Summer 6-21-2018

# STUDYING EGFR SIGNALING THROUGH SINGLE MOLECULE IMAGING AND COMPUTATIONAL MODELING

Emanuel Salazar Cavazos
*University of New Mexico*

Emanuel Salazar Cavazos
*Candidate*

Biomedical Sciences - Pathology
*Department*

This dissertation is approved, and it is acceptable in quality and form for publication:

*Approved by the Dissertation Committee:*

Diane S. Lidke, PhD , Chairperson

William S. Hlavacek, PhD

Bridget S. Wilson, PhD

Angela Wandinger-Ness, PhD

# STUDYING EGFR SIGNALING THROUGH SINGLE MOLECULE IMAGING AND COMPUTATIONAL MODELING

**by**

**EMANUEL SALAZAR CAVAZOS**

B.S., Genomic Biotechnology, Universidad Autónoma
de Nuevo León, México, 2009
M.S., Biomedical Engineering and Physics, Center for Research
and Advanced Studies of the National Polytechnic
Institute, México, 2012

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

**Doctor of Philosophy**
**Biomedical Sciences**

The University of New Mexico
Albuquerque, New Mexico

**July 2018**

# ACKNOWLEDGEMENTS

I would like to acknowledge my parents for all the support and love they have provided me throughout my entire life. They have always been there to encourage me to achieve whatever goals I decide to pursue. To the rest of my family, for always remaining a close and supportive family.

To my mentor, Diane Lidke, for giving me constant guidance during all these years, for being so patient and supportive, and for helping me become a better scientist. I have learned too many things from you that I could write pages about it. One very important trait I learned from you, and that I want to continue applying during my scientific career, is to have a strong commitment for rigorous and high-quality science.

To my dissertation committee, for always giving me valuable advice during all the committee meetings and whenever I needed help. Bill – first of all, for welcoming me in your lab before joining the PhD program, and showing me the power of collaboration between computational modelers and experimentalists; for helping me become a better computational modeler and making possible for this work to integrate computational modeling with quantitative biology. Bridget – For sharing with me your expertise about cell signaling, for your insightful comments that improved the quality of this work, and for helping me become a better science writer and presenter. Angela – For bringing a different perspective during our meetings, and for providing me with valuable advice related to this work and career advice for my future steps as a scientist.

To all the members of the Wilson/Lidke lab, for always being an incredibly supportive and collaborative group. From each of the current and past members I have learned different things, and without their help I could have never completed this work.

**STUDYING EGFR SIGNALING THROUGH SINGLE MOLECULE IMAGING AND COMPUTATIONAL MODELING**

**by**

**EMANUEL SALAZAR CAVAZOS**

**B.S., Genomic Biotechnology, Universidad Autónoma de Nuevo León, México, 2009**

**M.S., Biomedical Engineering and Physics, Center for Research and Advanced Studies of the National Polytechnic Institute, México, 2012**

**Ph.D., Biomedical Sciences, University of New Mexico, USA, 2018**

## ABSTRACT

Signaling through the Epidermal Growth Factor Receptor (EGFR) plays an important role in both physiological and cancer-related processes. In this work, single-molecule microscopy measurements and computational modeling were closely integrated to better understand the mechanisms that regulate EGFR signaling. Technical improvements were made over the previously described Single-Molecule Pull-down (SiMPull) assay to facilitate direct detection of the phosphorylation state of thousands of individual receptors, and thereby estimate both the fraction of receptors phosphorylated at specific tyrosine residues and the frequency of multisite phosphorylation. These improvements enabled the first direct detection of multisite phosphorylation on full-length Epidermal Growth Factor Receptor (EGFR), and revealed that the extent of phosphorylation varied by tyrosine residue (biased phosphorylation). To help in understanding the underlying processes giving rise to these observations, a rule-based model for EGFR signaling was developed. The model suggested that biased phosphorylation could be explained by variations in adaptor protein abundances.

This prediction arises from the structure of the model, in which a phospho-site that is bound by an adaptor protein is sterically protected from the action of phosphatases. Testing model predictions confirmed that overexpression of the adaptor protein Grb2 leads to phosphorylation levels enhanced specifically at a site where this protein binds. Finally, this model was extended to explore the possible mechanisms leading to differential signaling induced by EGFR ligands. Model results suggest that ligand-dependent differences in dimer lifetimes lead to differential multisite phosphorylation and ubiquitination, which in turn could influence signaling kinetics and cellular outcomes.

# Table of Contents

ACKNOWLEDGEMENTS ..... iii

ABSTRACT ..... iv

Table of Contents ..... vi

List of Figures ..... xi

List of Tables ..... xiii

Chapter 1 : Introduction ..... 1

1.1 Overview ..... 1

1.2 Signaling through the Epidermal Growth Factor Receptor ..... 2

1.2.1 Introduction ..... 2

1.2.2 Recruitment of adaptor proteins to activated receptors ..... 3

1.2.3 Protein tyrosine phosphatases (PTPs) and EGFR signaling ..... 5

1.2.4 Regulation of EGFR signaling by endocytosis ..... 6

1.3 Comparison between ensemble and single-molecule techniques ..... 7

1.4 Computational modeling of EGFR signaling ..... 9

1.5 Hypothesis ..... 10

1.6 Summary of Results ..... 11

Chapter 2 : Single Molecule Pull-down for Quantification of Membrane Receptor Multisite Phosphorylation ..... 13

2.1 Introduction ..... 13

2.2 Results ..... 17

2.2.1 Visualization of individual protein phosphorylation status by SiMPull . 17

2.2.2 Simplified sample chamber increases throughput and reduces sample volume ..... 18

2.2.3 Quenching with NaBH4 reduces background autofluorescence ..... 18

2.2.4 Antibody optimization is required for accurate phosphotyrosine detection ..... 20

2.2.5 Correction is required to account for non-surface localized receptors 22

2.2.6 Extent of phosphorylation varies by tyrosine residue ..... 25

2.2.7 Three-color SiMPull reveals multisite phosphorylation ..... 27

2.3 Discussion ..... 35

2.4 Methods ..... 40

2.4.1 Cell lines and reagents ..... 40

2.4.2 Labeling of antibodies ..... 41

2.4.3 Cell treatment and lysate preparation ..... 42

2.4.4 Fabrication of hydrophobic arrays and surface functionalization ..... 43

2.4.5 Labeling and quantification of surface receptors ..... 44

2.4.6 Single-Molecule Pulldown and phospho-site labeling ..... 45

2.4.7 SiMPull imaging ..... 46

2.4.8 Quantification of Receptor Phosphorylation ..... 47

2.4.9 Statistical Analysis ..... 49

2.4.10 Step-photobleaching Analysis ..... 49

Chapter 3 : Building a rule-based model of the initial events of EGFR signaling 50

3.1 Introduction ..... 50

3.2 Model development, Results and Discussion ..... 52

3.2.1 Example reaction rules ..... 52

3.2.2 Reaction rules ..... 53

3.2.3 Model parameters ..... 57

Chapter 4 : Insights on EGFR signaling by integrating computational modeling and single molecule data ..... 60

4.1 Introduction ..... 60

4.2 Results ..... 61

4.2.1 Model can fit experimental data with only allowing adaptor protein abundances to vary ..... 61

4.2.2 Predicted influence of Grb2 overexpression in phosphorylation levels is observed experimentally ..... 64

4.2.3 Model predicts cell-specific phosphorylation patterns based on differences in adaptor protein abundances ..... 66

4.3 Discussion ..... 68

4.4 Methods ..... 71

4.4.1 Simulations and parameter estimation ..... 71

4.4.2 Single Molecule Pull-down experiments ..... 71

Chapter 5 : A computational model of differential signaling induced by EGFR ligands ..... 72

5.1 Introduction ..... 72

5.2 Results ..... 75

5.2.1 Model reproduces Y1068-Y1173 dual phosphorylation measured experimentally ..... 75

5.2.2 Extended model for ubiquitination is able to reproduce experimental behavior ..... 76

5.2.3 Model predicts impaired ubiquitination in epiregulin-stimulated MCF7 cells ..... 82

5.2.4 Using the model to help identify the mechanisms contributing to epiregulin-induced sustained signaling ..... 83

5.3 Discussion ..... 85

5.4 Methods ..... 88

5.4.1 Simulations and parameter estimation ..... 88

Chapter 6 : Implications and Future Studies ..... 89

6.1 Implications ..... 89

6.1.1 Significance of improvements made on SiMPull technique ..... 89

6.1.2 Significance of understanding role of adaptor protein abundances in biased phosphorylation ..... 89

6.1.3 Significance of rule-based model for EGFR activation and ubiquitination by different ligands ..... 90

6.2 Future Studies ..... 91

6.2.1 Effects of adaptor protein abundances in downstream signaling and cellular outcomes ..... 91

6.2.2 Experimental testing of model predictions for differential ubiquitination induced by EGFR ligands ..... 91

6.2.3 Study phosphorylation and downstream signaling of EGFR mutant L858R ..... 93

APPENDIX A: MATLAB scripts for analysis of SiMPull data ..... 95

APPENDIX B: BNGL files of rule-based models ..... 109

APPENDIX C: Configuration files for parameter estimation (.conf) ..... 125

REFERENCES ..... 132

# List of Figures

Figure 1.1. Ligand-induced EGFR dimerization and activation. ..... 3

Figure 1.2. Recruitment capacity of activated EGFR. ..... 5

Figure 1.3. Comparison of traditional vs rule-based modeling approaches. ..... 10

