence data obtained in three different cell lines.

In agreement with our prior estimates of total EGFR abundances, the dose responses of tEGFR show that MCF-10A contains a higher number of receptors, with SKBR3 containing less receptors, and MCF-7 having a low number of receptors. The corresponding experimental estimates suggested that MCF-10A had approximately $3x10^5$-$6x10^5$, SKBR3 ~ 2.5 $x10^5$ and MCF-7 $5x10^4$ EGFR receptors per cell. The time-courses of total EGFR abundances in response to treatment with different doses of EGF over up to three hours show that in SKBR3 and MCF-7 total number of EGFR does not change significantly, even at higher concentrations of EGF doses. However, MCF-10A total EGFR measurements show that at higher EGF doses significant amount of receptors is degraded. This observation is in agreement with the pEGFR phosphorylation measurements – higher activation of EGFR leads to more of the receptors being degraded.

We observe that MCF-7 cells in response to EGF treatment did not show significant pEGFR activation; whereas in SKBR3 pEGFR signal increased with higher EGF doses and decayed over time. In MCF-10A pEGFR at higher doses was more pulse-like with a faster rate of decay compared to the signal decay rate in SKBR3 cells. The data also suggests that across all three cell lines pEGFR signal did not exhibit saturation, a feature which is manifested by flatness of dose-responses at higher EGF doses, as in example of pERK and pAKT signal saturation in dose responses shown in Figure 5.3.

We fit the model separately to each of the cell line-specific experimental data set. We impose the same priors on the parameters in model simulations for each of the three cell lines. With distinct parameter assignments, the model can fit reasonably well the experimental data for each of the three cell lines (Figure 5.4, Figure 5.5, and Figure 5.6). We then conduct a preliminary MCMC sampling of parameter space for data of each of the three cell lines. Although the MCMC chains in the preliminary simulations have not converge and need longer simulation time, we could draw some preliminary conclusions from certain features of model parameter space obtained for each of the three cell lines.

Obtaining the properly converged MCMC chains would then allow to better understand what basic signaling processes are fundamentally different between the cell lines.
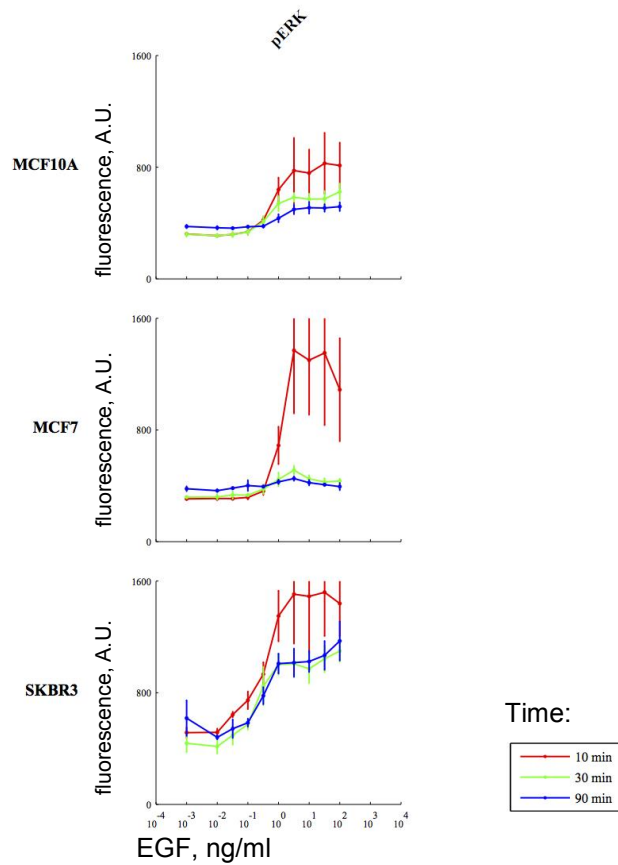
Figure 5.2: Experimentally measured pERK dose-responses over 1.5 hours of EGF treatment in three different cell lines MCF-10A, MCF-7, SKBR3.
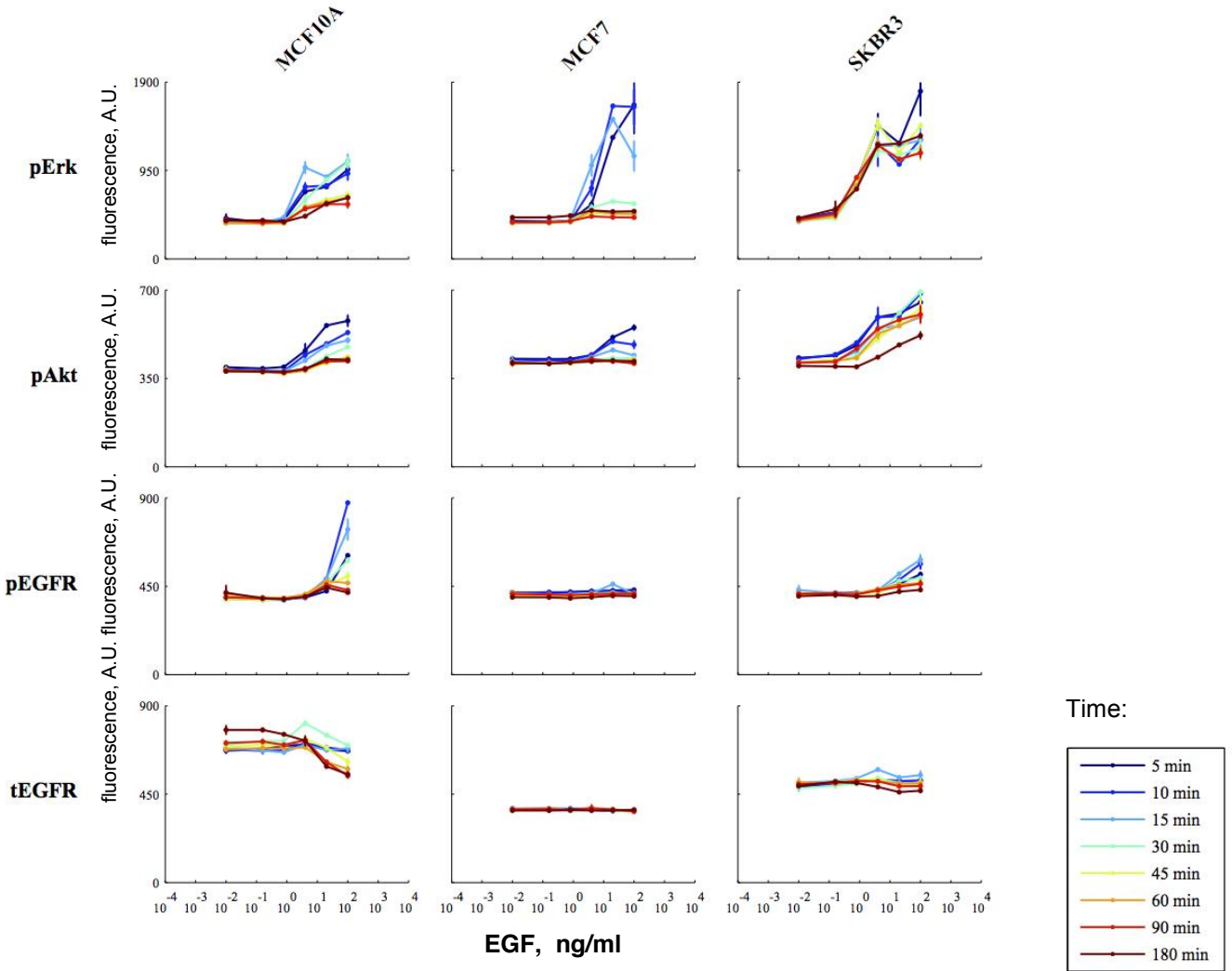
Figure 5.3: Experimentally measured phosphorylation dose-responses over three hours of EGF treatment in three cell lines MCF-10A, MCF-7, SKBR3. Measurements are reported in arbitrary units of fluorescence. The data shown is not normalized to allow for the comparison between the cell lines.

However, even from the preliminary MCMC chains we observe that the model correctly predicts difference in the average EGFR abundance in MCF-10A and MCF-7 cell lines (Figure A-6). Note that we did not impose different priors for each of these cell lines, although, we know that on average MCF-10A cells express about an order of magnitude more EGFR then do MCF-7 cells. Fitting the phospho-EGFR and tEGFR dynamic responses to different EGF doses allowed the model to recover this aspect of cell lines difference and to predict EGFR abundance in MCF-10A cells to be ~2.09 x 10^5 and in MCF-7 ~ 10^4 EGFR receptors per cell. The model failed to correctly estimate the number of EGFR receptors in SKBR3 cells and predicted it to be ~2.4x 10^4, whereas our

214

experimental measurements estimated EGFR abundance in SKBR3 cells to be on the order of 10^5-10^6. It is possible that the SKBR3 signaling dynamics is significantly influenced by the overexpression of ErbB2 receptors, which we do not implement in the current model.

We conduct a preliminarily analysis to find the differences in model parameter space obtained by fitting the model to the cell-line specific experimental data sets. We first look at parameters related to maintaining the basal levels of EGFR in the absence of EGF to try to explain the observed differences in EGFR abundances. We find that in MCF-10A cells the rate of basal EGFR production is higher than in the other two cell lines, in addition, the rate of degradation of basal receptors is slower than their recycling, thus allowing their accumulation on the cell surface. For MCF-7 cells, the rate of production of EGFR is low, and although the rate of receptors internalization is slower than in MCF-10A, internalized inactive receptors would preferentially be degraded, rather than recycled. The fraction of internalized receptors which is degraded is determined by the corresponding rates of recycling to degradation of inactive receptors. Overall, the steady state level of the total number of receptors can be shown to be as follows:

$$T = \frac{k_{prod}}{k_1 k_{do}}(k_{do} + k_i + k_r)$$

where k_{do} is the rate of basal degradation of receptors, k_i – rate of recycling of inactive receptors, k_i – the rate of internalization of inactive receptors, k_{prod} – rate of EGFR synthesis. Therefore the differences in the rates of EGFR synthesis predicted for MCF-10A and MCF-7 cells, and preferential degradation of receptors over their recycling in MCF-7, versus preferential recycling over degradation of receptors in MCF-10A cells, could explain different predicted levels of the initial EGFR in these two cell lines. Here we show how the model can help to explain the differences in the initial abundances of cell surface EGFR based on predicted parameters related to receptors turnover.

We find that *kcat* of activity of EGFR phosphatases in MCF-7 is predicted to be lower than in MCF-10A cells, as signal downregulation is possibly compensated by more rapid internalization and degradation of activated receptors. Therefore, one might hypothesize that signal downregulation in MCF-7 happens mostly due to rapid removal of activated receptors from the cell surface and their degradation. Given that the number of receptors initially is already low (~10^4 receptors per cell), this

can explain the observation that experimental measurements of tEGFR appear to be constant over time and doses. In fact, our results suggest that in MCF-7 cells the number of receptors does decrease over time at stimulation with higher doses of EGF, yet absolute amount of degraded receptors is low to be detected by experiments. It is possible, in the MCF-7 cells given the number of EGFR is low, total number of activated receptors is also low, which could lead to their rapid removal through the internalization machinery which is not-saturated. Saturation of cell internalization machinery by the overabundance of activated receptors has been previously reported in the literature (Sigismund et al., 2008).

We also find that compared to MCF-10A cells, in MCF-7 the rate of ERK phosphatase activity (kd5) is predicted to be lower, whereas the rate of activation of ERK (kd51) was predicted to be as high as in MCF-10A cell lines. A significantly higher rate of activation over deactivation rate, thus could explain the observed high amplification on pERK responses in MCF-7 experimental measurements, which could be a result of an overactivating mutation in this breast cancer derived cell line. Therefore, it is plausible that rapid pERK signal decay observed in MCF-7 is most likely due to rapid attenuation of signal at the EGFR receptors level.