Figure 2.1. SiMPull to quantify protein phosphorylation. ..... 16

Figure 2.2. Reduction of autofluorescence with Sodium Borohydride (NaBH4). 19

Figure 2.3. Optimization of antibody labeling for accurate quantification of receptor phosphorylation. ..... 21

Figure 2.4. Correction for cellular distribution of receptors. ..... 24

Figure 2.5. The extent of phosphorylation varies by tyrosine residue. ..... 26

Figure 2.6. Effect of phosphatase inhibition or cell lysate salt concentration on detected phosphorylation levels. ..... 27

Figure 2.7. Assessment and correction of steric hindrance in sequentially incubated antibodies for 3-color SiMPull. ..... 29

Figure 2.8. SiMPull reveals EGFR multisite phosphorylation. ..... 31

Figure 2.9. Importance of multi-color imaging for accurate quantification of phosphorylation percentages. ..... 34

Figure 3.1. Graphical description of model. ..... 51

Figure 4.1. Comparison of naïve model and experimental data. ..... 62

Figure 4.2. Fitting model to experimental data by varying adaptor protein abundances. ..... 63

Figure 4.3. Predicted and observed phosphorylation kinetics in cells overexpressing Grb2. Predicted ..... 65

Figure 4.4. Phosphorylation patterns predicted for different cell types and observed for HeLa S3 cells. ..... 67
Figure 5.1. Predicted and observed dual Y1068-Y1173 phosphorylation. ..... 76
Figure 5.2. Graphical description of reactions included in extended model to consider cooperative Cbl recruitment. ..... 78
Figure 5.3 Fitting model to experimental data. ..... 79
Figure 5.4. Model validation: Ubiquitination in cells with EGFR knockdown or Cbl overexpression. ..... 81
Figure 5.5. Predicted ubiquitination and Y1068 phosphorylation levels in MCF7 cells when stimulated with EGF vs epiregulin. ..... 83
Figure 5.6. Predicted EGF- and epiregulin-induced ubiquitination and phosphorylation different EGFR expression levels. ..... 85
Figure 6.1. Single- and multi-phosphorylation in WT vs L858R mutant receptor. 93

## List of Tables

Table 1. List of parameter values selected by fitting algorithm (Chapter 4) ..... 71

Table 2. List of parameter values selected by fitting algorithm (Chapter 5) ..... 88

# Chapter 1 : Introduction

## 1.1 Overview

Complex multicellular organisms, such as the human body, depend on the ability of its individual cells to respond in a precise and coordinated fashion to stimuli coming either from other cells or from the surrounding environment. The transmission of stimuli from the exterior of the cell to its interior occurs through a process called signal transduction or cell signaling. For example, binding of ligands to membrane receptors starts a cascade of molecular events, such as protein phosphorylation and interactions, that propagate the signal to the nucleus and eventually lead to cellular responses. Even though the overall picture of many signal transduction pathways is known, many critical details remain poorly understood. One such detail is the heterogeneity in the activation states of molecules participating in signal transduction, a feature that cannot be provided by commonly used ensemble techniques. To address this issue, a single-molecule microscopy technique was improved, allowing for access to this kind of information. The work presented here focuses in the study of activation states of the Epidermal Growth Factor Receptor (EGFR). Additionally, computational models were used to help understanding the underlying processes giving rise to the observed distribution of activated states.

In this introduction, some of the features of the EGFR signaling pathway are described, including signal initiation, transmission and regulation. Next, the advantages and limitations of ensemble and single-molecule techniques are

discussed. Then, some of the previous work relating to modeling of EGFR signaling and the current advances in modeling techniques is described. The final part of this section contains the central hypothesis for this work and a summary of the results obtained.

## 1.2 Signaling through the Epidermal Growth Factor Receptor

### 1.2.1 Introduction

The Epidermal Growth Factor Receptor (EGFR) belongs to the EGFR/erbB family of Receptor Tyrosine Kinases, which consists of four members: EGFR (erbB1, HER1), HER2 (erbB2), HER3 (erbB3), and HER4 (erbB4). In general, these proteins are single-pass transmembrane proteins with an extracellular ligand binding domain and a cytoplasmic tail containing a tyrosine kinase domain (Figure 1.1) (Lemmon and Schlessinger, 2010; Yarden and Sliwkowski, 2001). As exceptions, HER2 has no known ligand, and HER3 kinase activity is dependent on initial activation by HER2 (Steinkamp et al., 2014). Ligand binding to these transmembrane proteins leads to conformational changes, receptor homo- and hetero-dimerization, kinase activation and the transphosphorylation of multiple cytoplasmic tail tyrosines (Schlessinger, 2002). These phosphotyrosines in turn provide sites for the recruitment and activation of cytoplasmic proteins, initiating signaling cascades that control numerous cellular processes such as gene expression, cell migration and cell division. Dysregulation of EGFR signaling, commonly caused by receptor overexpression

and/or mutation, has been associated with development and progression of cancer.

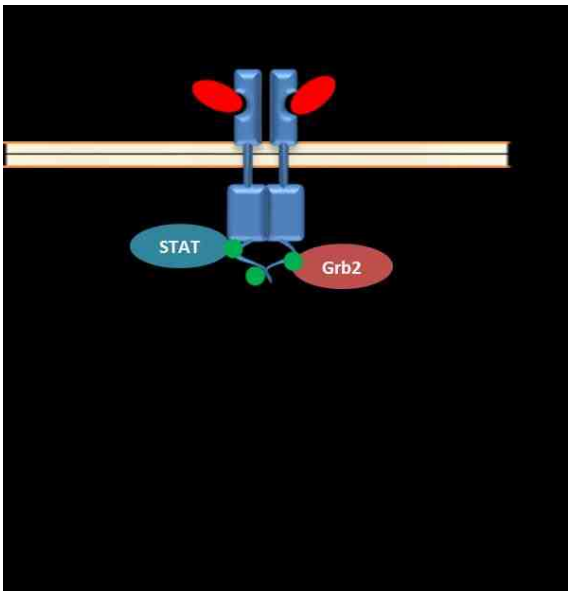


**Figure 1.1.** Ligand-induced EGFR dimerization and activation. EGFR ligands include EGF, TGF-α and epiregulin.

In recent years there has been growing evidence that suggests that EGFR activation may not only form dimers, but also tetramers and higher-order oligomers (Kozer et al., 2014, 2013a). Considering that most of the available information about EGFR activation relates to receptor dimerization, this study focuses on this form of signal initiation.

### 1.2.2 Recruitment of adaptor proteins to activated receptors

Activated EGFR is capable of recruiting a variety of adaptor proteins that have different roles in signal transduction and regulation of cellular outcomes. A few examples of these proteins and the specific sites to which they bind is displayed in Figure 1.2a. The first part of this dissertation focuses on the study of

the phosphorylation of tyrosines 1068 and 1173, and the recruitment of Grb2 and Shc1 to these sites (Figure 1.2b). Proteins recruited to the receptor can be phosphorylated either by EGFR or by other kinases such as Src. Grb2 can bind to phosphorylated Shc1, and therefore be recruited to the receptor indirectly through Shc1 (Batzer et al., 1994). Even though both Y1068 and Y1173 can recruit Grb2, either directly or indirectly, and lead to activation of MAPK pathway, these sites have distinct additional roles in signaling. For example, pY1173 can also recruit and activate PLC-γ and calcium signaling, or recruit the phosphatase Shp1, as shown in Figure 1.2a. On the other hand, pY1068 has been shown to be essential for efficient recruitment of the E-3 ubiquitin ligase Cbl, which ubiquitinates EGFR and mediates its downregulation (Sigismund et al., 2013).

The interaction of Grb2 and Cbl, and the proximity of their recruitment sites, seems to give rise to a cooperative behavior in which the stability of the Cbl-Grb2 complex is higher than that of the individual proteins (Figure 1.2c). Chapter 5 of this dissertation focuses on the computational modeling of this interaction, and how differences in stability of receptor dimers induced by distinct EGFR ligands could regulate this interaction, and eventually lead to different signaling behavior.

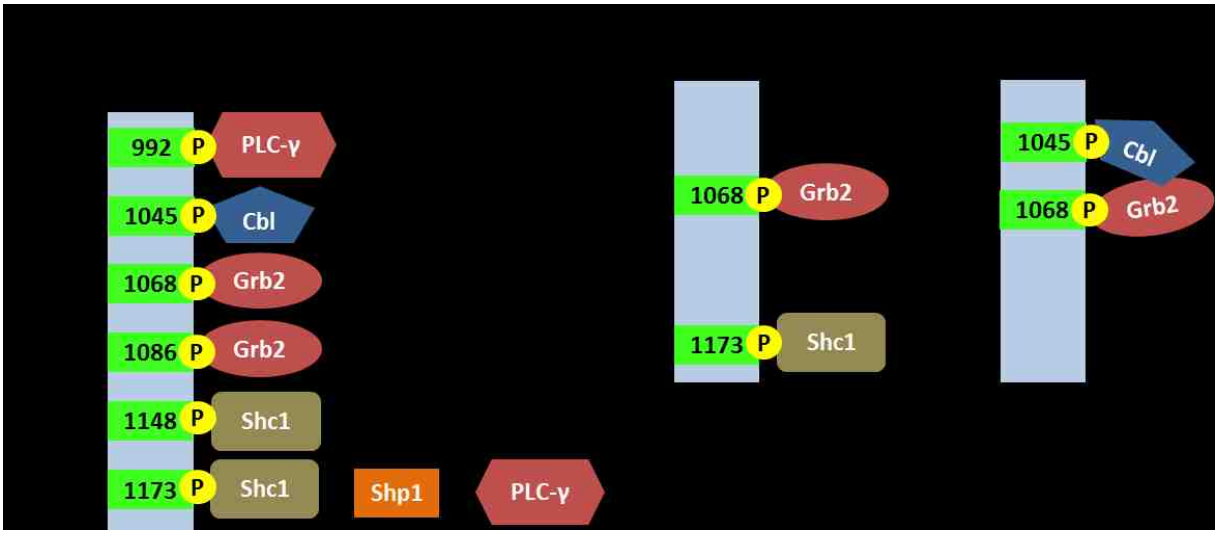


**Figure 1.2.** Recruitment capacity of activated EGFR. Description. **(a)** Representation of some of the adaptor proteins recruited to activated EGFR. Based on (Olayioye et al., 2000). **(b)** Sites and interactions considered for first part of this dissertation. **(c)** Sites and interactions considered in Chapter 5. Cbl and Grb2 can interact.

### 1.2.3 Protein tyrosine phosphatases (PTPs) and EGFR signaling

Phosphorylation levels are the net result of protein phosphorylation and dephosphorylation. A phosphate group can be added at the amino acid residues tyrosine, serine or threonine. In the case of phosphorylated tyrosines, in which this dissertation focuses, protein tyrosine phosphatases (PTPs) are the responsible for removing the phosphate groups from this amino acid. PTPs play a very important role in regulating EGFR signaling, both before and during ligand stimulation. For example, in many cell types phosphorylation of receptor tyrosine kinases (RTKs) is undetectable in the absence of ligand, but inhibition of PTPs

results in important levels of receptor phosphorylation (Jallal et al., 1992; Tarcic et al., 2009). This basal phosphorylation, which is thought to originate from short interactions between unliganded receptors, is suppressed by constantly active phosphatases. Phosphatase activity can be regulated both in time and space. For example, while DEP-1 causes receptor dephosphorylation at the membrane, PTP1B seems to mainly locate and cause EGFR dephosphorylation at endosomes (Tarcic et al., 2009; Eden et al., 2010; Yudushkin et al., 2007). Even though most PTPs are known to downregulate EGFR signaling, there are examples where a phosphatase can enhance downstream signaling (Yao et al., 2017).

The timescale of phosphorylation and dephosphorylation events was recently informed by a combination of experimental and computational modeling techniques (Kleiman et al., 2011). Kleiman *et al.* showed that activated EGFR is quickly dephosphorylated after a few seconds of adding a fast-binding kinase inhibitor. Their modeling results suggest that receptors can go through hundreds of phosphorylation/dephosphorylation cycles in the order of a few minutes. Each of these events require energy in the form of ATP. A fast dynamic interplay between these two processes is essential for allowing the cell to respond in a prompt manner to external stimuli and to prevent spurious and excessive signaling.

### 1.2.4 Regulation of EGFR signaling by endocytosis

EGFR signaling can be regulated both in time and space. After activation, receptors at the plasma membrane are internalized into endocytic vesicles.

Recent studies suggest that receptors at endosomes cannot activate MAPK pathway, the reason being that Ras, a protein required for MAPK activation, is mainly located at the plasma membrane and not in endosomes (Pinilla-Macua et al., 2016). The fate of internalized receptors is regulated by a complex network of proteins, including components of the ESCRT (endosomal sorting complexes required for transport) machinery and members of the Rab GTPase family (Sorkin and Goh, 2009; Schwartz et al., 2007). Depending on different factors, receptors can be recycled back to the plasma membrane, where they can continue signaling, or they can be targeted for lysosomal degradation, causing signal downregulation. One of the main factors regulating the fate of EGFR is receptor ubiquitination, which in the case of EGFR is mediated by the E3-ubiquitin ligase Cbl (Levkowitz et al., 1999). Receptor fate is determined by the levels and/or type of ubiquitination, which are recognized by Ubiquitin-interacting motifs (UIMs) present in proteins involved in receptor sorting (Raiborg and Stenmark, 2009; Huang et al., 2013). The level of ubiquitination and receptor fate seems to be dependent on ligand type and dose (Roepstorff et al., 2009; Sigismund et al., 2013). A more in depth discussion about the literature relating to ligand type- and dose-dependent regulation of endocytosis and signaling is presented in Chapter 5.

## 1.3 Comparison between ensemble and single-molecule techniques

There are two broad types of techniques that can be used to study protein interactions and PTMs (post-translational modifications): ensemble and single-

molecule techniques. In ensemble techniques, the average state of a (generally large) group of molecules is measured. Examples of this type of techniques include Western blot and ELISA (enzyme-linked immunosorbent assay), which measure the relative level of protein interaction- or PTM-state. Single-molecule techniques on the other hand are able to distinguish the state of individual molecules. Single-molecule microscopy and atomic force microscopy represent two of the most used techniques in this category.

Both ensemble and single-molecule techniques have their advantages and limitations, therefore choice of method depends on different factors such as considering the level of information that wants to be obtained and the availability of resources, to name a few. In terms of practicality and cost, ensemble techniques generally trump. Single-molecule assays generally require costly equipment and high-level of expertise. Nevertheless, they often provide information otherwise inaccessible. There are many examples of the power of single-molecule level measurements to reveal mechanistic details that could not be studied using ensemble techniques (Oh et al., 2012; Low-Nam et al., 2011; Munsky et al., 2012; Andrews et al., 2008). Recently, Jain et al. developed a technique called Single-Molecule Pull-down (SiMPull), which allows to assess the composition of individual molecular complexes (Jain et al., 2011). Later in this work, a series of modifications to this technique are described. These modifications allowed for the quantification of phosphorylation states from thousands of individual membrane receptors.

## 1.4 Computational modeling of EGFR signaling

The vast amount of experimental data available about EGFR activation and its downstream signaling events has made it a perfect system for mathematical and computational studies (Kholodenko et al., 1999; Blinov et al., 2006; Shankaran et al., 2012). In spite of this body of knowledge, many of the mechanisms involved in EGFR activation and its regulation are not completely understood. Modeling of EGFR signaling was generally performed using a series of simplifications. For example, even though it is known that different phosphorylated residues in EGFR recruit specific adaptor proteins and have different functions, these different sites were generally represented as a single one, for which multiple adaptor would compete for (Figure 1.3a). These simplifications were in most cases justified, not using them could easily result in the need of defining hundreds of ordinary differential equations, a process that would be error-prone and time-consuming. The relatively recent development of tools that allow for rule-based modeling of biochemical networks has made possible modeling of signaling systems without the need of employing the aforementioned simplifications (Faeder et al., 2009; Smith et al., 2012).

In this type of modeling, the user define rules containing only the protein components that are relevant for the reaction to happen, and the open source software does the work of creating a reaction for every possible species that is able to participate in the reaction. In the example presented in Figure 1.3, the user only needs to define 18 rules total, and the software generates ordinary differential equations (ODEs) for the 729 possible different species. In this work,

this powerful computational technique was used to include site-specific information about EGFR phosphorylation and recruitment of adaptor proteins.

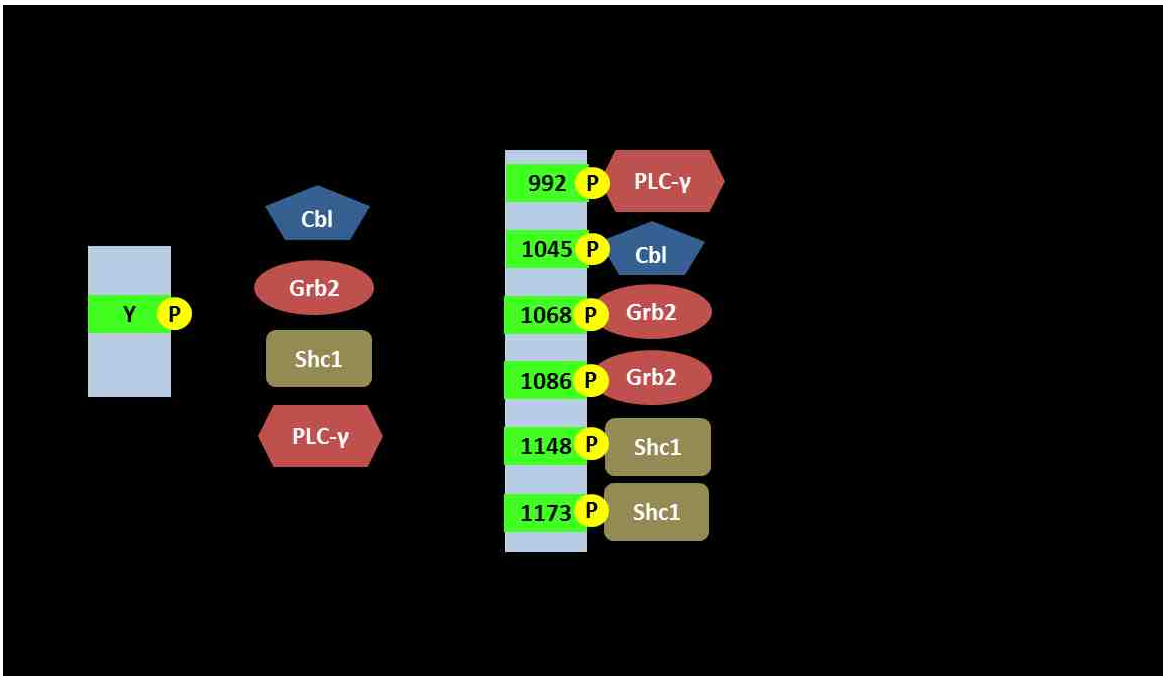


**Figure 1.3.** Comparison of traditional vs rule-based modeling approaches. **(a)** Traditional models generally lump sites together to prevent the need to define a high number of ordinary differential equations (ODEs). **(b)** Simplified representation of rule-based model considering phosphorylation, dephosphorylation and reversible adaptor protein binding to six tyrosine residues. The three possible states for each site are unphosphorylated, phosphorylated-free, or phosphorylated-occupied (by adaptor protein).