The experimental data collected in SKBR3 cells show slow EGFR receptors degradation and somewhat sustained pEGFR and pERK responses. pEGFR signals in SKBR3 cells had lower amplitudes than in MCF-10A cell. The preliminary model parameter scan predicts that in SKBR3 cells the rate of degradation of active receptors (k60) is slower than the rate of receptors recycling (kd91), possibly explaining slow pEGFR signal downregulation observed in SKBR3 cells. Interestingly, it is known that overexpression of ErbB2 in SKBR3 cells can lead to their dimerization with EGFR, and that the formed dimers are internalization-impaired compared to the EGFR homodimers (Hendriks, 2003), further supporting model prediction regarding the rates of degradation and recycling of active EGFR.
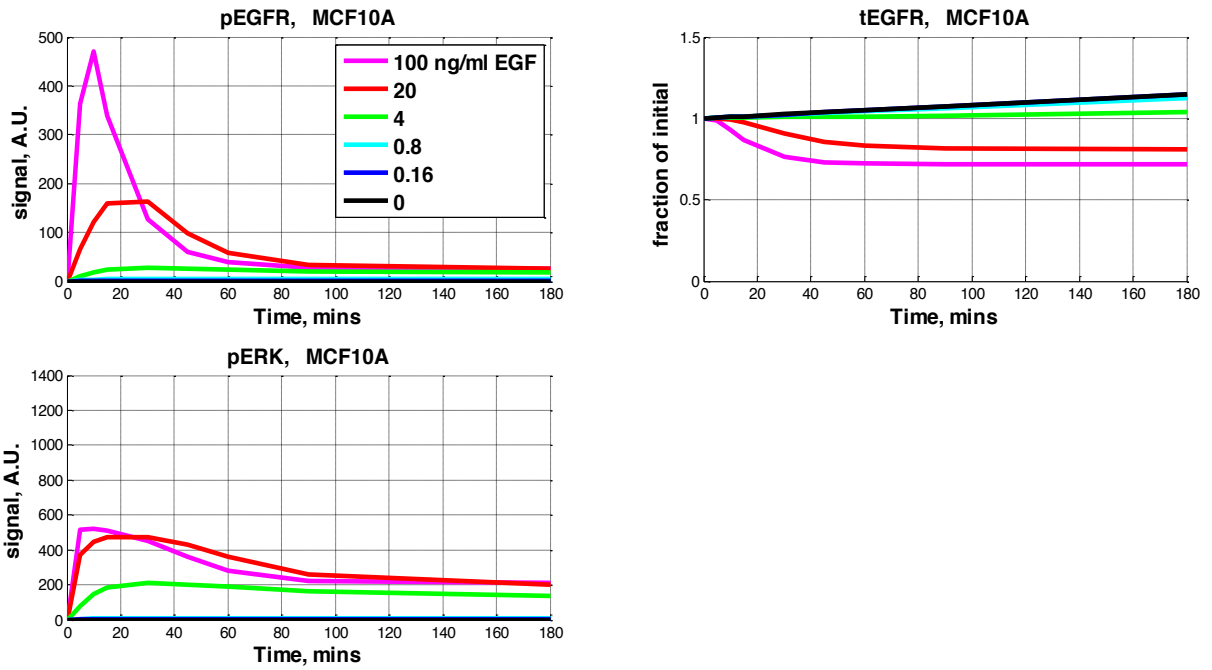
Figure 5.4: Predicted fit to the MCF-10A experimental measurements.
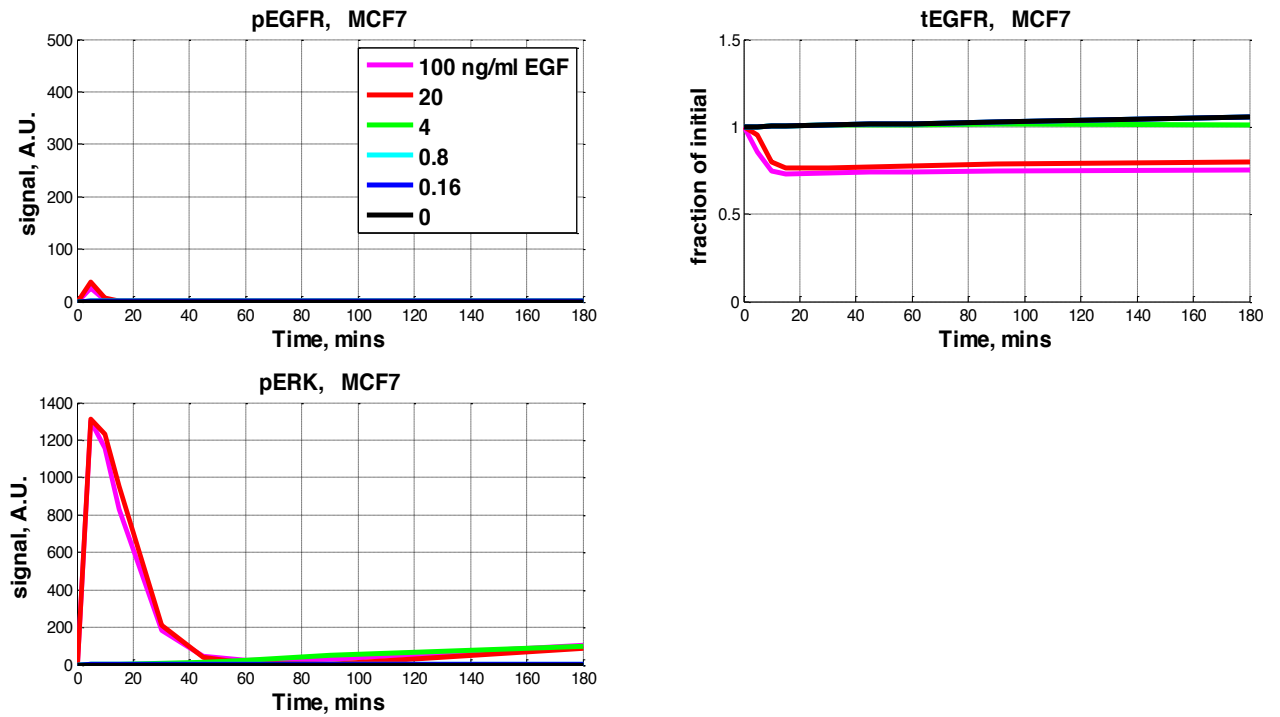
Figure 5.5: Predicted fit to the MCF-7 experimental measurements.
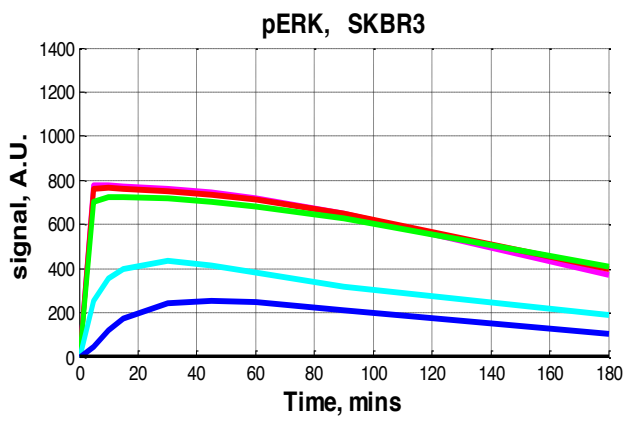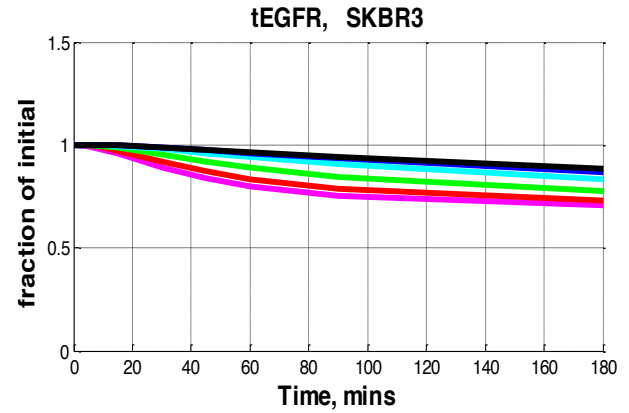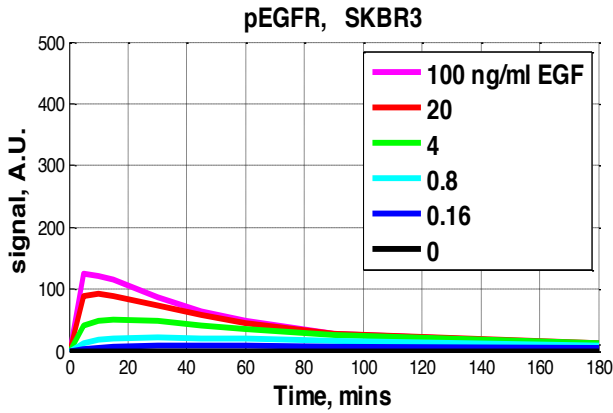
Figure 5.6: Predicted fit to the SKBR3 experimental measurements.

Our model and probabilistic framework can be used to comprehensively sample model parameter space which agrees with the experimental data from each of the three cell lines. The obtained parameter distributions can then be compared to identify what fundamental signaling processes are disrupted in cancer cell lines and what are their possible implications to cell signaling.

We find that in all three cell lines, after transient increase pEGFR signal decays nearly to its pre-stimulus level. It is interesting to investigate how differences in receptors down-regulation dynamics although all produce signal decay at pEGFR levels, could determine different system responses to repeated stimulations with EGF. One next step in this direction could be to investigate experimentally if MCF-7 and SKBR3 cell lines also are able to sense relative, not absolute, changes in EGF levels in downstream kinase cascades leading to ERK activation. Do cancer cells adjust their sensitivity to EGF after they were exposed to chronic stimulation with EGF? What are the quantitative features of EGF memory in cancer cells and what possible impact such memory might have on the

cancer cells' ability to process important EGF signals? These are some questions which could be interesting to investigate with the experimental measurements in cancer cells.

Table 7: Parameter's description

k1/ kd1	EGF binding/unbinding from EGFR monomers
k31/ kd31	EGF binding/unbinding from EGFR dimers
k2/kd2	EGFR dimerization with one EGF molecule bound
k2a/kd2a	EGFR dimerization with two EGF molecule bounds
kd123	kcat of EGFR phosphorylation
kd950	kcat of EGFR dephosphorylation
k60	degradation of activated EGFR
k77	basal production of EGFR
k910	basal internalization of EGFR
kd910	recycling of inactive EGFR
k91	internalization of EGF-bound receptors
kd91	recycling of EGF-bound receptors
k61	degradation of inactive EGF
k50	binding of kinase activating ERK
kd50	unbinding of kinase activating ERK
kd51	kcat of activation of pERK by pEGFR
kd5	kcat of dephosphorylation of pERK

Section 5.4: Fitting the model to single cell distributions of cell responses

Most of the results presented throughout Sections 4.2 - 4.6 were obtained by fitting the model to experimental data representing average cell response to a particular ligand treatment. Each experimental point to which the model is fitted was obtained by taking the average of the averages of single cell distributions of responses across several independent biological replicates of the experiments. To minimize contribution of experimental error and to derive the most probable system responses of the overall ensemble of cells, the biological replicates of the experiment were repeated on different days. The error used to fit such experimental data is standard error in the mean. Such error reflects uncertainty in detecting the average cell population behavior between different replicates. The error is most likely to be attributed to the experimental error of measuring day-to-day variability in average cell population responses and is assumed to be normally distributed. To summarize, fitting the model to averages of the distributions of single cell responses allows to derive

knowledge about most likely behavior of cell population and is a good approximation if single cell distributions appear to have Gaussian shape.

The variability in the predicted distributions of responses in model fits to the ensemble averages data captures the variation arising due to day-to-day variability in measurement error in the experimental data of average cell behavior. In addition, the variability in produced fits arises due to unidentifiability in model parameters space – multiple parameter assignments can fit the data similarly well with some deviation in certain predicted points.

It can be interesting to understand how responses of individual cells can vary between within cell population and to gain an insight about the cell-to-cell variability in the underlying parameter values. To the best of our knowledge, fitting mechanistic models to dynamic data of single cell distributions has not been previously done, most likely due to the experimental challenge it presents, but also, due to the absence of a principled computational and probabilistic framework that the interpretation of such data and the results require.

Most of the single cell distributions we obtain can be approximated with the Gaussian distribution. However, in some cases the distributions of single cell responses might deviate from the normal distributions in their shape (Figure 5.7), and hence, other than the normal distributions might be better for approximating single cell responses. The Gamma distribution has been proposed to approximate skewed single cell distributions as in Figure 5.7B(Birtwistle et al., 2012).

Our experimental data shows that distributions of single cell responses of the same target can vary in shape depending on EGF dose of stimulation and time when response is measured. For example, the variance of pERK responses to EGF tends to be smaller at earlier times (Figure 5.12) compared to later times. Interestingly, for pEGFR response (Figure 5.14) variances in single cell distributions increase with time of measurement of the response and are the highest at the peak of the cell responses, decreasing to the variances in the pre-stimulus single cell responses distributions. Shapes of single cell distributions at a given time point might depend on EGF dose of the treatment (Figure 5.9). For example, we observe that at lower EGF early pERK responses might be double-peaked in distributions, suggesting heterogeneity in the rates of ErbB signaling pathway activation between different cells in the same well. Therefore, modeling signal response dynamics at levels of

single cell distributions may provide a valuable insight into cell population behaviors which can be important in the context of cells functioning in tissues in an organism.

In practice, there are certain possible limitations to using the distributions of single cell response data. For example, at certain conditions the distributions of single cell responses may be spread in shape due to the fact that the single cell response distributions might have large variances compared to the usually lower errors in the mean used in the ensemble averages data. Therefore, the parameter distributions predicted by fitting the model to single cell responses distributions may be not as constrained as they may be in case the model is fitted to the cell population ensemble averages data. Still, such distributions could be informative about the degree of cell-to-cell variability in values of various parameters, such as, for example various protein abundances.

Overall, the predicted distributions of single cell responses obtained from fitting the model directly to single cell responses distributions allows to capture the true cell-to-cell difference in their gene expression, as well as differences in their other parameters (such as, for example, effective *in vivo* kinetic rate constants of reactions), and an experimental error due to detection error of the phosphorylation of the single cell data.