## 1.5 Hypothesis

My central hypothesis is that upon EGFR activation there are subsets of receptors having distinct phosphorylation patterns and therefore playing different

roles in signal propagation. To test this hypothesis, a series of improvements were made to the SiMPull technique, allowing the analysis of phosphorylation states from thousands of receptors. Since phosphorylation kinetics is the result of a variety of molecular processes, computational modeling was used to aid in the quest to understand how the interrelation between these events give rise to the observed behavior.

## 1.6 Summary of Results

Chapter 2 provides a detailed description of the improvements made to the SiMPull technique for quantification of receptor multisite phosphorylation. Results show that only a subpopulation of EGFR become phosphorylated under what is considered maximal activation conditions and that the extent of phosphorylation varies by tyrosine residue. Three-color imaging of EGFR-GFP with antibodies directed to two distinct phospho-sites revealed the presence of a subset of receptors with simultaneous phosphorylation at the two sites probed.

Chapter 3 describes the development of a rule-based model for the initial steps of EGFR signaling. Particularly, it considers ligand-induced receptor activation and the recruitment of the adaptor proteins Grb2 and Shc1 to phosphorylated tyrosines 1068 and 1173 in EGFR, respectively. In Chapter 4, the model was fine-tuned with experimental data obtained with the improved SiMPull technique. Experimental testing of model predictions confirmed that adaptor proteins are able to protect the sites to which they bind from dephosphorylation, and therefore modulate the phosphorylation patterns

observed *in vivo*. Chapter 5 describes an extended model developed to help explain differential signaling induced by EGFR ligands. The model provides testable predictions to help dissect the roles of dimer lifetimes and ubiquitination in this differential signaling. Finally, Chapter 6 provides a brief discussion about the implications that these results have in the understanding of signaling pathways, and some of the possible future directions.

# Chapter 2 : Single Molecule Pull-down for Quantification of Membrane Receptor Multisite Phosphorylation

Emanuel Salazar-Cavazos[1,2], Keith A Lidke[2,3], Diane S. Lidke[1,2]

[1]Department of Pathology, [2]Comprehensive Cancer Center and [3]Department of Physics & Astronomy, University of New Mexico, Albuquerque, NM 87131

## 2.1 Introduction

The ability of a cell to respond rapidly and specifically to changes in the surrounding environment is controlled by protein-protein interactions at the plasma membrane and along the signaling cascade. While much is known about the biochemical events that govern signaling pathways, this information has mostly been derived from population-based measurements that typically average over millions of cells and/or proteins. However, there is growing evidence that the heterogeneity of the system contributes to how cellular information is processed (Lahav et al., 2004; Feinerman et al., 2008; Coba et al., 2009; Spencer et al., 2009). To better understand the role of protein phosphorylation heterogeneity in directing signaling outcomes, the single molecule pull down (SiMPull) assay was adapted to identify the phosphorylation state of individual receptors.

SiMPull is a powerful technique that allows for interrogation of macromolecular complexes at the individual protein level. Jain *et al.* first demonstrated the ability of this technique to capture macromolecular complexes (Jain et al., 2011). SiMPull samples are prepared in a manner similar to IP/Western Blot protocols, but the sample is interrogated using single molecule

microscopy. Briefly, cells are lysed and the protein of interest is captured by antibodies bound to the coverglass. If the proteins are fluorescently tagged, either by fluorescent proteins or subsequent antibody labeling, their presence will be quantified by single molecule imaging (Figure 2.1A).

Here, a modification of SiMPull for the study of phosphorylation patterns of transmembrane receptors is described. Traditionally, protein phosphorylation has been measured using ensemble techniques, such as Western Blot analysis or flow cytometry, which provide information on the relative changes of a protein phosphorylation amount. However, these techniques cannot determine the fractions of proteins in a specific phosphorylation state, much less identify when an individual protein contains multiple sites of phosphorylation. While mass spectrometry has the potential to detect multisite phosphorylation, the residues of interest must be found in the same small peptide that is generated by enzymatic digestion (typically 7-35 amino acids) or the protein of interest must be small (Swaney et al., 2010; Curran et al., 2015; Brunner et al., 2015). Therefore, new techniques are needed to better understand the phosphorylation status of individual proteins. Recently, Kim *et al.* used a modified SiMPull approach, termed SiMBlot, to pull-down surface biotinylated proteins and identify phosphorylation using denaturing conditions and phosphorylation-specific antibody labeling (Kim et al., 2016). Our approach differs in several significant ways from SiMBlot and provides important improvements over previous protocols (Jain et al., 2012; Kim et al., 2016), including the reduction of autofluorescence in the green spectral channel and a simplified imaging chamber that accommodates

higher sample number with lower sample volume. Results demonstrate the importance of optimizing antibody labeling and fixation conditions. To quantify receptor phosphorylation, two- and three-color imaging were used to identify individual proteins and their corresponding phosphorylation status. Corrections to account for membrane receptor surface expression and steric hindrance in the case of dual antibody labeling are described.

This method was applied to the study of the classical Epidermal Growth Factor Receptor (EGFR). EGFR has 20 tyrosines in its cytoplasmic tail, at least 12 of which are known to recruit specific adaptor proteins (Schulze et al., 2005). The potential for multisite phosphorylation provides a mechanism through which the cell might differentially respond to extracellular cues, depending on the extent and combination of receptor phosphorylation (Gibson et al., 2000; Salazar and Höfer, 2009; Coba et al., 2009; Lau et al., 2011). Results showed that only a subpopulation of EGFR become phosphorylated under what is considered maximal activation conditions and that the extent of phosphorylation varies by tyrosine residue. Multiplex imaging of the GFP-tagged receptors and antibodies directed to two distinct phosphotyrosines revealed that multisite phosphorylation frequently occurs. The extent of phosphorylation at individual tyrosines along with the existence of multisite phosphorylation has implications for how EGFR translates extracellular cues into downstream signaling outcomes.

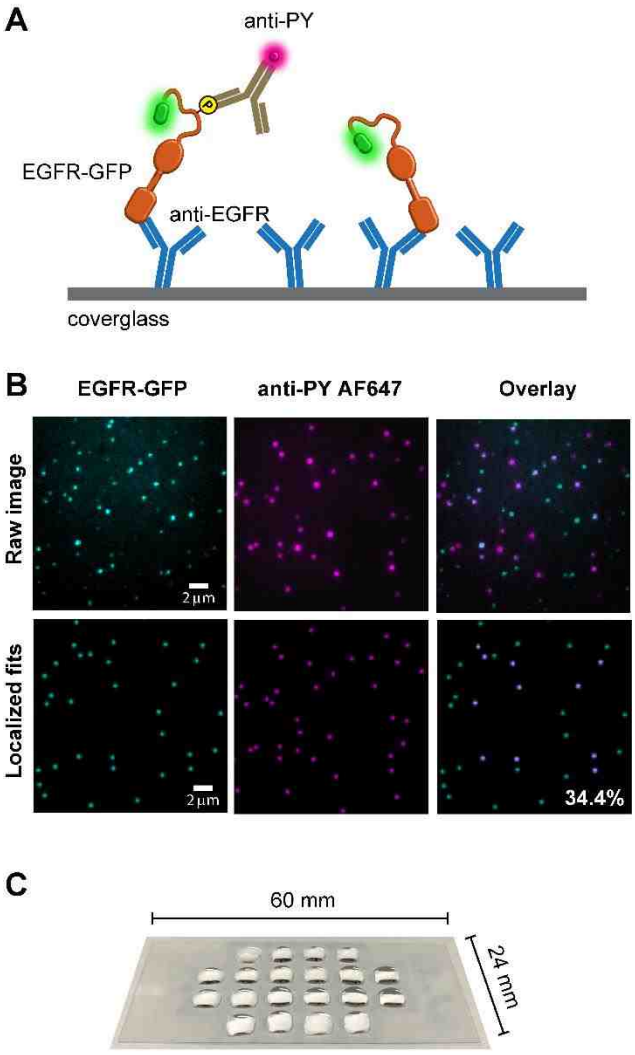


**Figure 2.1.** SiMPull to quantify protein phosphorylation. **(A)** Illustration depicting overall principle for assessing phosphorylation at the single molecule level using GFP-tagged EGFR (EGFR-GFP) as an example. **(B)** Representative images showing raw data (top) and blob-reconstructed localized molecules (bottom). CHO-EGFR-GFP cells were stimulated for 5 min with 25 nM EGF at 37°C before lysis for SiMPull. Raw images are brightness and contrast enhanced for visualization. The EGFR-GFP fits were filtered based on their fit to the

microscope point spread function and the GFP-channel used as a mask to create the overlay. The number in the bottom right image represents the phosphorylation percentage estimated for this field of view. **(C)** Hydrophobic array for preparation of SiMPull samples.

## 2.2 Results

### 2.2.1 Visualization of individual protein phosphorylation status by SiMPull

Phosphorylation of EGFR was assayed at the single molecule level using the SiMPull concept depicted in Figure 2.1A, where the GFP-tagged receptors from cell lysates are immunoprecipitated by antibodies bound to the coverglass and subsequently labeled with fluorescently-tagged antibodies detecting phosphorylated tyrosines (anti-PY). Figure 2.1B shows the capture of single EGFR-GFP from cell lysate on the coverglass surface and the corresponding labeling of phosphorylated EGFR using a pan-phosphotyrosine antibody conjugated to Alexa Fluor 647 (anti-PY AF647; PY99 antibody). Individual molecules are identified in each image (Figure 2.1B). Images are then overlaid to identify phosphorylated receptors. Colocalization of receptor and PY localizations provide an initial estimate of the fraction of phosphorylated receptors (Figure 2.1B). Optimization of the experimental process and image analysis, including reduction in autofluorescence, corrections for the level of receptor surface expression and the appropriate antibody labeling conditions, are described in the following sections.

### 2.2.2 Simplified sample chamber increases throughput and reduces sample volume

Jain *et al*. originally described the use of a fluidic chamber for imaging that consisted of 4-6 channels generated between a coverglass and microscope slide using epoxy (Jain et al., 2012). While these types of flow chambers are straightforward to produce, the protocol is time consuming (30-60 min) and larger volumes (~70 µL) are required to fill each channel. To overcome these limitations, a hydrophobic barrier pen was used to create an array of isolated sample regions on a coverglass (Figure 2.1C). Rectangular (24x60mm, #1.5) coverglasses are treated as described and an array of up to 20 squares can be drawn with the hydrophobic ink pen in a matter of minutes. As little as 10 µL of sample is needed to fill each region, which is seven times less than for the original flow channels. This is particularly useful considering the high cost of fluorescently labeled antibodies and that for some applications sample availability may be limited. Time for sample preparation is also reduced as washing and labeling steps are simplified without the need for flowing of buffers through channels.

### 2.2.3 Quenching with NaBH4 reduces background autofluorescence

It has been previously noted that autofluorescent background is detected in the spectral region corresponding to green emitting fluorophores (Jain et al., 2012). Autofluorescent puncta were also observed in the green spectral channel (503-548 nm), which were identified as single GFP molecules in the absence of

cell lysate. Since the experimental approach used relies on GFP to identify the location of EGFR on the coverglass, it was important to reduce this background to avoid over-counting of receptors. Incubating the PEG-coated coverglass with Sodium Borohydride (10 mg/mL $NaBH_4$ for 4 min) was found to significantly reduce the number of background fluorescent molecules (Figure 2.2). Despite the improvement in background signal, background measurements were routinely acquired for each coverglass preparation to enable background correction for each experiment (see Methods for details).

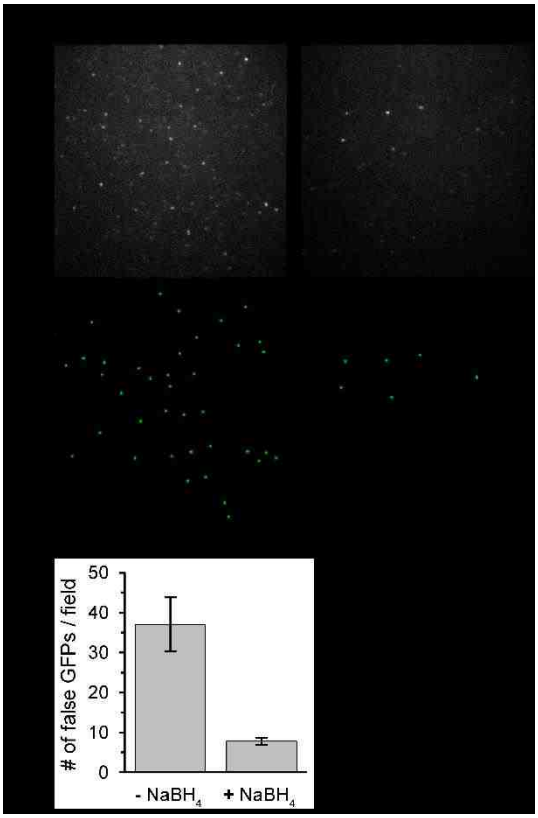


**Figure 2.2.** Reduction of autofluorescence with Sodium Borohydride (NaBH4). **(A)** Raw images and blob-reconstructions from a typical field of view of a PEG/PEG-biotin functionalized surface without (left) and with (right) $NaBH_4$-treatment. **(B)** Quantification of the average number of false-positive localizations

per field of view in surfaces with or without treatment with $NaBH_4$. For each condition N > 12 fields of view were analyzed. Error bars represent mean +/- S.E.M.

### 2.2.4 Antibody optimization is required for accurate phosphotyrosine detection

Since antibodies are used to quantify protein phosphorylation, it is critical to optimize the antibody labeling protocol. Figure 2.3 presents the results from optimization of anti-EGFR-pY1068. In these experiments, EGF stimulated cells were co-treated with a phosphatase inhibitor (pervanadate, PV) to increase the amount of receptor phosphorylation. Incubation of EGFR-pY1068 on ice for 60 minutes was needed to ensure maximal labeling (Figure 2.3A). Importantly, results show that over time the antibody dissociates from EGFR, with ~37% reduction after 1 hr at room temperature (Figure 2.3B, no fixative). Complete imaging of the sample array can take up to 1 hr, therefore, loss of antibody over this period would lead to an underestimate of receptor phosphorylation for samples imaged later in time. Multiple fixation protocols to minimize antibody unbinding were tested. Results show that fixation with 4% Paraformaldehyde/0.1% Glutaraldehyde (PFA/GA) for 10 min stabilized the antibody levels for at least 1 hr (Figure 2.3B, PFA/GA).

To ensure that saturating levels of antibody are used, concentration curves for each antibody were generated, using the PFA/GA fixation for optimal results. The example in Figure 2.3C shows the titration curve for anti-EGFR-

pY1068, which saturates at ~20 μg/mL. Consequently, 20 μg/mL were used for all experiments. Binding affinity will vary for each antibody and fluorescent-conjugation may also alter antibody affinity. Therefore, it is necessary to perform a binding curve for each antibody and for each new antibody conjugation. Another important consideration is the specificity of the antibody for its binding site. Kim et al demonstrated an elegant way to determine specificity by using purified proteins with individual tyrosines mutated to alanine. The same EGFR-pY1068 and EGFR-pY1173 antibodies that they found to have high specificity from their *in vitro* measurements are used in this work (Kim et al., 2016).

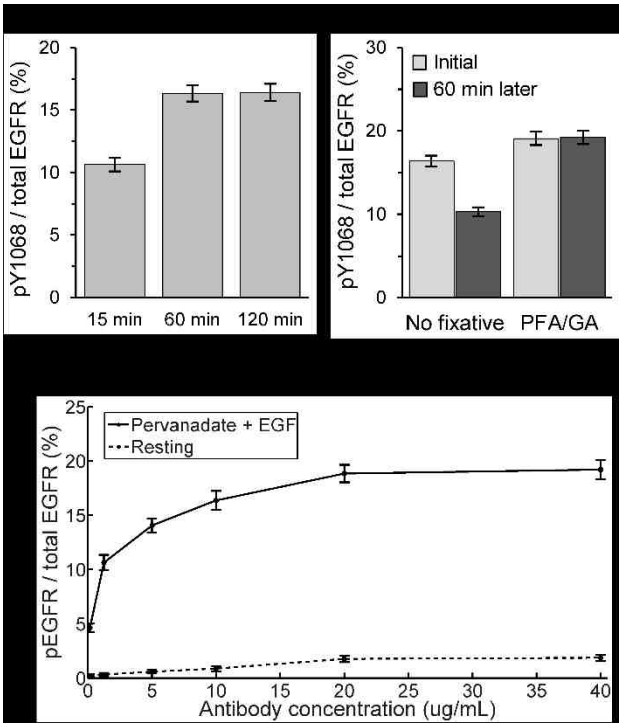


**Figure 2.3. Optimization of antibody labeling for accurate quantification of receptor phosphorylation**. CHO-EGFR-GFP cells were pre-treated with 1 mM PV for 15 min and stimulated with 50 nM EGF+1mM PV for 5 min at 37°C to enhance receptor phosphorylation and interrogated for anti-EGFR-pY1068-

CF555 labeling. **(A)** Antibody labeling with anti-pY1068 requires 60 min to reach maximal labeling. A 20 μg/mL antibody concentration was used. Number of receptors analyzed per condition, N>3400. **(B)** Addition of PFA/GA post-fixation prevents loss of antibody over time. N>2700 per condition. **(C)** Increase in labeling as a function of antibody dose. EGFR-pY1068-CF555 saturates at ~20 μg/mL. Antibody was incubated for 1 hour on ice and post-fixed with PFA/GA. Resting cells were used as a control for non-specific labeling. N>1700 per data point. Error bars are standard error of measured phosphorylation percentages.