In this work we obtained experimental measurements of single cell responses to EGF treatments in cancer cell lines and developed a principled method to normalize and scale independent biological replicates (see Section 4.1). The data can then be used to fit the kinetic ODE model of ErbB signaling we implemented using probabilistic framework (Section 3.2) to derive posterior parameters distributions that are most consistent with the observed experimental data. It can be interesting to explore how various subpopulations of cells with different signaling dynamics map onto the subregions of model parameter space. In addition, the predicted distributions can then be used to generate distributions of single cell responses to various perturbations, such as, for example gene deletion, overactivating mutation, drug perturbations, and others.
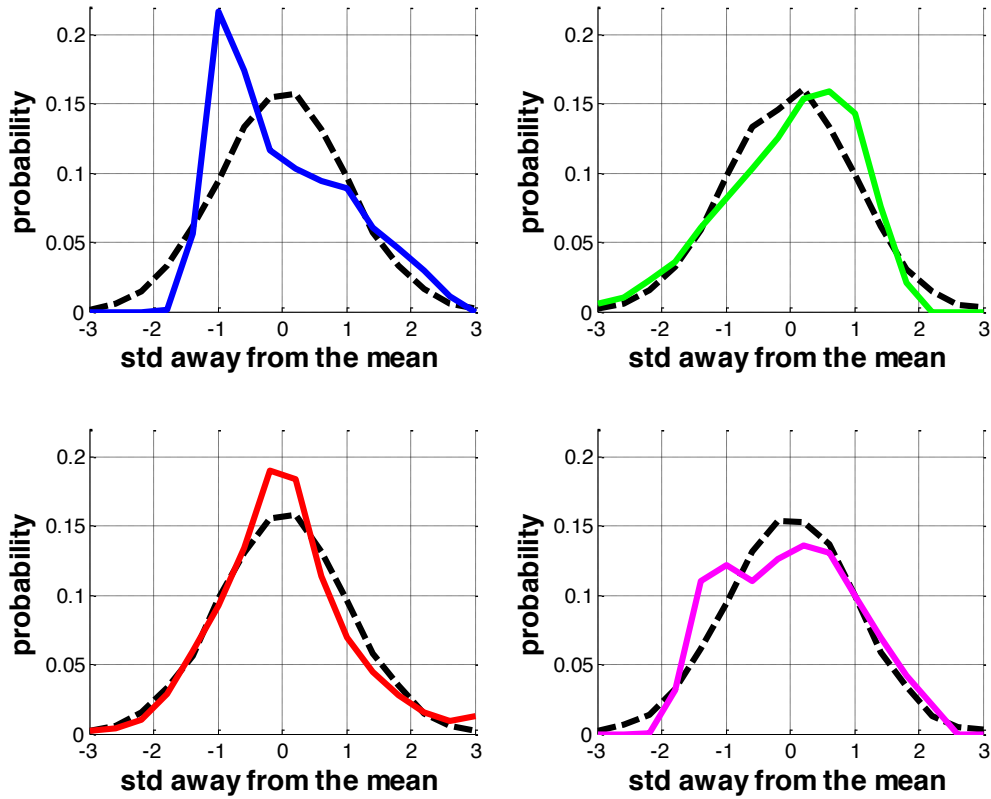
Figure 5.7: Examples of experimentally measured single cell distributions of the phosphorylation responses. For comparison, shown in black are normal Gaussian distributions. Each distribution is plotted normalized to its own mean and standard deviation. Data is derived from MCF-10A cells treated with EGF.
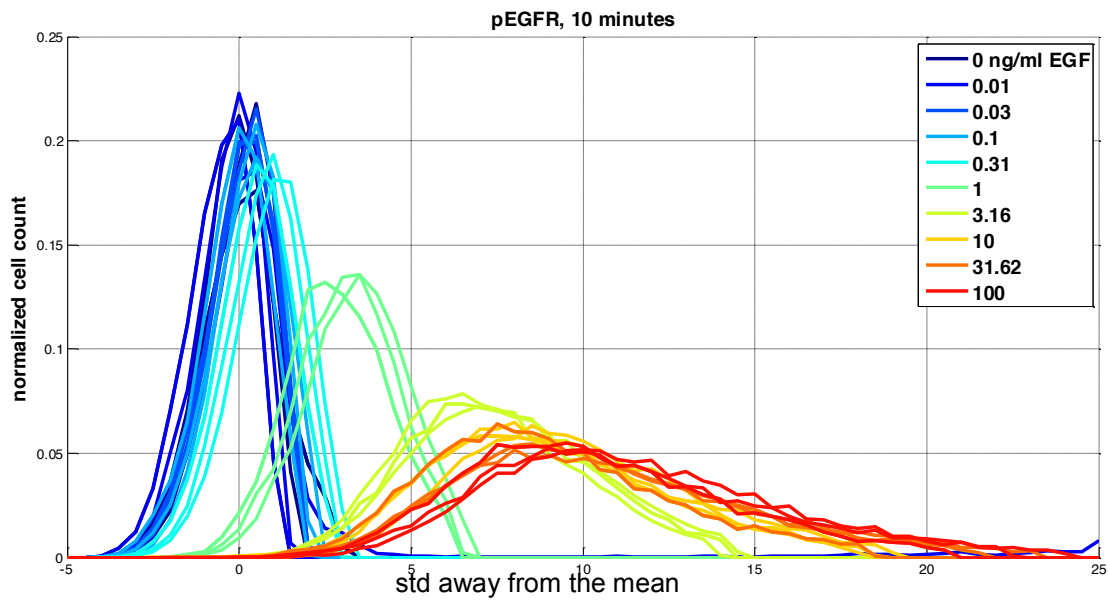
Figure 5.8: Single cell distributions of the pEGFR dose-responses 10 minutes after treatment with EGF of MCF-10A cells. The distributions are standardized to the corresponding 0 EGF control distribution.
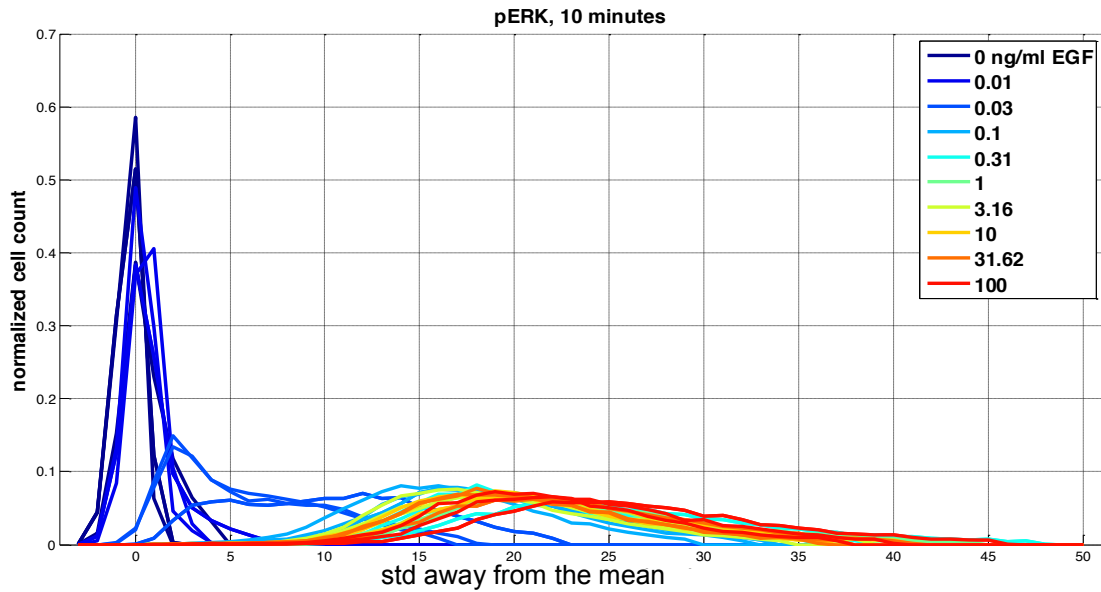
Figure 5.9: Single cell distributions of pERK dose-responses 10 minutes after treatment with EGF of MCF-10A. The distributions are standardized to the corresponding 0 EGF control distribution.
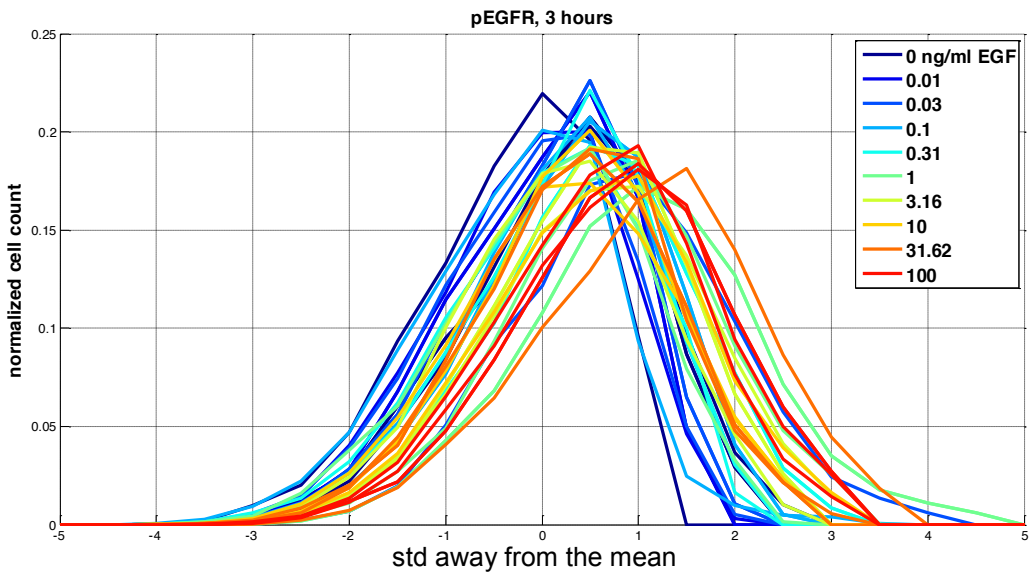
Figure 5.10: Single cell distributions of pEGFR dose-responses three hours after EGF addition in MCF-10A cells. The distributions are standardized to the corresponding 0 EGF control distributions.
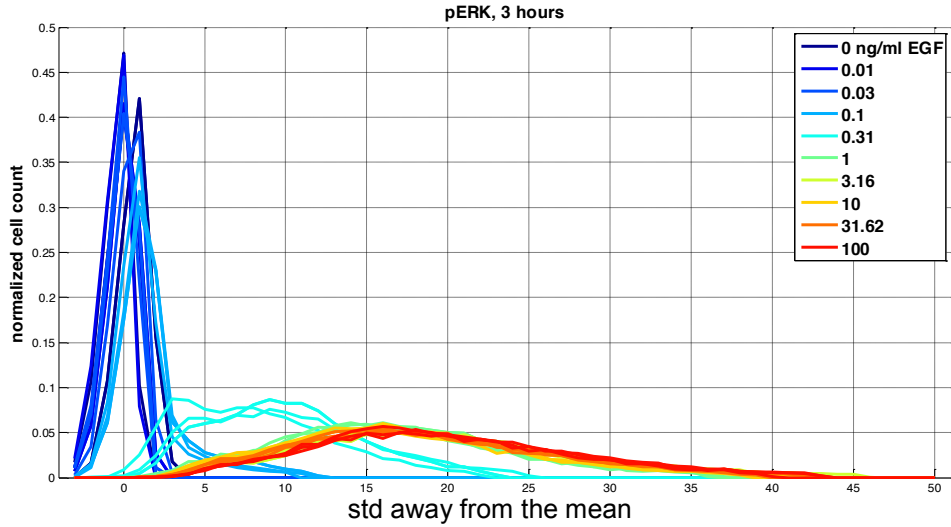
Figure 5.11: Single cell distributions of pERK dose-responses after three hours of EGF treatment. The distributions are standardized to the corresponding 0 EGF control distribution.
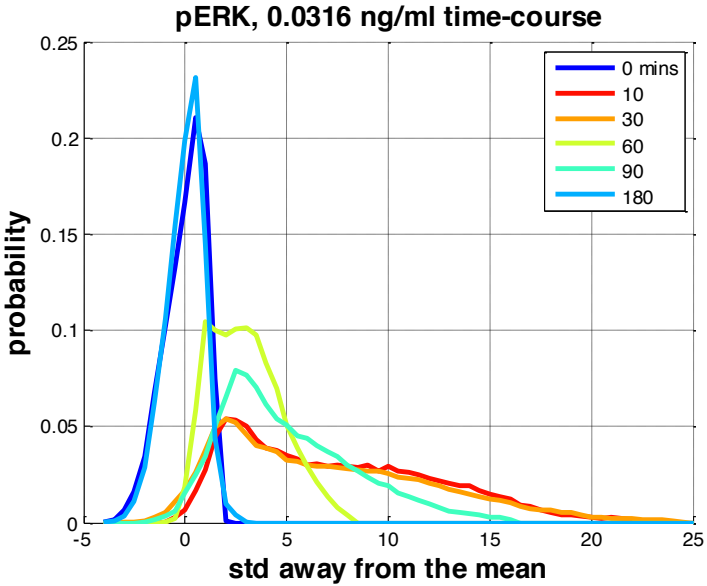
Figure 5.12: Time-course of single cell distributions of pERK responses to 0.0316 ng/ml EGF treatment in MCF-10A cells.
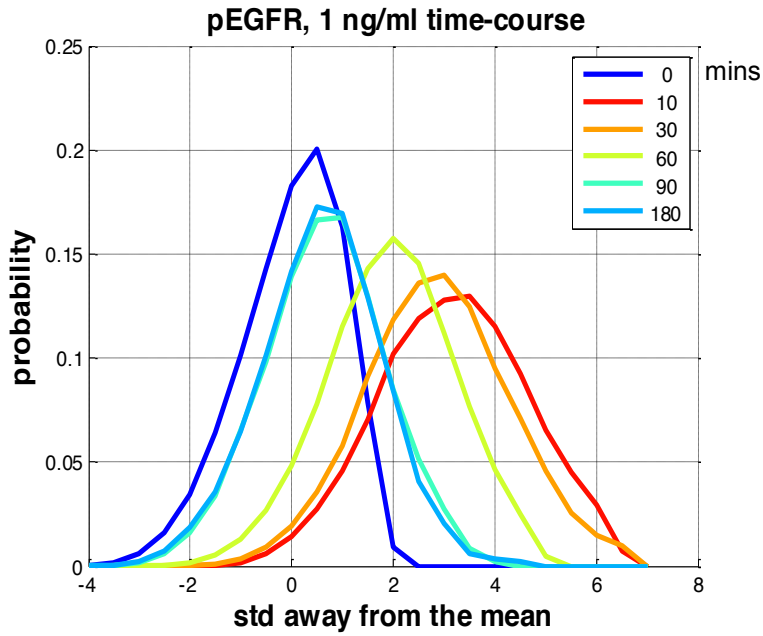
Figure 5.13: Time-course of single cell distributions of pEGFR responses to 1 ng/ml EGF in MCF-10A cells.
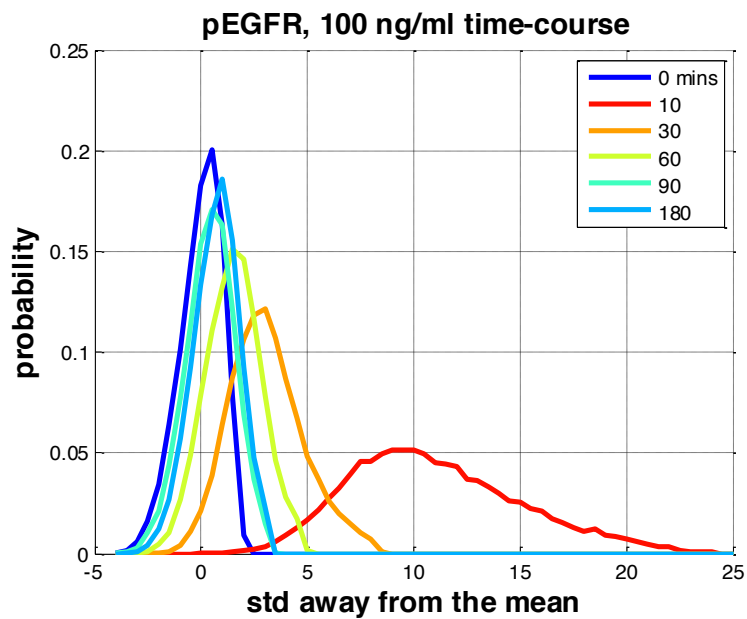
Figure 5.14: Time-course of single cell distributions of pEGFR responses to 100 ng/ml of EGF in MCF-10A cells.

Section 5.5: Exploring effects of Weber's Law in sensing relative EGF signals on cell phenotype

How can the ability to decode relative changes in EGF by the phosphorylated targets affect cell behavior? The hypothesis that Weber's Law in the phosphorylation responses is important to cell functioning could be validated through observing and quantifying various cellular phenotypes directly controlled by the considered phosphorylation targets. The phenotypic features can be directly regulated (or strongly influenced) by the downstream targets of pEGFR signaling through regulating gene transcription.

For example, some studies of Weber's Law in other biological systems have reported a correlation between fold-change responses of the target of interest, namely b-catenin levels, and cell phenotypes related to development (Goentoro & Kirschner, 2009). Similar approach could be applied to explore the relationship between the Weber's Law in phospho-ERK responses and cellular behavior.

For instance, cells relying on EGF-induced MAPK signaling for proliferation could be subjected to various fold-changes in EGF, starting with the pretreatments with EGF doses from the reported here range likely to exhibit Weber's Law (0.03-0.3ng/ml EGF). The correlation between cell proliferation rate and the response to the applied fold-change in EGF dose could then be assessed. If it is found that the cell proliferation rate does not depend on the initial dose in the EGF pretreatment, but instead, on the fold-change of the EGF applied at the step increase, this could suggest that indeed, fold-changes, and not absolute levels, in EGF are important for cell proliferation. To establish that the observed effect is mediated through pERK, the pERK responses would also need to be quantified in these experiments. However, cell proliferation is a process shaped by multiple regulators, and thus, the effect of pERK may not be uniquely determining cell decisions.

A possible method to assess physiological consequences of pEGFR signaling in cells treated with EGF could be to quantify protrusions of lamellipodia, feature of cell-migration initiation. It has been shown that lamellipodia protrusion occur rapidly in MCF-10A cells in response to EGF treatment and are mediated through phosphorylation of EGFR, ERK and Akt (Kleiman, 2010). Since lamelipodia protrusions are experimentally quantifiable, and there is not significant time delay between EGF

treatment and the elicited cellular response (lamellipodia protrusion) this cellular phenotype may be convenient to explore the effects of Weber's Law in phosphorylated responses to step-inputs in EGF on cell behaviors.

Section 5.6: Comparing apples and oranges: cell memory of multiple ligands and multiple receptors

Section 5.6.1: How EGFR can store the memory of combinations of different ligands

EGFR can bind multiple types of ligands, such as, EGF and TGF-alpha, heregulin (HER or Neu) and others (Zahnow, 2006). It can be interesting to investigate how EGFR receptors can remember treatments of combinations of different types of ligands, especially, given that in *in vivo* cells are usually exposed to multiple different stimuli over time. EGF and TGF-alpha have been suggested to co-occur in human mammary epithelium and to play an important role in development (Fisher & Lakshmanan, 1990).

How can cells remember and discriminate between different types of ligands? For example, it has been suggested that yeast is able to compare relative ratios of glucose, its preferred nutrient, to galactose, and the ratio between the levels of two nutrients then determines cell state (Savir et al., 2005). Similarly, since EGFR can bind multiple ligands which can co-occur together *in vivo*, it can be interesting to study what strategies receptors use to remember the combinations of the ligands and what affects these strategies have on cell responses to repeated stimulations with these ligands.

A recent theoretical work modeled system responses to simultaneous fold-change increases in two different ligands in a system capable of fold-change detection in levels of each of the two ligands (Hart et al., 2013). The study suggested that under certain circumstances, the generated response would be proportional to the product of the two fold changes, implying that the system would compare change in each ligand to its own background level and then respond proportionally to the product of the two folds. However, the presented theoretical study analyzed ligand-receptor interactions via Monod-Wyman-Changeux model (Changeux, 2012), a model which may not be applicable to describe other ligand-receptor interactions.

Contrary to the suggestions of Hart et al., 2013, that the combinations of ligand may be processed by comparing each corresponding ligand dose to its own background, and the resulting response would be proportional to the product of the two fold-changes, it is possible that for EGFR measuring fold-changes in multiple simultaneous signals is more complicated. Although the modes of EGFR activation by various ligands may be similar (through induction of receptors dimerization and phosphorylation), the kinetic properties of ligand-receptor interactions may differ and thus may produce different physiological outcomes of signaling.

For example, both EGF and TGF-alpha can bind and activate EGFR. However, unlike EGF, TGF-alpha is known to dissociate from the internalized ligand-receptor complex in the endosomes which have lower pH levels compared to the pH levels on the cell surface. Rapid ligand unbinding in endosomes then leads to preferential EGFR recycling rather than its degradation (Ebner & Derynck, 1991). Therefore, TGF-alpha, compared to EGF, induces lesser receptor degradation, and thus can be a more potent mitogenic stimulant. However, some studies show that EGF and TGF-alpha can elicit similar mitogenic responses, even though, the dynamics of receptor activation they induce differ (Reddy et al., 1998).

Hence, kinetic properties of ligand-receptor interactions can affect not only the immediate signaling responses, but also the number of receptors which remain on the cell surface after several hours of persistent ligand stimulation. Thus, unlike the case of EGF stimulation, in which the number of EGFR receptors remaining on the cell surface displays the memory of EGF doses, TGF-stimulated system might not exhibit the memory of ligand dose through the number of cell surface receptors. Alternatively, the memory of the two ligands might have different time-span, as replenishment of degraded receptors (as in the case of EGF signaling), may require more time for *de novo* synthesis of receptors, compared to replenishment of receptors removed through internalization, which would rely on their recycling to the cell surface.

We can use our model to simulate phosphorylation responses to treatments with various combinations of different fold-changes in EGF and TGF-alpha. We can then compare how EGFR stores the memory of EGF and TGF-alpha separately and explore the time-scales of each of these two memories. We can then test our predictions experimentally by measuring phosphorylation

responses to the combinations of doses and fold-changes predicted to be most informative by the model. To have significant predictive power, the model may need to be fitted to the two types of time-courses of signal responses to the treatments with either EGF or TGF-alpha. The model allows to implement several possible modes of action of the two ligands: for example, the interactions of each of the ligands with receptors can be assumed independent, mutually exclusive, or negatively co-operative. The experimental measurements of cell responses to combinations of treatments with different fold-changes in ligand doses could then help to identify the correct mode of interaction.

Another possible example of combinations of ligands to verify fold-change detection properties could be EGF and HRG. Such combinatorial treatments have been previously modeled in Birtwistle et al., 2007, however the study did not consider repeated stimulations with different fold-changes in ligand combinations.

Our model can be used to study the effects of various other ligand mixtures, and is not limited to EGF and TGF-alpha, as the identity of the ligand in the model is encoded mainly through parameters of its binding and unbinding to receptor on the cell surface, and in the endosomes.

Section 5.6.2: How ErbB and other receptors together can remember the same ligand

It can be interesting to investigate how various receptors of the ErbB family, namely ErbB1, 2, 3 and 4, co-expressed on the cell surface can together remember the same ligand and if having multiple types of receptors involved in responding and memorizing one ligand could provide any advantages in cell signaling.

For example, EGF binding to EGFR can lead to formation of dimers not only between EGFR receptors themselves, but also lead to heterodimerization with other members of the ErbB family, such as ErbB2, ErbB3, ErbB4. Therefore, in the course of EGF stimulation, not only EGFR levels may be affected, but the number of receptors co-dimerizing with EGFR can change as well through receptors trafficking and degradation. As a result of initial stimulation, changes in the numbers of different types of cell surface receptors could then alter cell sensitivity to consecutive stimulations with the same ligand.

Another scenario to consider is when different receptors can directly bind the same ligand. For instance, ErbB3 and ErbB4 can bind heregulin, a growth factor molecule important in cell proliferation, invasion, and differentiation (Breuleux, 2007). Although ErbB3 receptors themselves are kinase-impaired, they can form dimers with EGFR and thus lead to activation of EGFR phosphorylation signaling cascade in response to heregulin (Chen et al., 2009). One can explore how the memory of heregulin stimulations could be distributed between ErbB3 and ErbB4 (receptor species able to bind heregulin) and what implications this can provide for cell responses to consequent stimulations with heregulin.

Section 5.7: Generalizing mechanism of Weber's Law to other receptors sensory systems

The ubiquity of EGF molecule throughout multiple tissues in the organism suggests that the presented here Weber's Law in sensing relative changes in EGF levels discovered in mammalian epithelial cells could be generalizable to many other types of cells relying on EGF-activated EGFR signaling. In addition, EGFR receptors can be activated by multiple other ligands (Birtwistle et al., 2007) and therefore, the discovered here Weber's Law could facilitate sensing relative changes in levels of many other agonists of EGFR. Moreover, other receptors signaling cascades for which the derived here conditions hold (rapid endocytosis of activated receptors and a fast post-translational inactivating mechanism) could potentially realize receptors-mediated relative sensing of extracellular stimuli. Overall, the discovered here mechanism of Weber's Law could be very general and represent a result of convergent evolution in cell functioning.

In particular, the presented here mechanism of Weber's Law in sensing relative changes in EGF concentrations relies on general enough assumptions about EGFR receptors signaling, specifically, that ligand-bound receptors are internalized significantly faster than ligand-free receptors, and that activated receptors are deactivated by a rapid post-translational mechanism, such as dephosphorylation by phosphatases. Given that endocytosis is a general enough cell mechanism which can internalize various other receptors (Sigismund et al., 2012), and that various receptors can be rapidly inactivated via multiple other post-translational modification mechanisms (methylation, phosphorylation) (He, Fong, Zastrow, & Whistler, 2001; Vladimirov & Sourjik, 2009), the derived here

requirements for Weber's Law can be applicable to other receptors sensory systems, and therefore, the discovered Weber's Law mechanism may be generalizable from EGFR receptors signaling to other mammalian receptor-signaling systems with rapid endocytosis of activated receptors.

For example, relative sensing in extracellular stimuli levels can be present in signaling through the G-protein coupled receptors (GPCR). GPCR receptors constitute a large and diverse family of cell surface receptors and in mammals are responsible for various senses (Kobilka, 2007), including vision (Luo et al., 2008). In vision, dependence of the GPCR receptors sensitivity on background light has been suggested to follow Weber's Law (Luo et al., 2008), however, there are not many quantitative studies on the role of the GPCR receptors in this phenomenon. Since, the GPCR receptors can be desensitized through many various mechanisms, including endocytosis of receptors (Ferguson & Caron, 1998) and their rapid deactivation by phosphorylation, the mechanism of Weber's Law derived in this work (Section 4.7.2) may be applicable to relative sensing of light through GPCR in mammalian vision. Moreover, given the expanse of GPCR in realizing multiple other functions of perception, the same Weber's Law mechanism may apply to other human senses.

Another interesting example where Weber's Law in sensing relative changes in levels of extracellular stimuli can be realized through the presented in this work mechanism, is signaling through opioid receptors and the associated with it opioid tolerance and addiction (Nestler & Aghajanian, 1997). Opioid receptors belong to the GPCR family of receptors (Al-Hasani, Ream, 2011). The opioid receptors can be rapidly desensitized via phosphorylation and arrestin binding, preventing activation of G proteins. Endocytosis of the opioid receptors has been implicated in the development of tolerance to analgesic drugs (He et al., 2001). Although, the tolerance may arise through multiple other mechanisms, downstream of the receptor-level, endocytosis of receptors may play an important role in regulating cell sensitivity to the stimulant. Given that the assumptions used to derive Weber's Law mechanism presented in this work apply to the regulation of the opioid receptors, the opioid receptors may realize Weber's Law through the presented here receptors-based mechanism. A quantitative experimental validation of this phenomenon potentially could have important implications for the understanding of drug tolerance and addiction.

Lastly, degradation of ligand-bound internalized receptors may not be required for realizing Weber's Law through a receptors-dependent mechanism. Biochemical properties of ligand-receptor interactions can determine the fate of internalized ligand-bound receptors, by direct them either to the lysosomal degradation (as is the case of EGFR and opioid receptors) or to recycling to the cell surface (EpoR) (Alberts B, Johnson A, Lewis J, et al, 2002). It is possible that the removal of ligand-bound receptors from the cell surface could alter cell sensitivity to consecutive stimulations with the same ligand, and the fate of internalized receptors (degraded or recycled) and speed of recovery of functional receptors to the cell surface could influence the duration of the cell memory of past ligand doses and thus the ability to realize Weber's Law in sensing relative changes in stimulus levels.

Going beyond the receptor-mediated mechanisms of Weber's Law in sensing relative changes in extracellular stimuli, one might explore other methods of cellular memory of past stimuli treatments and a consequent possibility of the relative sensing in the stimuli levels. For example, one may investigate the presence of Weber's Law in signaling through Epo receptors, which activate red blood cell production and can distinguish the levels of Epo ligand across several orders of magnitude. A recent study (Becker et al., 2010) indicated that the Epo receptors signaling may rely on efficient depletion of the extracellular Epo from the cell surface via ligand binding to receptors and subsequent internalization of the molecule-bound receptor. If this is a prominent mechanism of EpoR signaling attenuation, the Epo receptors in this way erase the memory of the ligand stimulation from their cell surface, and therefore would not be able to retain the memory of ligand concentration through establishing the equilibrium in the number of receptors on the cell surface which would be related to the Epo dose. The Epo signaling leads to mobilization of pools of Epo receptors to the cell surface if high doses of ligand are present. Hence, if Weber's Law is indeed present in EpoR signaling, its ligand memory would have to be realized through a mechanism distinct from the cell surface receptor-based mechanism presented here for the EGFR system.

A more general direction of inquiry could be to investigate how physiological environment of specific cells and their biological functions shape the receptors sensory mechanisms realized in the cells. In particular, this work demonstrates that the discovered here Weber's Law in EGFR signaling is observed in the physiological range of EGF doses. Therefore, it is plausible that Weber's Law, and

possibly other mechanisms of regulation of the EGFR signaling system, arose under evolutionary pressure to be attuned to the EGF levels characteristic to various tissues *in vivo*.

Section 5.8: Endocytosis in the context of cell signaling - a mechanism to integrate information about cell environment into cell fate decisions

In vivo cells are being constantly exposed to various stimuli over time. How cells process multiple stimuli from their environment and integrate them into their cell fate decision making remains an active area of research (Albeck et al., 2013). This work emphasizes the role of endocytosis as a mechanism by which the cells can regulate abundances of their cell surface receptors in response to stimuli exposure, therefore, adjusting their sensitivity to consequent ligand stimulations. Thus, in the context of cell signaling endocytosis can be considered as a mechanism for cells to integrate the information about their past stimuli exposures into their cell fate decisions.

Traditionally, endocytosis has been considered as a tool to terminate cell signaling through EGFR and to render cells unresponsive to consecutive stimulations with EGF (Kholodenko et al., 2010). However, such view was derived based on experiments with bolus-stimulations of cells with very high, non-physiological levels of EGF which lead to saturation of EGFR phosphorylation responses and therefore removal of most of the functioning EGFR receptors from the cell surface (Becker et al., 2010). An alternative view suggests that the role of endocytosis may be to deliver signaling closer to the downstream signaling targets, such as, for example, phospho-ERK (Kholodenko, 2009). The presented here results suggest a more nuanced role of endocytosis: in particular, we think that endocytosis is instrumental in regulating and maintaining the dynamic steady state of the cells and their sensitivity to the time-varying environment. By adjusting the number of functioning cell surface receptors according to the dose of the applied chronic ligand stimulation, endocytosis thus affects cellular signaling responses and ultimately, cell fate decision making.

Interplay of endocytosis with other mechanisms of receptors signalling downregulation provides the cells with flexibility in regulating their sensitivity to repeated extracellular stimuli. In particular, using solely receptors endocytosis and degradation to terminate signalling would render cells unresponsive to consequent ligand stimulations and cells resensitization would then rely on a slow and costly receptors *de novo* synthesis. Hence, action of phosphoatases, or other mechanisms

of receptors downregulation, in addition to receptors endocytosis would allow cells not to completely degrade the receptors in response to initial pre-stimulation with EGF, but to maintain a certain number of ligand-binding ready receptors on the cell surface, ensuring the cells are sensitive to ligand restimulations. As an alternative to receptors degradation, activation of transcription of various negative regulators of EGFR (such as phosphatases) could be a more advantageous strategy of signalling downregulation for the cells in the context of time-varying ligand concentrations.

In vivo, the stimuli to which the cells are exposed may not necessarily have a constant profile over time. For example, due to molecular noise, the stimulus level can fluctuate around a particular constant value. Since, the process of endocytosis is not instantaneous and operates with a certain time delay, it allows the cells to integrate the applied stimulus over a certain time frame and thus adjust the number of cell surface receptors according to some average value of the applied stimulus, allowing robustness against the noise in stimulus levels.

The solutions of the analytical model in Section 4.7.2 can be modified to replace chronic stimulation with a constant dose of EGF, u, into a time-varying function of time $u \to u + f(u,t)$, where the term $f(u,t)$ represents the noisy fluctuations which, for example, could be modeled by a sinusoid function. In such case, given that the activation term of the derivative of the total number of receptors (Section 4.7.2, expression (8)) is linear in u, the phosphorylation term will linearly depend on $u + f(u,t)$. The total number of remaining receptors still depends on rapid endocytosis of activated receptors (Section 4.7.2, expression (8)), and therefore, the information about ligand stimulation will still be reflected in the number of remaining receptors, controlled via integral of the produced phosphorylation signal over time. Therefore, if the stimulus is given as $u + f(u,t)$, endocytosis would integrate the applied signal for some time and the contribution of the sinusoid noise term $f(u,t)$ would be insignificant compared to the integral of the signal in u. In case the ligand input profile deviates from step-input in constant levels of EGF, the cells ability to respond to fold changes in the average levels of the extracellular stimuli would be preserved, given the change in the averages happens sufficiently fast, and the cells do not slowly adjust their sensitivity to the change, but rather rapidly respond to the step increase in EGF.

CONCLUSIONS

This work presents an experimental evidence of the ability of EGFR expressing cells to sense relative changes in levels of extracellular EGF. In addition, we reveal a mechanism behind such property. The phenomenon of relative sensing in stimuli levels is referred to as Weber's Law. For the first time, this work addressed Weber's Law in relative sensing of extracellular ligands in mammalian phosphorylation signaling cascades. We find that the relative sensing in EGF ligand in MCF-10A cells occurs exactly within the reported physiological range of EGF doses. Given the ubiquity of EGF and of the EGFR receptors throughout mammalian tissues, the presented here mechanism might exist across many mammalian cell signaling systems. The discovered mechanism of Weber's Law relies on fast endocytosis of ligand-bound receptors and on additional rapid deactivation of receptors through a post-translational modification. Such Weber's Law enabling features of receptors regulation are common to many receptors sensory systems and therefore the discovered mechanism of relative sensing might be generalizable to many other mammalian receptors sensory systems. For example, for the G-protein coupled receptors responsible for multiple human senses there is an evidence of relative sensing of light stimuli. The key challenge in the future will be to quantitatively validate this phenomenon in the GPCR and other mammalian receptor sensory systems. If applicable to multiple receptor signaling systems, the receptor-mediated mechanism of relative sensing of stimuli could then present a result of convergent evolution in functioning of cell signaling cascades.