### 2.2.5 Correction is required to account for non-surface localized receptors

At any point in time, a fraction of membrane receptors is trafficking through intracellular compartments. These internal receptors are not accessible during addition of extracellular ligand, but will be captured by the antibody during SiMPull sample preparation and result in an underestimate of receptor phosphorylation. In CHO-EGFR-GFP cells, a fraction of the receptors are located in intracellular compartments (Figure 2.4A, left). To determine the fraction of EGFR accessible to ligand, all surface proteins on the CHO-EGFR-GFP cells were labeled with membrane-impermeable AF647-NHS Ester (Figure 2.4A, right) and used SiMPull to visualize the amount of EGFR-GFP colocalized with AF647. By increasing the concentration of AF647-NHS until saturation is achieved, it was estimated that ~65% of the receptors are located at the plasma membrane (Figure 2.4B). With this information, measurements were corrected to account for only those receptors available to bind ligand. After correction, ~14% of the

receptors are phosphorylated at Y1068 after 1 min stimulation with 50 nM EGF (Figure 2.4C). While surface labeling of receptors with AF647-NHS ester allows for identification of surface proteins, this modification reduced EGF binding (data not shown). Therefore, pre-labeling of receptors was not used for the study of EGFR activation.

To validate the correction method, the phosphorylation levels of receptors were measured in CHO cells expressing ACP-tagged EGFR. EGFR localized at the plasma membrane was directly labeled using membrane-impermeable CoA-Atto488 as describe previously (Valley et al., 2015; Ziomkiewicz et al., 2013). Cells were then exposed to EGF and probed for EGFR phosphorylation with SiMPull, this time using Atto488 as the marker for plasma membrane EGFR. The percent of phosphorylated EGFR was similar when comparing the membrane-localized ACP-EGFR and the membrane-corrected EGFR-GFP samples (Figure 2.4C). Therefore, the effects of EGF binding to EGFR on the plasma membrane can be accurately determined from whole cell lysates using the correction, which was applied for the remainder of the results.

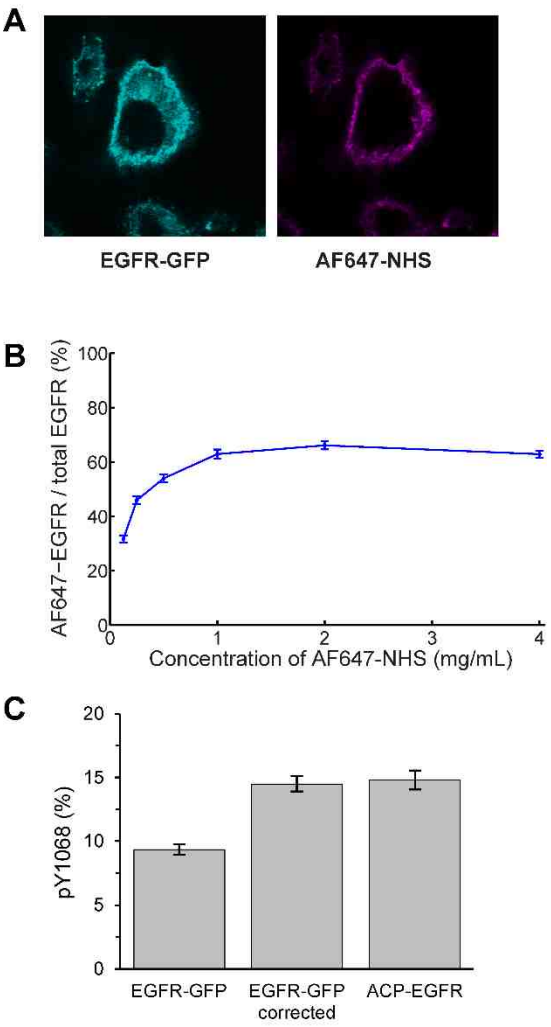


**Figure 2.4.** Correction for cellular distribution of receptors. **(A)** Confocal images showing typical distribution of EGFR-GFP in CHO cells (left) and the labeling of surface proteins achieved with the AF647-NHS ester (right). **(B)** Cells were incubated with increasing concentrations of AF647-NHS and assayed by SiMPull to determine the percentage of EGFR-GFP molecules labeled with AF647. Number of receptors analyzed per data point, 850 < N < 1550. **(C)** Percentage of pY1068+ receptors estimated for EGFR-GFP before and after correcting for surface expression. The corrected phosphorylation percentage for EGFR-GFP corresponds to the value measured for ACP-EGFR, which only includes plasma

membrane localized receptors. N>2400 for each EGFR type. Error bars are standard error of measured phosphorylation percentages.

### 2.2.6 Extent of phosphorylation varies by tyrosine residue

SiMPull was used to characterize the kinetics and dose response of EGFR activation. The multi-well hydrophobic array (Figure 2.1C) made it possible to efficiently examine a full dose response or time course of activation in a single imaging session. Total EGFR tyrosine phosphorylation (PY) and the phosphorylation patterns for two specific tyrosine sites (Y1068 and Y1173) were quantified. Cells simulated for 5 min (Figure 2.5A) with increasing concentrations of EGF showed the expected increase in total phosphorylation with ligand dose (Figure 2.5A, PY, blue bars). This fraction reached 64% with 50 nM EGF, a dose that is considered saturating. While both specific tyrosines show less phosphorylation than total PY, the fraction of EGFR with phosphorylation at Y1173 was consistently higher than at Y1068 (Figure 2.5A). The kinetics of phosphorylation between PY, pY1068 and pY1173 are similar (Figure 2.5B). These results indicate several important outcomes. First, phosphorylation detection by SiMPull is sensitive, capable of detecting receptor phosphorylation at low ligand dose and early time points. Second, even under saturating ligand conditions, only a fraction of receptors is phosphorylated, reaching a maximum of 64% with 5 min stimulation. Third, the extent of phosphorylation varies by tyrosine residue. The detected phosphorylation levels are not due to limitations in antibody labeling, since cells stimulated in the presence of phosphatase

inhibitors showed increased receptor phosphorylation (Figure 2.6A).The use of high salt (500 mM NaCl) concentration during cell lysis did not change the detected phosphorylation, indicating that adaptor proteins are not interfering with antibody recognition (Figure 2.6B).

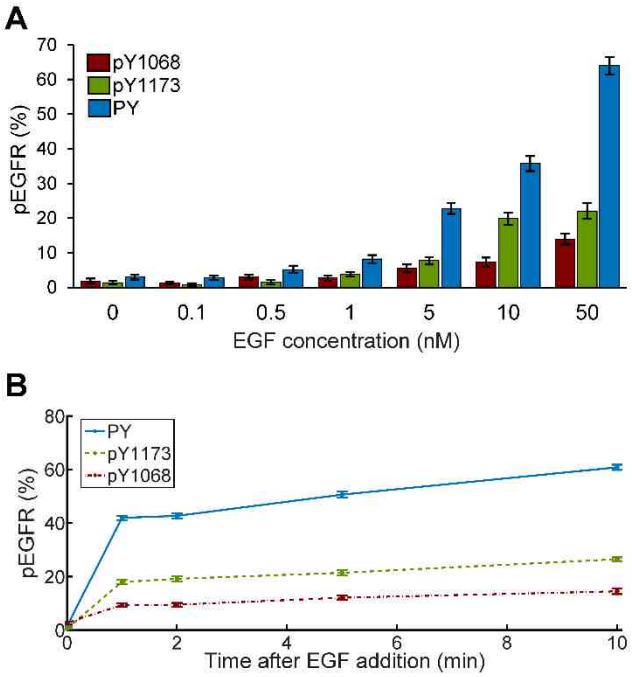


**Figure 2.5.** The extent of phosphorylation varies by tyrosine residue. **(A)** Dose response curve for CHO-EGFR-GFP cells after 5 min of EGF addition at 37°C. Number of receptors analyzed per condition, 800 < N < 1800. **(B)** Site-specific EGFR phosphorylation kinetics. Phosphorylation time course for CHO-EGFR-GFP cells stimulated with 25 nM EGF at 37°C. N>1800. Error bars are standard error of measured phosphorylation percentages.

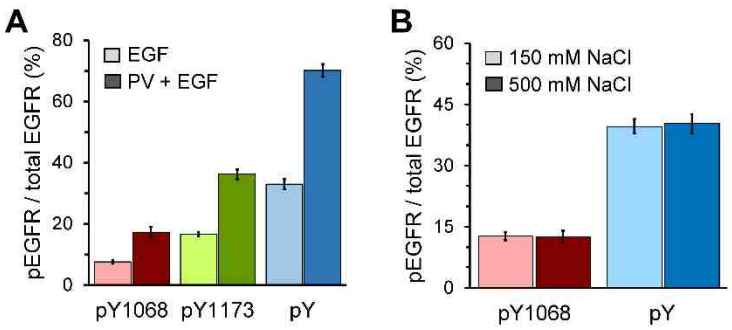


**Figure 2.6.** Effect of phosphatase inhibition or cell lysate salt concentration on detected phosphorylation levels. **(A)** CHO-EGFR-GFP cells were stimulated at 37°C with either 50 nM EGF for 5 min or pre-treated with 1 mM pervanadate (PV) for 15 min and then stimulated with 50 nM EGF and 1mM PV (PV + EGF) for 5 min. Considering that pervanadate treatment induces EGFR phosphorylation that may not be restricted to the plasma membrane, no surface correction was applied for this figure. Number of receptors per condition, 690 < N < 3400. **(B)** CHO-EGFR-GFP cells were stimulated at 37°C with 25 nM EGF for 1 min and protein extraction was performed with either regular lysis buffer containing 150 mM NaCl (see Methods) or 500 mM NaCl. High NaCl concentrations have been shown to promote disruption of interactions between SH2-containing proteins and their phosphorylated binding partner sites (Grucza, R. A., et al., *Biochemistry*, 39(33), 10072-10081). 670 < N < 1600. Error bars are standard error of measured phosphorylation percentages.