In addition, the importance of the EGFR receptors endocytosis in the discovered mechanism of Weber's Law suggests a novel, unconventional framework of understanding the role of endocytosis in cell functioning *in vivo*: enabling cells to dynamically adjust and maintain their sensitivity to time-varying stimuli, endocytosis allows cells to incorporate the information about their environment into their cell fate decision making.

REFERENCES

Albeck, J. G., Mills, G. B., & Brugge, J. S. (2013). Frequency-modulated pulses of ERK activity transmit quantitative proliferation signals. *Molecular cell*, *49*(2), 249–61. doi:10.1016/j.molcel.2012.11.002

Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002

Al-Hasani, Ream, M. R. B. (2011). Molecular Mechanisms of Opioid Receptor-Dependent Signaling and Behavior, *115*(6), 1363–1381. doi:10.1097/ALN.0b013e318238bba6.Molecular

Alon, U., Surette, M. G., Barkai, N., & Leibler, S. (1999). Robustness in bacterial chemotaxis. *Nature*, *397*(6715), 168–71. doi:10.1038/16483

Andrabi, S., Gjoerup, O. V, Kean, J. A., Roberts, T. M., & Schaffhausen, B. (2007). Protein phosphatase 2A regulates life and death decisions via Akt in a context-dependent manner, 2–7.

Aoki, K., Kumagai, Y., Sakurai, A., Komatsu, N., Fujita, Y., Shionyu, C., & Matsuda, M. (2013). Stochastic ERK activation induced by noise and cell-to-cell propagation regulates cell density-dependent proliferation. *Molecular cell*, *52*(4), 529–40. doi:10.1016/j.molcel.2013.09.015

Aoki, K., Yamada, M., Kunida, K., Yasuda, S., & Matsuda, M. (2011). Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proceedings of the National Academy of Sciences*, 1–6. doi:10.1073/pnas.1104030108

Apgar, J. F., Toettcher, J. E., Endy, D., White, F. M., & Tidor, B. (2008). Stimulus design for model selection and validation in cell signaling. *PLoS computational biology*, *4*(2), e30. doi:10.1371/journal.pcbi.0040030

Arroyo, J. D., & Hahn, W. C. (2005). Involvement of PP2A in viral and cellular transformation. *Oncogene*, *24*(52), 7746–55. doi:10.1038/sj.onc.1209038

Arteaga, C. L., & Engelman, J. a. (2014). ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer cell*, *25*(3), 282–303. doi:10.1016/j.ccr.2014.02.025

Avraham, R., & Yarden, Y. (2011). Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nature reviews. Molecular cell biology*, *12*(2), 104–17. doi:10.1038/nrm3048

Baselga, J., & Swain, S. M. (2009). Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. *Nature reviews. Cancer*, *9*(7), 463–75. doi:10.1038/nrc2656

Batchelor, E., Silhavy, T., & Goulian, M. (2004). Continuous control in bacterial regulatory circuits. *Journal of bacteriology*, *1655*(31), 7618–7625. doi:10.1128/JB.186.22.7618

Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., … Aebersold, R. (2011). The quantitative proteome of a human cell line. *Molecular systems biology*, *7*(549), 549. doi:10.1038/msb.2011.82

Becker, V., Schilling, M., Bachmann, J., Baumann, U., Raue, A., Maiwald, T., … Klingmüller, U. (2010). Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science (New York, N.Y.)*, *328*(5984), 1404–8. doi:10.1126/science.1184913

Bermudez, O., Jouandin, P., Rottier, J., Bourcier, C., Pagès, G., & Gimond, C. (2011). Post-transcriptional regulation of the DUSP6/MKP-3 phosphatase by MEK/ERK signaling and hypoxia. *Journal of cellular physiology*, *226*(1), 276–84. doi:10.1002/jcp.22339

Birk, D., Gansauge, F., Gansauge, S., & Formentini, A. (1999). Serum and Correspondent Tissue Measurements of Epidermal Growth Factor (EGF) and Epidermal Growth Factor Receptor (EGF-R): Clinical Relevance in Pancreatic Cancer and Chronic Pancreatitis. *International journal of gastrointestinal cancer*, *25*(2), 89–96. doi:10.1385/IJGC:25:2:89

Birtwistle, M. R., Hatakeyama, M., Yumoto, N., Ogunnaike, B. a, Hoek, J. B., & Kholodenko, B. N. (2007). Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. *Molecular systems biology*, *3*(144), 144. doi:10.1038/msb4100188

Birtwistle, M. R., Rauch, J., Kiyatkin, A., Aksamitiene, E., Dobrzyński, M., Hoek, J. B., … Kholodenko, B. N. (2012). Emergence of bimodal cell population responses from the interplay between analog single-cell signaling and protein expression noise. *BMC systems biology*, *6*, 109. doi:10.1186/1752-0509-6-109

Blüthgen, N. (2006). Sequestration shapes the response of signal transduction cascades. *IUBMB life*, *58*(11), 659–63. doi:10.1080/15216540600994340

Breuleux, M. (2007). Role of heregulin in human cancer. *Cellular and molecular life sciences : CMLS*, *64*(18), 2358–77. doi:10.1007/s00018-007-7120-0

Brooks, S. P., & Gelman, A. (2010). General Methods for Monitoring Convergence of Iterative Simulations. *America*, *7*(4), 434–455.

Changeux, J.-P. (2012). Allostery and the Monod-Wyman-Changeux model after 50 years. *Annual review of biophysics*, *41*, 103–33. doi:10.1146/annurev-biophys-050511-102222

Chen, D., Waters, S. B., Holt, K. H., & Pessin, J. E. (1996). SOS Phosphorylation and Disassociation of the Grb2-SOS Complex by the ERK and JNK Signaling Pathways *. *Biochemistry*, *271*(11), 6328–6332.

Chen, W. W., Niepel, M., & Sorger, P. K. (2010). Classic and contemporary approaches to modeling biochemical reactions. *Genes & development*, *24*(17), 1861–75. doi:10.1101/gad.1945410

Chen, W. W., Schoeberl, B., Jasper, P. J., Niepel, M., Nielsen, U. B., Lauffenburger, D. a, & Sorger, P. K. (2009). Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Molecular systems biology*, *5*(239), 239. doi:10.1038/msb.2008.74

Cheong, R., Rhee, A., Wang, C. J., Nemenman, I., & Levchenko, A. (2011). Information transduction capacity of noisy biochemical signaling networks. *Science (New York, N.Y.)*, *334*(6054), 354–8. doi:10.1126/science.1204553

Cirit, M., Wang, C.-C., & Haugh, J. M. (2010). Systematic quantification of negative feedback mechanisms in the extracellular signal-regulated kinase (ERK) signaling network. *The Journal of biological chemistry*, *285*(47), 36736–44. doi:10.1074/jbc.M110.148759

Citri, A., & Yarden, Y. (2006). EGF-ERBB signalling: towards the systems level. *Nature reviews. Molecular cell biology*, *7*(7), 505–16. doi:10.1038/nrm1962

Cobleigh, B. M. A., Vogel, C. L., Tripathy, D., Robert, N. J., Scholl, S., Fehrenbacher, L., … Slamon, D. J. (1999). M u l t i n a t i o n a l S t u d y o f t h e Effi c a c y a n d S a f e t y o f H u m a n i z e d Anti-HER2 Monoclonal Antibody in Women Who Have HER2-Overexpressing Metastatic Breast Cancer That Has Progressed After Chemotherapy for Metastatic Disease, *9*.

Cohen, a a, Geva-Zatorsky, N., Eden, E., Frenkel-Morgenstern, M., Issaeva, I., Sigal, A., … Alon, U. (2008). Dynamic proteomics of individual cancer cells in response to a drug. *Science (New York, N.Y.)*, *322*(5907), 1511–6. doi:10.1126/science.1160165

Cohen-Saidon, C., Cohen, A. a, Sigal, A., Liron, Y., & Alon, U. (2009). Dynamics and variability of ERK2 response to EGF in individual living cells. *Molecular cell*, *36*(5), 885–93. doi:10.1016/j.molcel.2009.11.025

Collins, S., Caron, M. G., & Lefkowitz, R. J. (1992). From ligand binding to gene expression: new insights into the regulation of G-protein-coupled receptors. *Trends in biochemical sciences*, *17*(1), 37–9.

Corder, G. W.; Foreman, D. I. (2014). Nonparametric Statistics: A Step-by-Step Approach. Wiley. ISBN 978-1118840313.

Countaway, J. L., Nairn, a C., & Davis, R. J. (1992). Mechanism of desensitization of the epidermal growth factor receptor protein-tyrosine kinase. *The Journal of biological chemistry*, *267*(2), 1129–40.

Ebisuya, M., Kondoh, K., & Nishida, E. (2005). The duration, magnitude and compartmentalization of ERK MAP kinase activity: mechanisms for providing signaling specificity. *Journal of cell science*, *118*(Pt 14), 2997–3002. doi:10.1242/jcs.02505

Ebner, R., & Derynck, R. (1991). Epidermal growth factor and transforming growth factor-a : differential intracellular routing and processing of ligand-receptor complexes, (August), 599–612.

Ferguson, S. S., & Caron, M. G. (1998). G protein-coupled receptor adaptation mechanisms. *Seminars in cell & developmental biology*, *9*(2), 119–27. doi:10.1006/scdb.1997.0216

Fisher, D. a, & Lakshmanan, J. (1990). Metabolism and effects of epidermal growth factor and related growth factors in mammals. *Endocrine reviews*, *11*(3), 418–42. doi:10.1210/edrv-11-3-418

Fujioka, A., Terai, K., Itoh, R. E., Aoki, K., Nakamura, T., Kuroda, S., … Matsuda, M. (2006). Dynamics of the Ras/ERK MAPK cascade as monitored by fluorescent probes. *The Journal of biological chemistry*, *281*(13), 8917–26. doi:10.1074/jbc.M509344200

Gibbons, J. (1993). *Nonparametric statistics: An introduction Corder. Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki.*

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J., eds. (1996), Markov Chain Monte Carlo in Practice, London: Chapman and Hall

Goentoro, L., & Kirschner, M. W. (2009). Evidence that fold-change, and not absolute level, of beta-catenin dictates Wnt signaling. *Molecular cell*, *36*(5), 872–84. doi:10.1016/j.molcel.2009.11.017

Goentoro, L., Reeves, G. T., Kowal, C. P., Martinelli, L., Schüpbach, T., & Shvartsman, S. Y. (2006). Quantifying the Gurken morphogen gradient in Drosophila oogenesis. *Developmental cell*, *11*(2), 263–72. doi:10.1016/j.devcel.2006.07.004

Goentoro, L., Shoval, O., Kirschner, M. W., & Alon, U. (2009). The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular cell*, *36*(5), 894–9. doi:10.1016/j.molcel.2009.11.018

Goh, L. K., Huang, F., Kim, W., Gygi, S., & Sorkin, A. (2010). Multiple mechanisms collectively regulate clathrin-mediated endocytosis of the epidermal growth factor receptor. *The Journal of cell biology*, *189*(5), 871–83. doi:10.1083/jcb.201001008

Goldbeterl, A., & Koshland, E. (1984). Ultrasensitivity in Biochemical Systems Controlled by Covalent Modification, 14441–14447.

Goldsmith, Z. G., & Dhanasekaran, D. N. (2007). G protein regulation of MAPK networks. *Oncogene*, *26*(22), 3122–42. doi:10.1038/sj.onc.1210407

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS computational biology*, *3*(10), 1871–78. doi:10.1371/journal.pcbi.0030189

Hart, Y., Mayo, A. E., Shoval, O., & Alon, U. (2013). Comparing apples and oranges: fold-change detection of multiple simultaneous inputs. *PloS one*, *8*(3), e57455. doi:10.1371/journal.pone.0057455

He, L., Fong, J., Zastrow, M. Von, & Whistler, J. L. (2001). Regulation of Opioid Receptor Trafficking and Morphine Tolerance by Receptor Oligomerization, 271–282.

Heinrich, R., Neel, B. G., & Rapoport, T. a. (2002). Mathematical models of protein kinase signal transduction. *Molecular cell*, *9*(5), 957–70.

Hemmings, B. a, & Restuccia, D. F. (2012). PI3K-PKB/Akt pathway. *Cold Spring Harbor perspectives in biology*, *4*(9), a011189. doi:10.1101/cshperspect.a011189

Hendriks, B. S. (2003). EGFR & HER2 Trafficking and Signaling DYnamics: Experimental and Modeling Studies. *Chemical Engineering*.

Hendriks, B. S., Opresko, L. K., Wiley, H. S., & Lauffenburger, D. a. (2003). Quantitative analysis of HER2-mediated effects on HER2 and epidermal growth factor receptor endocytosis: distribution of homo- and heterodimers depends on relative HER2 levels. *The Journal of biological chemistry*, *278*(26), 23343–51. doi:10.1074/jbc.M300477200

Herbst, R. S. (2004). Review of epidermal growth factor receptor biology. *International journal of radiation oncology, biology, physics*, *59*(2 Suppl), 21–6. doi:10.1016/j.ijrobp.2003.11.041

Herskowitz, I. (1995). MAP kinase pathways in yeast: for mating and more. *Cell*, 187–197.

Hirai M, Gamou S, Minoshima S, S. N. (1988). Two Independent Mechanisms for Escaping Epidermal Growth Factor- mediated Growth Inhibition in Epidermal Growth Factor Receptor-hyperproducing Human Tumor Cells Flow Cytometric Analysis of Cell Volume. *The Journal of cell biology*, *107*(August), 791–799.

Hornberg, J. J., Bruggeman, F. J., Binder, B., Geest, C. R., de Vaate, a J. M. B., Lankelma, J., … Westerhoff, H. V. (2005). Principles behind the multifarious control of signal transduction. ERK phosphorylation and kinase/phosphatase control. *The FEBS journal*, *272*(1), 244–58. doi:10.1111/j.1432-1033.2004.04404.x

Huyer, G., Liu, S., Kelly, J., Moffat, J., Payette, P., Kennedy, B., … Ramachandran, C. (1997). Mechanism of Inhibition of Protein-tyrosine Phosphatases by Vanadate and Pervanadate. *Journal of Biological Chemistry*, *272*(2), 843–851. doi:10.1074/jbc.272.2.843

Junttila, M. R., Li, S.-P., & Westermarck, J. (2008). Phosphatase-mediated crosstalk between MAPK signaling pathways in the regulation of cell survival. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, *22*(4), 954–65. doi:10.1096/fj.06-7859rev

Kajikawa, K., Yasui, W., Sumiyoshi, H., Yoshida, K., Nakayama, H., Ayhan, A., … Tahara, E. (1989). Expression of epidermal growth factor in human tissues Immunohistochemical and biochemical analysis, (1991), 27–32.

Karasarides, M., Chiloeches, A., Hayward, R., Niculescu-Duvaz, D., Scanlon, I., Friedlos, F., … Marais, R. (2004). B-RAF is a therapeutic target in melanoma. *Oncogene*, *23*(37), 6292–8. doi:10.1038/sj.onc.1207785

Keymer, J. E., Endres, R. G., Skoge, M., Meir, Y., & Wingreen, N. S. (2006). Chemosensing in Escherichia coli: two regimes of two-state receptors. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(6), 1786–91. doi:10.1073/pnas.0507438103

Keyse, S. M. (2008). Dual-specificity MAP kinase phosphatases (MKPs) and cancer. *Cancer metastasis reviews*, *27*(2), 253–61. doi:10.1007/s10555-008-9123-1

Kholodenko, B. N. (2009). Spatially distributed cell signalling. *FEBS letters*, *583*(24), 4006–12. doi:10.1016/j.febslet.2009.09.045

Kholodenko, B. N., Demin, O. V, Moehren, G., & Hoek, J. B. (1999). Quantification of short term signaling by the epidermal growth factor receptor. *The Journal of biological chemistry*, *274*(42), 30169–81.

Kholodenko, B. N., Hancock, J. F., & Kolch, W. (2010). Signalling ballet in space and time. *Nature reviews. Molecular cell biology*, *11*(6), 414–26. doi:10.1038/nrm2901

Kim, E. K., & Choi, E.-J. (2010). Pathological roles of MAPK signaling pathways in human diseases. *Biochimica et biophysica acta*, *1802*(4), 396–405. doi:10.1016/j.bbadis.2009.12.009

Kleiman, L. B. (2010). Experimental and Computational Analysis of Epidermal Growth Factor Receptor Pathway Phosphorylation Dynamics By. *Systems Biology*.

Kleiman, L. B., Maiwald, T., Conzelmann, H., Lauffenburger, D. a, & Sorger, P. K. (2011). Rapid phospho-turnover by receptor tyrosine kinases impacts downstream signaling and drug binding. *Molecular cell*, *43*(5), 723–37. doi:10.1016/j.molcel.2011.07.014

Kobilka, B. K. (2007). G protein coupled receptor structure and activation. *Biochimica et biophysica acta*, *1768*(4), 794–807. doi:10.1016/j.bbamem.2006.10.021

Kohberger, R. (1978). A method for parameter sensitivity analysis in differential equation models. *Water Resources …*, *14*.

Koshland, D. E., Goldbeter, A., & Stock, J. B. (1982). Amplification and Adaptation in Regulatory and Sensory Systems, *217*(July).

Kullback, Solomon. "The kullback-leibler distance." American Statistician 41.4 (1987): 340-340.

Kuppuswamy, D., & Pike, L. J. (1991). Desensitization of the EGF receptor alters its ability to undergo EGF-induced dimerization. *Cellular signalling*, *3*(2), 107–17.

Lazova, M., & Ahmed, T. (2011). Response rescaling in bacterial chemotaxis. *Proceedings of the ….* doi:10.1073/pnas.1108608108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1108608108

Lazzara, M. J., & Lauffenburger, D. a. (2009). Quantitative modeling perspectives on the ErbB system of cell regulatory processes. *Experimental cell research*, *315*(4), 717–25. doi:10.1016/j.yexcr.2008.10.033

Lee, R. E. C., Walker, S. R., Savery, K., Frank, D. a, & Gaudet, S. (2014). Fold Change of Nuclear NF-κB Determines TNF-Induced Transcription in Single Cells. *Molecular cell*, 1–13. doi:10.1016/j.molcel.2014.01.026

Lemmon, M. a, & Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. *Cell*, *141*(7), 1117–34. doi:10.1016/j.cell.2010.06.011

Lequin, R. M. (2005). Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA). *Clinical chemistry*, *51*(12), 2415–8. doi:10.1373/clinchem.2005.051532

Levchenko, A., Bruck, J., & Sternberg, P. W. (2000). Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(11), 5818–23.

Liepe, J., Filippi, S., Komorowski, M., & Stumpf, M. P. H. (2013). Maximizing the information content of experiments in systems biology. *PLoS computational biology*, *9*(1), e1002888. doi:10.1371/journal.pcbi.1002888

Luo, D., Kefalov, V., Yau, K., Rods, P., In, C., Basbaum, A. I., … Albright, T. D. (2008). The Senses: A comprehensive Reference.

Ma, W., Trusina, A., El-Samad, H., Lim, W. a, & Tang, C. (2009). Defining network topologies that can achieve biochemical adaptation. *Cell*, *138*(4), 760–73. doi:10.1016/j.cell.2009.06.013

Macdonald, J. L., & Pike, L. J. (2008). Heterogeneity in EGF-binding affinities arises from negative cooperativity in an aggregating system. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(1), 112–7. doi:10.1073/pnas.0707080105

Macdonald-Obermann, J. L., & Pike, L. J. (2014). Different epidermal growth factor (EGF) receptor ligands show distinct kinetics and biased or partial agonism for homodimer and heterodimer formation. *The Journal of biological chemistry*, *289*(38), 26178–88. doi:10.1074/jbc.M114.586826

Manning, B. D., & Cantley, L. C. (2007). AKT/PKB signaling: navigating downstream. *Cell*, *129*(7), 1261–74. doi:10.1016/j.cell.2007.06.009

Marquez-Lago, T. T., & Leier, A. (2011). Stochastic adaptation and fold-change detection: from single-cell to population behavior. *BMC systems biology*, *5*(1), 22. doi:10.1186/1752-0509-5-22

Martelli, C., Carlson, J. R., & Emonet, T. (2013). Intensity invariant dynamics and odor-specific latencies in olfactory receptor neuron response. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *33*(15), 6285–97. doi:10.1523/JNEUROSCI.0426-12.2013

McEwen, B. S. (2001). *The Endocrine System-VOLUME IV: Coping With The Environment: Neural and Endocrine Mechanisms* (pp. doi:10.1097/00024382–200116010–00017).

Meloche, S., & Pouysségur, J. (2007). The ERK1/2 mitogen-activated protein kinase pathway as a master regulator of the G1- to S-phase transition. *Oncogene*, *26*(22), 3227–39. doi:10.1038/sj.onc.1210414

Mesibov, R., & Adler, J. (1972). Chemotaxis toward amino acids in Escherichia coli. *Journal of bacteriology*, *112*(1), 315–26.

Mesibov, R., Ordal, G. W., & Adler, J. (1973). The range of attractant concentrations for bacterial chemotaxis and the threshold and size of response over this range. Weber law and related phenomena. *The Journal of general physiology*, *62*(2), 203–23.

Milde-Langosch, K. (2005). The Fos family of transcription factors and their role in tumourigenesis. *European journal of cancer (Oxford, England : 1990)*, *41*(16), 2449–61. doi:10.1016/j.ejca.2005.08.008

Moelling, K., Schad, K., Bosse, M., Zimmermann, S., & Schweneker, M. (2002). Regulation of Raf-Akt Cross-talk. *The Journal of biological chemistry*, *277*(34), 31099–106. doi:10.1074/jbc.M111974200

Murphy, L. O., Smith, S., Chen, R.-H., Fingar, D. C., & Blenis, J. (2002). Molecular interpretation of ERK signal duration by immediate early gene products. *Nature cell biology*, *4*(8), 556–64. doi:10.1038/ncb822

Murphy, M. S. (1998). Growth factors and the gastrointestinal tract. *Nutrition (Burbank, Los Angeles County, Calif.)*, *14*(10), 771–4.

Nachev, V., Stich, K. P., & Winter, Y. (2013). Weber's law, the magnitude effect and discrimination of sugar concentrations in nectar-feeding animals. *PloS one*, *8*(9), e74144. doi:10.1371/journal.pone.0074144

Nestler, E. J., & Aghajanian, G. K. (1997). Molecular and Cellular Basis of Addiction.

Olsen, D. A., Bechmann, T., Østergaard, B., Wamberg, P. a, Jakobsen, E. H., & Brandslund, I. (2012). Increased concentrations of growth factors and activation of the EGFR system in breast cancer. *Clinical chemistry and laboratory medicine : CCLM / FESCC*, *50*(10), 1809–18. doi:10.1515/cclm-2011-0823

Osborne, C. K., Hamilton, B., Titus, G., & Livingston, R. B. (1980). Epidermal Growth Factor Stimulation of Human Breast Cancer Cells in Culture Epidermal Growth Factor Stimulation of Human Breast Cancer Cells in, 2361–2366.

Östman, A., & Böhmer, F. (2001). Regulation of receptor tyrosine kinase signaling by protein tyrosine phosphatases, *2*(June), 258–266.

Pennock, S., & Wang, Z. (2003). Stimulation of Cell Proliferation by Endosomal Epidermal Growth Factor Receptor As Revealed through Two Distinct Phases of Signaling Stimulation of Cell Proliferation by Endosomal Epidermal Growth Factor Receptor As Revealed through Two Distinct Phases of. *Molecular and cellular biology*. doi:10.1128/MCB.23.16.5803

Reddy, C., Wells, A., & Lauffenburger, D. a. (1998). Comparative mitogenic potencies of EGF and TGFα and their dependence on receptor-limitation versus ligand-limitation. *Medical and Biological …*, 499–507.

Reynolds, A. R., Tischer, C., Verveer, P. J., Rocks, O., & Bastiaens, P. I. H. (2003). EGFR activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation. *Nature cell biology*, *5*(5), 447–53. doi:10.1038/ncb981

Riddle, V. G. H., Pardee, A. B., & Rossow, P. W. (1979). Growth Control of Normal and Transformed Cells, *538*, 529–538.

Roepstorff, K., Grandal, M. V., Henriksen, L., Knudsen, S. L. J., Lerdrup, M., Grøvdal, L., … van Deurs, B. (2009). Differential effects of EGFR ligands on endocytic sorting of the receptor. *Traffic (Copenhagen, Denmark)*, *10*(8), 1115–27. doi:10.1111/j.1600-0854.2009.00943.x

Rogers, K. W., & Schier, A. F. (2011). Morphogen gradients: from generation to interpretation. *Annual review of cell and developmental biology*, *27*, 377–407. doi:10.1146/annurev-cellbio-092910-154148

Rommel, C., Clarke, B. a, Zimmermann, S., Nuñez, L., Rossman, R., Reid, K., … Glass, D. J. (1999). Differentiation stage-specific inhibition of the Raf-MEK-ERK pathway by Akt. *Science (New York, N.Y.)*, *286*(5445), 1738–41.

Rosenfeld, N., Perkins, T. J., Alon, U., Elowitz, M. B., & Swain, P. S. (2006). A fluctuation method to quantify in vivo fluorescence data. *Biophysical journal*, *91*(2), 759–66. doi:10.1529/biophysj.105.073098

Rouger, V., Goaillard, J., & Marguet, D. (2014). Physiological Epidermal Growth Factor Concentrations Activate High Affinity Receptors to Elicit Calcium Oscillations, (September). doi:10.1371/journal.pone.0106803

Sasagawa, S., Ozaki, Y., Fujita, K., & Kuroda, S. (2005a). Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nature cell biology*, *7*(4), 365–73. doi:10.1038/ncb1233

Sasagawa, S., Ozaki, Y., Fujita, K., & Kuroda, S. (2005b). Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nature cell biology*, *7*(4), 365–73. doi:10.1038/ncb1233

Savir, Y., Escalante-chong, R., & Springer, M. (2005). Yeast as a ratio sensor : the ratio of glucose and galactose determines cell state, 1.

Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D., & Müller, G. (2002). Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature biotechnology*, *20*(4), 370–5. doi:10.1038/nbt0402-370

Segel, L. a, Goldbeter, A., Devreotes, P. N., & Knox, B. E. (1986). A mechanism for exact sensory adaptation based on receptor modification. *Journal of theoretical biology*, *120*(2), 151–79.

Selimkhanov, J., Taylor, B., Yao, J., & Pilko, A. (2014). Accurate information transmission through dynamic biochemical signaling networks. *Science*, (December).

Sella, G., & Hirsh, A. (2005). The application of statistical physics to evolutionary biology. *Proceedings of the National Academy of …*, (Track II).

Shankaran, H., Resat, H., & Wiley, H. S. (2007). Cell surface receptors for signal transduction and ligand transport: a design principles study. *PLoS computational biology*, *3*(6), e101. doi:10.1371/journal.pcbi.0030101

Shankaran, H., Zhang, Y., Chrisler, W. B., Ewald, J. a, Wiley, H. S., & Resat, H. (2012). Integrated experimental and model-based analysis reveals the spatial aspects of EGFR activation dynamics. *Molecular bioSystems*, *8*(11), 2868–82. doi:10.1039/c2mb25190f

Shaul, Y. D., & Seger, R. (2007). The MEK/ERK cascade: from signaling specificity to diverse functions. *Biochimica et biophysica acta*, *1773*(8), 1213–26. doi:10.1016/j.bbamcr.2006.10.005

Shin, S.-Y., Rath, O., Choo, S.-M., Fee, F., McFerran, B., Kolch, W., & Cho, K.-H. (2009). Positive- and negative-feedback regulations coordinate the dynamic behavior of the Ras-Raf-MEK-ERK signal transduction pathway. *Journal of cell science*, *122*(Pt 3), 425–35. doi:10.1242/jcs.036319

Shinar, G., Milo, R., & Alon, U. (2007). Input – output robustness in simple bacterial signaling systems, (36), 19931–19935.

Shoval, O., Goentoro, L., Hart, Y., Mayo, A., Sontag, E., & Alon, U. (2010). Fold-change detection and scalar symmetry of sensory input fields. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(36), 15995–6000. doi:10.1073/pnas.1002352107

Sigismund, S., Algisi, V., Nappo, G., Conte, A., Pascolutti, R., Cuomo, A., … Di Fiore, P. P. (2013). Threshold-controlled ubiquitination of the EGFR directs receptor fate. *The EMBO journal*, *32*(15), 2140–57. doi:10.1038/emboj.2013.149

Sigismund, S., Argenzio, E., Tosoni, D., Cavallaro, E., Polo, S., & Di Fiore, P. P. (2008). Clathrin-mediated internalization is essential for sustained EGFR signaling but dispensable for degradation. *Developmental cell*, *15*(2), 209–19. doi:10.1016/j.devcel.2008.06.012

Sigismund, S., Confalonieri, S., Ciliberto, A., Polo, S., Scita, G., & Di Fiore, P. P. (2012). Endocytosis and signaling: cell logistics shape the eukaryotic cell plan. *Physiological reviews*, *92*(1), 273–366. doi:10.1152/physrev.00005.2011

Sigismund, S., Woelk, T., Puri, C., Maspero, E., Tacchetti, C., Transidico, P., … Polo, S. (2005). Clathrin-independent endocytosis of ubiquitinated cargos. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(8), 2760–5. doi:10.1073/pnas.0409817102

Singh, A. B., & Harris, R. (2005). Autocrine, paracrine and juxtacrine signaling by EGFR ligands. *Cellular signalling*, *17*(10), 1183–93. doi:10.1016/j.cellsig.2005.03.026

Skataric, M., & Sontag, E. (2012). A characterization of scale invariant responses in enzymatic networks. *PLoS computational biology*, *8*(11), e1002748. doi:10.1371/journal.pcbi.1002748

Skouta, R., Hayano, M., Shimada, K., & Stockwell, B. R. (2012). Design and synthesis of Pictet-Spengler condensation products that exhibit oncogenic-RAS synthetic lethality and induce non-apoptotic cell death. *Bioorganic & medicinal chemistry letters*, *22*(17), 5707–13. doi:10.1016/j.bmcl.2012.06.077

Sontag, E. (2001). Protein phosphatase 2A : the Trojan Horse of cellular signaling, *13*.