### 2.2.7 Three-color SiMPull reveals multisite phosphorylation

The observation that Y1068 and Y1173 have different phosphorylation levels suggests that there are subpopulations of receptors with differing

phosphorylation patterns. However, examining a single tyrosine site at a time cannot address the coincidence of phosphotyrosines. To assess the potential of multisite protein phosphorylation, simultaneous three-color SiMPull imaging was developed. To test whether receptors phosphorylated at Y1068 were also phosphorylated at other tyrosine residues, receptors were co-labeled with anti-pY1068 and anti-PY antibodies. When labeling a single protein with two or more antibodies, the effects of steric hindrance must be considered. In this case, labeling first with anti-pY1068 followed by anti-PY did not alter PY levels (Figure 2.7A). However, labeling with anti-PY first did cause a loss of pY1068 signal (Figure 2.7B). Therefore, the experiments were performed with sequential labeling, anti-pY1068 followed by anti-PY. As before, the addition of EGF resulted in increased phosphorylation, with the PY antibody showing more labeling than the site-specific antibody (Figure 2.8A,B) and the presence of multi-phosphorylated receptors was observed in the three-color images (Figure 2.8A, white circles). Figure 2.8B shows quantification of the three-color colocalization, which revealed that ~12% of EGFR were labeled by both antibodies (pY1068+PY, orange bar). Importantly, nearly 76% of the receptors phosphorylated at Y1068 were co-labeled with PY. Therefore, multisite phosphorylation is a prevalent outcome in EGFR activation.

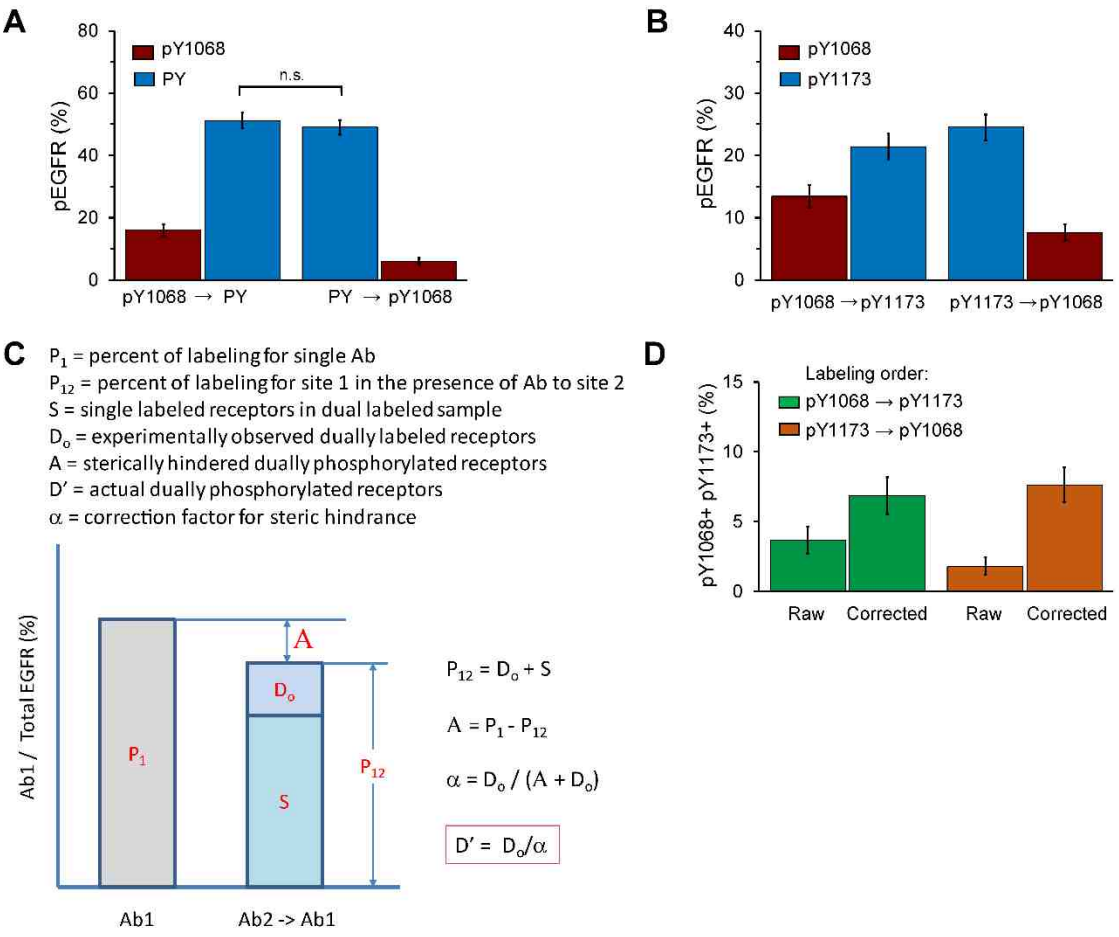


**Figure 2.7.** Assessment and correction of steric hindrance in sequentially incubated antibodies for 3-color SiMPull. **(A)** Evaluation of steric hindrance between anti-pY1068-CF555 and anti-PY-AF647 (PY) antibodies. CHO-EGFR-GFP cells were stimulated with 25 nM EGF for 5 min at 37°C and EGFR phosphorylation quantified using 3-color SiMPull. Labeling with anti-pY1068 first did not reduce subsequent labeling by anti-PY. However, a reduction in pY1068+ receptors is seen when the labeling order is reversed. Number of receptors analyzed per measurement, N>800. n.s. not significant, P = 0.5187. **(B)** Evaluation of steric hindrance between anti-pY1068-CF555 and anti-pY1173-CF640R antibodies. Cells were stimulated as described in (A) and receptor

phosphorylation assayed by 3-color SiMPull. A reduction in labeling was observed for the antibody that is applied second in the labeling sequence. N>780 per measurement. **(C)** Diagram describing estimation of correction factor ($\alpha$) to calculate actual fraction of receptors with dual phosphorylation ($D'$). The observed reduction in labeling with Antibody 1 alone (left bar) as compared to Antibody 1 following Antibody 2 (right bar) indicates the level of steric hindrance. From this information, the correction factor can be calculated. **(D)** Validation of the correction factor by exchanged labeling order. After applying the correction factor ("Corrected" bars), the percentage of pY1068+pY1173+ receptors is similar. Error bars are standard error of measured phosphorylation percentages.

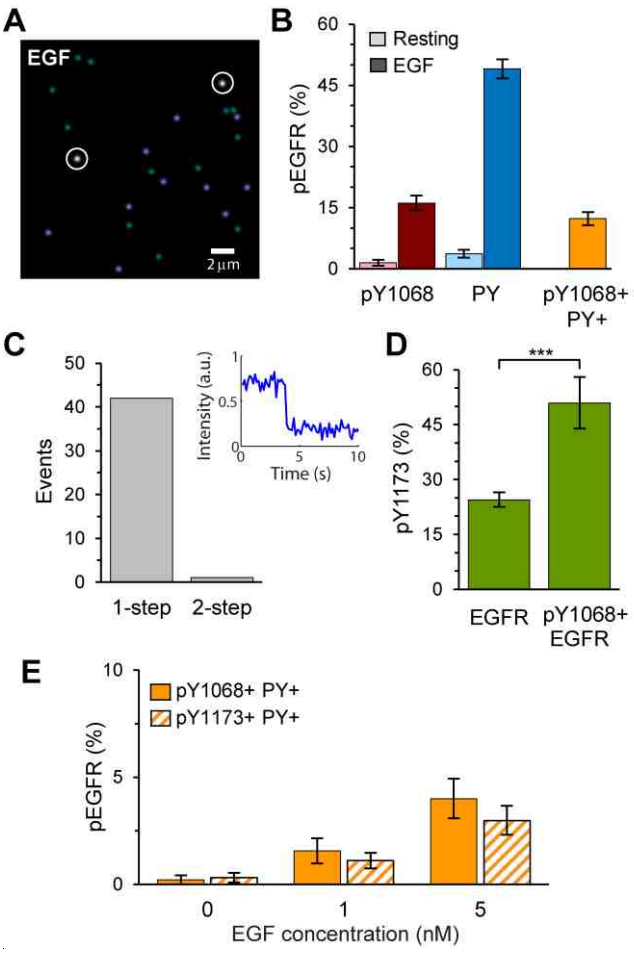


**Figure 2.8.** SiMPull reveals EGFR multisite phosphorylation. **(A)** Representative 3-color SiMPull image showing detection of EGFR-GFP (cyan), where receptors positive for PY labeling appear purple and white circles mark receptors labeled for both PY and pY1068. This image does not contain receptors labeled with pY1068 alone. Cells were treated with 25 nM EGF for 5 min. **(B)** Quantification of single and multi-phosphorylation in EGFR. Number of receptors analyzed per condition, N>500 for resting condition and N>840 for EGF condition. **(C)** Step-photobleaching analysis of multi-phosphorylated EGFR-GFP from (B). The majority (98%) of diffraction limited GFP spots show single-step bleaching, consistent with the pull-down of receptors as monomers. Inset shows example

GFP-intensity trace of a multi-phosphorylated EGFR-GFP. **(D)** Percentage of Y1173 phosphorylation in overall population of surface receptors compared to that in pY1068+ receptors. N>780 for EGFR and N=51 for pY1068 EGFR. *** P < 0.001. **(E)** Multisite phosphorylation is also observed at lower EGF dose. Cells stimulated for 5 min with indicated EGF dose. 970 < N < 1700 per condition. Error bars are standard error of measured phosphorylation percentages.