Soulsby, M., & Bennett, A. M. (2009). Physiological signaling specificity by protein tyrosine phosphatases. *Physiology (Bethesda, Md.)*, *24*, 281–9. doi:10.1152/physiol.00017.2009

Sturm, O. E., Orton, R., Grindlay, J., Birtwistle, M. R., Vyshemirsky, V., Gilbert, D., … Kolch, W. (2010). The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. *Science signaling*, *3*(153), ra90. doi:10.1126/scisignal.2001212

Sulis, M., & Parsons, R. (2003). PTEN: from pathology to biology. *Trends in Cell Biology*, *13*(9), 478–483. doi:10.1016/S0962-8924(03)00175-2

Swat, A., Dolado, I., Rojas, J. M., & Nebreda, A. R. (2009). Cell density-dependent inhibition of epidermal growth factor receptor signaling by p38alpha mitogen-activated protein kinase via Sprouty2 downregulation. *Molecular and cellular biology*, *29*(12), 3332–43. doi:10.1128/MCB.01955-08

Takeda, K., Shao, D., Adler, M., Charest, P. G., Loomis, W. F., Levine, H., … Firtel, R. a. (2012). Incoherent feedforward control governs adaptation of activated ras in a eukaryotic chemotaxis pathway. *Science signaling*, *5*(205), ra2. doi:10.1126/scisignal.2002413

Teeffelen, S. van, Shaevitz, J., & Gitai, Z. (2012). Image analysis in fluorescence microscopy: Bacterial dynamics as a case study. *Bioessays*, *34*(5), 427–436. doi:10.1002/bies.201100148.Image

Tiganis, T. (2002). Protein tyrosine phosphatases: dephosphorylating the epidermal growth factor receptor. *IUBMB life, 53*(1), 3–14. doi:10.1080/15216540210811

Toettcher, J. E., Weiner, O. D., & Lim, W. a. (2013). Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/Erk module. *Cell*, *155*(6), 1422–34. doi:10.1016/j.cell.2013.11.004

Tu, Y., Shimizu, T. S., & Berg, H. C. (2008). Modeling the chemotactic response of Escherichia coli to time-varying stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(39), 14855–60. doi:10.1073/pnas.0807569105

Uyemura, T., Takagi, H., Yanagida, T., & Sako, Y. (2005). Single-molecule analysis of epidermal growth factor signaling that leads to ultrasensitive calcium response. *Biophysical journal*, *88*(5), 3720–30. doi:10.1529/biophysj.104.053330

Van de Poll, M. L. M., van Rotterdam, W., Gadellaa, M. M., Jacobs-Oomen, S., & van Zoelen, E. J. J. (2005). Ligand depletion negatively controls the mitogenic activity of epidermal growth factor. *Experimental cell research*, *304*(2), 630–41. doi:10.1016/j.yexcr.2004.12.011

Vijver, M. Van de, Kumar, R., & Mendelsohn, J. (1991). Ligand-induced activation of A431 cell epidermal growth factor receptors occurs primarily by an autocrine pathway that acts upon receptors on the surface rather than. *Journal of Biological Chemistry*, 7503–7508.

Vladimirov, N., & Sourjik, V. (2009). Chemotaxis: how bacteria use memory. *Biological chemistry*, *390*(11), 1097–104. doi:10.1515/BC.2009.130

Wang, C.-C., Cirit, M., & Haugh, J. M. (2009). PI3K-dependent cross-talk interactions converge with Ras as quantifiable inputs integrated by Erk. *Molecular systems biology*, *5*(246), 246. doi:10.1038/msb.2009.4

Wang, F., Travins, J., DeLaBarre, B., Penard-Lacronique, V., Schalm, S., Hansen, E., … Yen, K. E. (2013). Targeted Inhibition of Mutant IDH2 in Leukemia Cells Induces Cellular Differentiation. *Science* , *340* (6132), 622–626. doi:10.1126/science.1234769

Weber, EH (1905). Tatsinn and Gemeingefuhl. Verl Von Wilhelm Englemann Leipz Ger

Weemen, B. Van, & Schuurs, A. (1971). Immunoassay using antigen—enzyme conjugates. *FEBS letters*, (June 1971), 232–236.

Wiley, H. S., & Cunningham, D. (1981). A Steady State Model for Analyzing the, (August 1981), 433–440.

Wiley, H. S., Herbst, J. J., Walsh, B. J., Lauffenburger, D. a, Rosenfeld, M. G., & Gill, G. N. (1991). The role of tyrosine kinase activity in endocytosis, compartmentation, and down-regulation of the epidermal growth factor receptor. *The Journal of biological chemistry*, *266*(17), 11083–94.

Wiley, H. S., Shvartsman, S. Y., & Lauffenburger, D. a. (2003). Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends in cell biology*, *13*(1), 43–50.

Yarden, Y. (2001). The EGFR family and its ligands in human cancer. signalling mechanisms and therapeutic opportunities. *European journal of cancer (Oxford, England : 1990)*, *37 Suppl 4*, S3–8.

Yarden, Y., & Pines, G. (2012). The ERBB network: at last, cancer therapy meets systems biology. *Nature reviews. Cancer*, *12*(8), 553–63. doi:10.1038/nrc3309

Zahnow, C. a. (2006). ErbB receptors and their ligands in the breast. *Expert reviews in molecular medicine*, *8*(23), 1–21. doi:10.1017/S146239940600010X

Zhou, X., & Agazie, Y. M. (2012). The signaling and transformation potency of the overexpressed HER2 protein is dependent on the normally-expressed EGFR. *Cellular signalling*, *24*(1), 140–50. doi:10.1016/j.cellsig.2011.08.015

Zi, Z., & Klipp, E. (2007). Steady state analysis of signal response in receptor trafficking networks. *Genome Inform*.

APPENDIX

Section A.1: Model Equations, Parameter descriptions, Priors on the parameter values

This section provides description of model parameters, differential equations, and initial conditions,
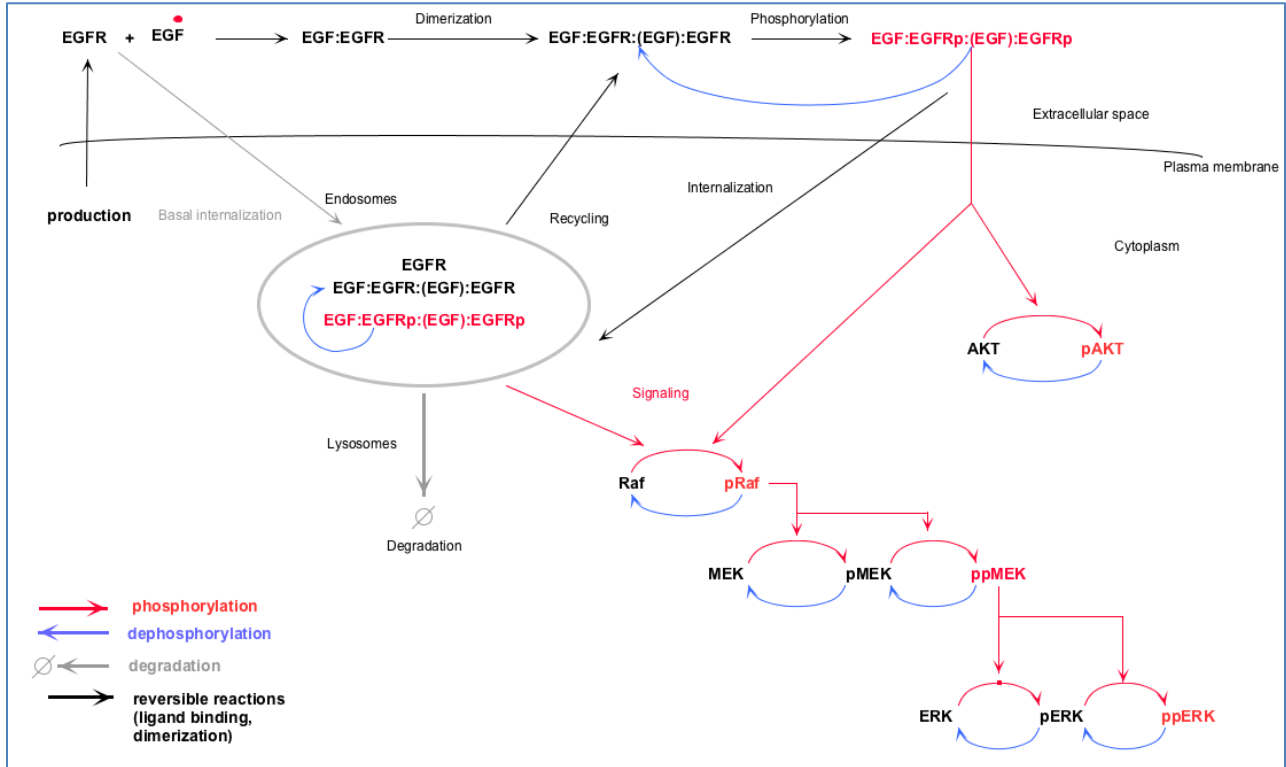
for the ErbB signaling model.

Figure A-1: Schematics of the extended model of the ErbB signaling cascade.
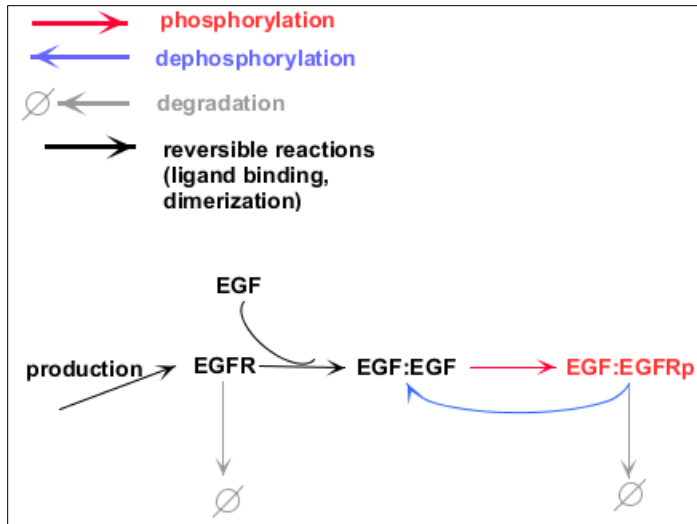
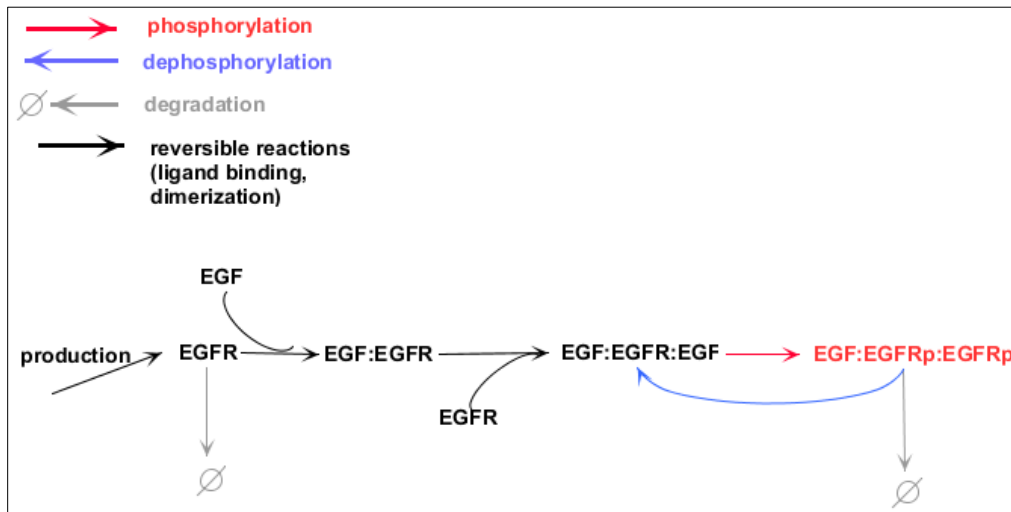**Figure A-2: Simplified analytical kinetic ODE model of the ErbB signaling: linear in EGF activation of EGFR.**

Figure A-3:Simplified analytical kinetic ODE model of ErbB signaling: one type of EGFR dimer.
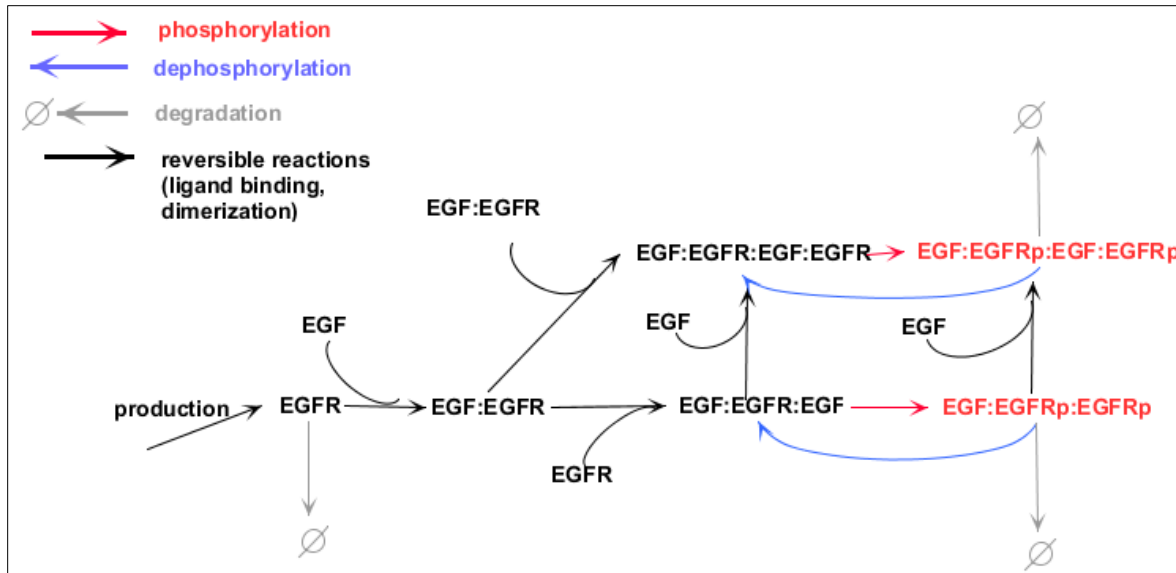
Figure A-4: Simplified analytical kinetic ODE model of the ErbB signaling: two types of EGFR dimers.
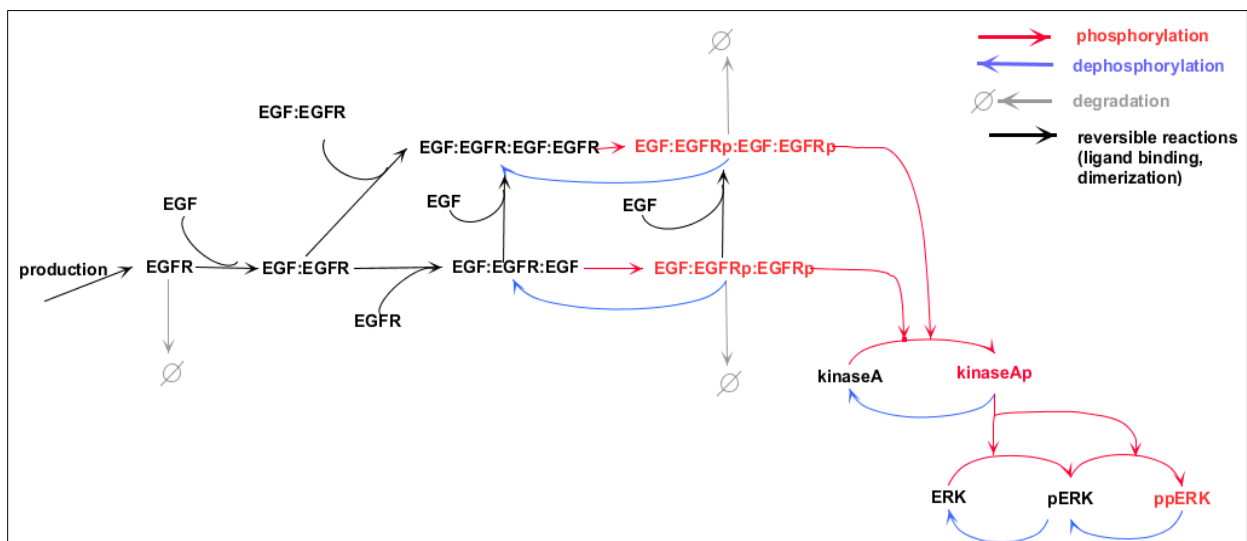
Figure A-5: Simplified analytical kinetic ODE model: two types of EGFR dimers, kinase cascade leading to activation of ERK simplified into one effective component.

Section A.1.1: Description of model parameters and imposed priors

This section presents a list of protein species, their corresponding variable names and initial conditions (in number of molecules) and a list of parameters in model's differential equations, and their corresponding initial values.

parameter name	initial condtions	Description	priors (bounds in log-uniform space)
c1	variable	EGF	
c531	c531	EGFR	[4..6.5]
c10	0	EGF	
c3	0	EGF:EGFR	
c4	0	EGF:EGFR:EGFR	
c2	0	EGF:EGFR:EGF:EGFR	
c5	0	EGF:EGFR#p:EGF:EGFR#p	
c6	0	EGF:EGFR#p:EGFR#p	
c5312	0	EGFR	
c22	0	EGF:EGFR:EGF:EGFR	
c42	0	EGF:EGFR:EGFR	
c52	0	EGF:EGFR#p:EGF:EGFR#p	
c62	0	EGF:EGFR#p:EGFR#p	
c86	0	degraded receptors	
c61	0	EGF:EGFR#p:EGF:EGFR#p:ERK	
c63	0	EGF:EGFR#p:EGFR#p:ERK	
c610	0	EGF:EGFR#p:EGF:EGFR#p:ERK	
c630	0	EGF:EGFR#p:EGFR#p:ERK	
c612	0	EGF:EGFR#p:EGF:EGFR#p:ERK#p	
c632	0	EGF:EGFR#p:EGF:EGFR#p:ERK#p	
c6102	0	EGF:EGFR#p:EGF:EGFR#p:ERK#p	
c6302	0	EGF:EGFR#p:EGFR#p:ERK#p	
c8	c8	ERK, cytoplasm	[3..7]
c80	0	ERK#p, cytoplasm	
c800	0	ERK#p#p, cytoplasm	

Rate constants:

rate constants	descriptions	priors
k1	EGF binds to ligand-unbound monomer	[-14..-4]
kd1	EGF unbinds from ligand-bound monomer	[-3..1]
k31	EGF binds to one-EGF bound dimer	[-14..4]
kd31	EGF unbinding from two-EGF bound dimers	[-3..1]
k2	dimerization of one-EGF bound and one free monomers	[-8..-2]
kd2	undimerization of one-EGF bound and one free monomers	[-2.5..2]
k2a	dimerization of two EGF-bound monomers	[-8..-2]
kd2a	undimerization of two EGF-bound monomers	[-2.5..2]

251

kd123	kcat of EGFR activation	[-4..2]
kd950	kcat of EGFR dephosphorylation	[-1.75..-1.15]
k60	degradation of active receptors	[-6..-2]
k77	production of monomers	[-2..2]
k910	internalization of inactive receptors	[-8..-4]
kd910	recycling of inactive receptors	[-4..-2]
k91	internalization of active receptors	[-4.5..-2]
kd91	recycling of active receptors	[-4.5..-2.5]
k61	degradation of inactive receptors	[-6..-2]
k50	activated receptors binding to inactive ERK	[-8..-4]
kd50	unbinding of activated receptors from inactive ERK	[-2..0]
kd51	kcat of ERK activation	[-3..-1]
kd5	kcat of ERK dephosphorylation by phosphtatase	[-3...-1]

Section A.1.2: Equations of the detailed dynamic ODE model of ErbB signaling

Differential equations of the ErbB model: the left-hand side of an equation is the time derivative of the number of molecules of a species. The right-hand side is a collection of first or second order mass action reaction terms between species that contribute to the generation or degradation of the species on the left-hand side.

```
dc531/dt =k77 - vk1 * c1  * c531 + vkd1 * c3  -k2 *c531 *c3 +kd2 *c4-k910*c531 +
kd910 *c5312
dc3/dt = vk1 * c1  * c531 - vkd1 * c3-k2 *c531 *c3 + kd2 *c4  -2*k2 *c3 *c3 +2*
kd2a *c2
dc4/dt =k2 *c531 *c3 -kd2 *c4 -vk31*c1 *c4 +  vkd31*c2 - kd123 *c4 +kd950 *c6-k91
*c4  + kd91  *c42
dc2/dt = k2 *c3 *c3 -  kd2a *c2 +vk31*c1 *c4 -  vkd31*c2-kd123 *c2 +kd950 *c5 -k91
*c2  +kd91  *c22
dc5/dt = -kd950 *c5 +kd123 *c2 +vk31*c1 *c6 - vkd31*c5 -k91  *c5 + kd91  *c52 -k50
*c8 *c5  + kd50 *c61+kd51 *c61  -k50 *c80 *c5 + kd50 *c612 +kd51 *c612
dc6/dt =  kd123 *c4  -kd950 *c6-vk31*c1 *c6 - vkd31*c5 -k91  *c6 + kd91  *c62  -k50
*c8 *c6 + kd50 *c63  +kd51 *c63  -k50 *c80 *c6+  kd50 *c632  -kd51 *c632
dc5312/dt  = k910  *c531 - kd910 *c5312-k61 *c5312
dc22/dt  = - k61 *c22  +kd950 *c52 -kd123 *c22  +k91  *c2  -kd91  *c22
dc42/dt  = +k91  *c4  - kd91  *c42   -kd123 *c42  +kd950 *c62
dc52/dt  = -k60 *c52 +k91  *c5 - kd91  *c52  -kd950 *c52 +kd123 *c22 - k50 *c8 *c52
+ kd50 *c610 +kd51 *c610  -k50 *c80 *c52 + kd50 *c6102 +kd51 *c6102
dc62/dt  =  - k60 *c62 +k91  *c6 - kd91  *c62  -kd950 *c62+kd123 *c42 -k50 *c8 *c62
+ kd50 *c630+kd51 *c630 -k50 *c80 *c62  + kd50 *c6302 +kd51 *c6302
dc86/dt  = k60 *c52 + k60 *c62+ k61 *c22
dc61/dt  = k50 *c8 *c5  - kd50 *c61-kd51 *c61
dc63/dt  = k50 *c8 *c6 - kd50 *c63- kd51 *c63
dc610/dt  = k50 *c8 *c52  - kd50 *c610-kd51 *c610
dc630/dt  = k50 *c8 *c62 - kd50 *c630-kd51 *c630
dc612/dt  = k50 *c80 *c5 - kd50 *c612 -kd51 *c612
dc632/dt  = k50 *c80 *c6-  kd50 *c632 -kd51 *c632
dc6102/dt  =k50 *c80 *c52 - kd50 *c6102-kd51 *c6102
```

```
dc6302/dt =k50 *c80 *c62  - kd50 *c6302+kd51 *c6302
dc8/dt =-k50 *c8 *c5  + kd50 *c61-k50 *c8 *c6 + kd50 *c63 - k50 *c8 *c52  + kd50
*c610-k50 *c8 *c62 + kd50 *c630+kd5 *c80
dc80/dt = +kd51 *c61 +kd51 *c63 +kd51 *c610-k50 *c80 *c5 + kd50 *c612  -k50 *c80
*c6+ kd50 *c632 -k50 *c80 *c52 + kd50 *c6102-k50 *c80 *c62  + kd50 *c6302   + kd5
*c800  -kd5 *c80
dc800/dt =kd51 *c612  +kd51 *c632+kd51 *c6102+kd51 *c6302 -kd5 *c800
```

Algebraic variables for counting total amount of species:
```
total_egfr= 2*c612 +2*c632 +2*c6102 +2*c6302 +2*c61 +2*c63 +2*c610 +2*c630
+ c531 +c5312 +c3 +2*c2 +2*c4 +2*c5 +2*c6 +2*c22 +2*c42 +2*c52 +2*c62
total_egfr_surf=2*c612 +2*c632 + 2*c61 +2*c63 +  c531 +c3 +2*c2 +2*c4
+2*c5 +2*c6
total_pegfr= 2*c612 +2*c632 +2*c6102 +2*c6302 + 2*c61 +2*c63 +2*c610
+2*c630 +2*c5 +2*c6 +2*c52 +2*c62
total_egf =c1
total_pperk = c800
total_perk =  c6102 +c6302 +c80
total_erk = c612 +c632 + c8
```

Section A.1.3: Detailed model of ErbB signaling from Chen et al., 2009

The detailed model of ErbB signaling simulated in Section 4.2 has been previously published

in (Chen et al., 2009) and can be found at:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2644173/bin/msb200874-s2.zip

Section A.2: Gelman-Rubin criterion for MCMC chains convergence

To assess convergence of MCMC chains we apply Gelman and Rubin criterion for multiple

sequence diagnostic (Gilks, Richardson, and Spiegelhalter, 1996). The criterion is applied to

trajectories of MCMC chains in each individual parameter dimension.

The following algorithm steps (for each parameter):

1. Run m ≥ 2 chains of length 2n started from overdispersed starting points found by

simulated annealing

2. As a burnoff, discard the first n steps in each chain

3. Calculate the within-chain and between-chain variance as follows:

Compute within chain variance for the jth chain, sj, and obtain W, the mean of the variances

of each chain:

$$W = \frac{1}{m}\sum_{j=1}^{m} s_j^2 \qquad\qquad B = \frac{n}{m-1}\sum_{j=1}^{m}(\bar{\theta}_j - \bar{\bar{\theta}})^2$$

$$s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\theta_{ij} - \bar{\theta}_j)^2 \qquad\qquad \bar{\bar{\theta}} = \frac{1}{m}\sum_{j=1}^{m}\bar{\theta}_j$$

4. Estimated variance of the parameter as a weighted sum of the within-chain and between-chain variance:

$$\hat{\mathrm{Var}}(\theta) = (1 - \frac{1}{n})W + \frac{1}{n}B$$

5. Compute the potential scale reduction factor, Rhat:

$$\hat{R} = \sqrt{\frac{\hat{\mathrm{Var}}(\theta)}{W}}$$

The MCMC chains are ran until Rhat is reduced to approximately 1.1-1.2 (heuristically derived estimate).

Section A.3: Preliminary comparison of predicted parameter distributions of cell lines MCF-10A, MCF-7, SKBR3
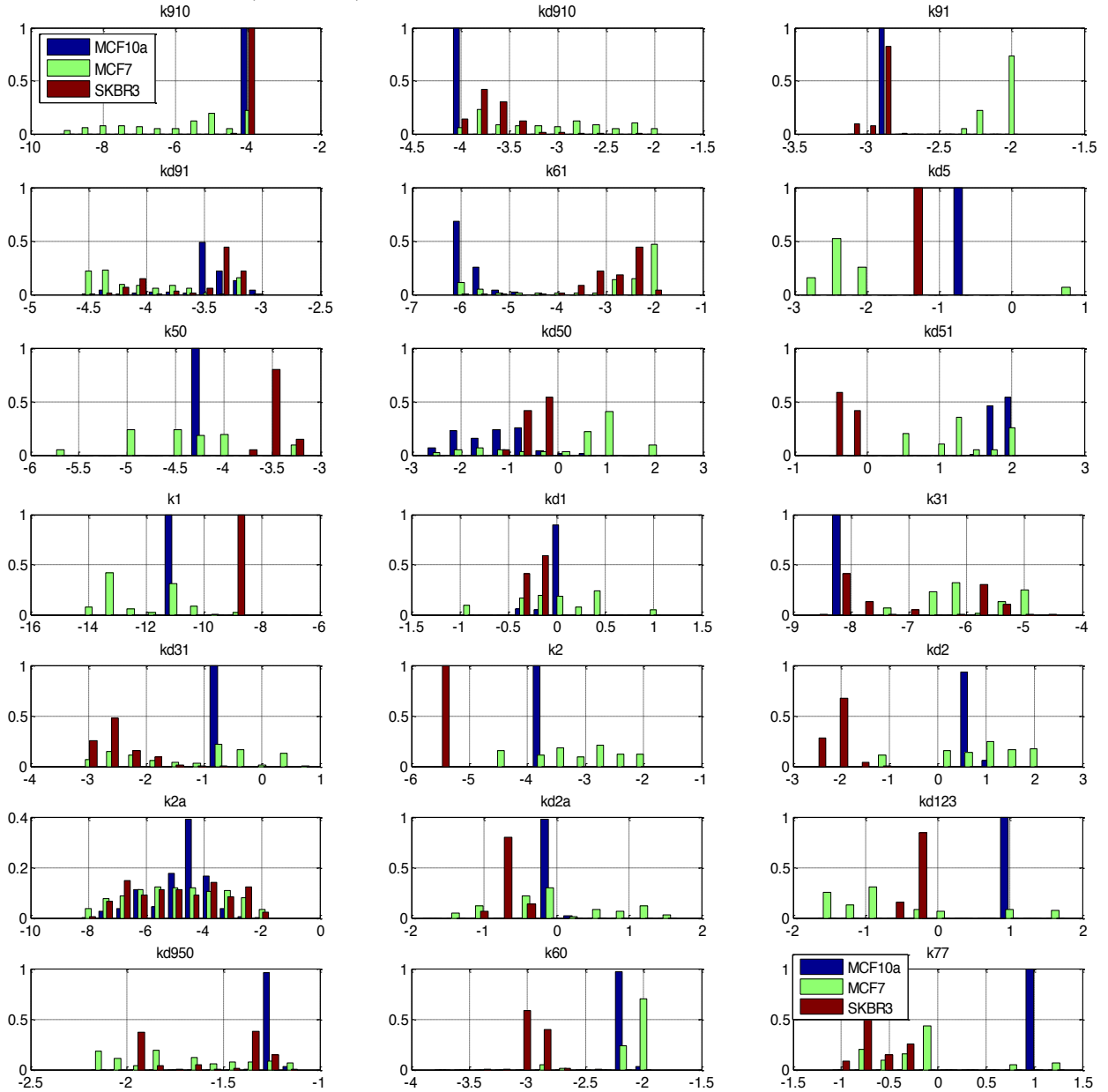

Figure A-6: Comparison of the preliminary predicted distributions of parameters obtained from fitting the model to the experimental data from three different cell lines: MCF-10A, MCF-7, SKBR3.