To ensure receptors detected as multi-phosphorylated were individual receptors labeled with both antibodies rather than two nearby labeled receptors detected as one in a diffraction-limited spot, step-photobleaching analysis was performed (Figure 2.8C). Analysis showed that the majority of doubly labeled receptors (~98%) were associated with a single EGFR-GFP molecule (Figure 2.8C, right). It is important to note that the number of GFP spots demonstrating two-step photobleaching increased as the sample density increased (data not shown). Therefore, a pulldown protein density in the range of 0.04-0.08/$\mu m^2$ is recommended. Alternatively, photobleaching traces can be performed in each measurement to exclude those spots showing more than one-step photobleaching.

With the knowledge that the majority of pY1068+ receptors are also phosphorylated in at least one other tyrosine residue, the pairwise phosphorylation of Y1068 and Y1173 was examined. In contrast to dual labeling with pY1068 and PY (pan-pTyr) antibodies, the close proximity of these two tyrosines did result in steric hindrance of antibody binding. This is seen as a reduction in labeling when the antibodies are applied second as compared to first

(Figure 2.7B). The reduced labeling efficiency measured in sequential labeling was used to correct for steric blocking (see Figure 2.7C,D and Methods). Results show that approximately 50% of the pY1068+ receptors are co-phosphorylated at Y1173. This is an enrichment of approximately two-fold as compared to pY1173+ in the total EGFR population (Figure 2.8D). Stimulation of cells with lower doses of EGF also resulted in multisite EGFR phosphorylation (Figure 2.8E). Notably, at 1 nM EGF, multi-phosphorylation is already considerable within the site-specific subpopulations, with 58 +/- 14% of pY1068+ receptors and 29 +/- 8% of pY1173+ receptors being co-labeled with PY. Therefore, multisite phosphorylation is not merely a consequence of saturating ligand conditions. The use of a three-color imaging scheme to correlate phospho-antibody labeling directly with GFP-tagged receptors was critical, due to the relatively high non-specific binding of the antibodies (Figure 2.9A,B). In the absence of the GFP channel to remove the non-specific binding, the values for dual labeling are underestimated (Figure 2.9C,D)**.** These results show that SiMPull, when performed using the improvements described here, can be used to quantify the extent and coincidence of phosphorylation at multiple tyrosines.

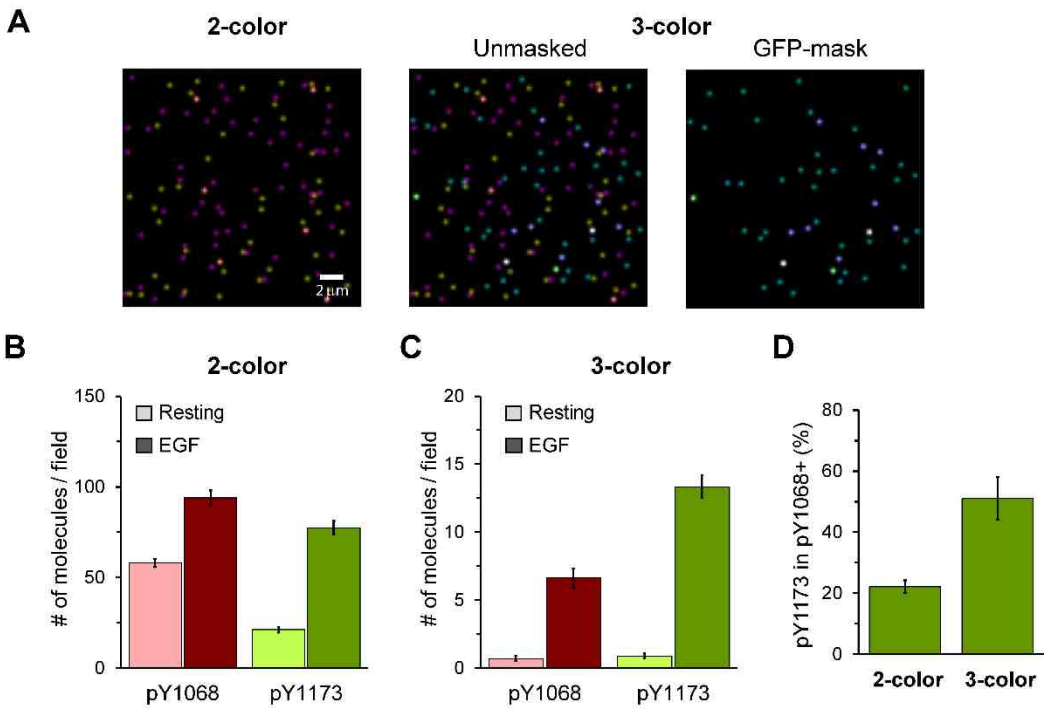


**Figure 2.9.** Importance of multi-color imaging for accurate quantification of phosphorylation percentages. **(A)** Representative images displaying raw data and blob-reconstructed localized molecules from a 3-color SiMPull experiment. CHO-EGFR-GFP cells were stimulated with 25 nM EGF for 5 min at 37°C and assayed using anti-pY1068-CF555 (yellow) and anti-pY1173-CF640R (pink) antibodies. **(B)** Quantification of total number of pY1068 and pY1173 localizations per field of view when only those two channels are examined. EGFR-GFP channel was ignored for this quantification to emulate a 2-color SiMPull experiment. **(C)** Quantification of total number of pY1068 and pY1173 localizations per field of view using 3-color SiMPull. Here, the EGFR-GFP channel was used to identify pY1068 and pY1173 localizations overlapping with EGFR molecules, removing contributions from non-specific antibody binding. **(D)** In the absence of the EGFR-GFP channel to identify receptor locations, the 2-

color SiMPull underestimates protein multi-phosphorylation. Number of receptors per condition, N>2400. Error bars are standard error of measured phosphorylation percentages.

## 2.3 Discussion

In this Chapter, a series of modifications to SiMPull that allow this technique to obtain quantitative information about multiple post-translational modifications (PTMs) at the single-protein level were described. SiMPull was used to monitor EGFR phosphorylation patterns, quantify subpopulations of phosphorylated receptors, and directly observe the existence multisite phosphorylation. This approach holds distinct advantages over other techniques. Detailed information on protein PTMs is not accessible by traditional biochemical methods that can only determine relative changes from an average of the population. While mass spectrometry has the potential to detect multisite phosphorylation, the residues of interest must be found in close proximity (Swaney et al., 2010). In SiMBlot, which is also a single molecule approach to detecting PTMs, surface proteins are first biotinylated and then pulled-down via streptavidin-coated surface, rather than by a specific antibody (Kim et al., 2016). The SiMPull method is not restricted to surface proteins and is therefore compatible with the interrogation of intracellular proteins. PTMs other than phosphorylation can also be studied as long as a suitable antibody is available. Thus, detection of PTMs by SiMPull enables measurements that were previously difficult to perform.

In addition to its advantages, SiMPull has a number of caveats that must be considered to ensure rigorous quantification. As with any antibody-dependent technique, the affinity and specificity of the antibody must be determined. Results showed that it is important to establish proper concentrations and labeling times for each antibody used, as well as the importance of post-fixation to prevent antibody dissociation during imaging. In addition, directly labeling the primary antibody with the fluorophore eliminates the need for secondary antibodies, which may add additional labeling efficiency artifacts and restrict options due to the limited availability of species used to generate primary antibodies. The phosphotyrosines probed in EGFR are located in an intrinsically disordered region of the C-terminal tail, therefore these sites are likely to be more accessible to antibodies than if they were located in structured regions. If the PTMs of interest are found in structured regions, a protein denaturation step may be used (Kim et al., 2016). Steric hindrance of two or more antibodies binding to the same protein is another possible complication. Steric hindrance was found in the case of co-labeling pY1068 and pY1173. However, a simple control experiment and mathematical correction are described to avoid undercounting of dual-phosphorylated receptors. For future studies, the use of fluorescently-conjugated Fab fragments may reduce the impact of steric hindrance. It is worth noting that while detailed information on the phosphorylation status of individual proteins status is obtained, the cell-to-cell variability is lost with SiMPull.

Using SiMPull, EGFR phosphorylation patterns were quantified. The new level of detail afforded by SiMPull has provided several important insights. First,

only a fraction of EGFR was found to be phosphorylated, even under saturating ligand conditions. Second, the phospho-EGFR is further divided into subpopulations that vary in the extent of phosphorylation at individual tyrosine residues. Third, the use of three-color imaging allowed to probe for multisite phosphorylation. Comparisons of pY1068 with a pan-phosphotyrosine antibody revealed that many receptors are indeed phosphorylated at more than one tyrosine simultaneously. Strikingly, the majority of pY1068+ receptors are co-labeled with PY antibody and ~50% of pY1068+ are also positive for pY1173. These results are in contrast to recent SiMBlot studies of EGFR concluding that multisite phosphorylation was not a common occurrence. These differences may be explained by optimization of our labeling protocol that provided the sensitivity needed to detect multisite phosphorylation, including the use of fluorescently-conjugated primary antibodies, labeling under saturating antibody conditions and post-fixation to prevent antibody dissociation.

Notably, our results are consistent with previous work indicating that multisite phosphorylation is important in the efficient recruitment of certain adaptor proteins to activated EGFR (Sigismund et al., 2013; Fortian and Sorkin, 2014). The existence of multisite phosphorylation holds significant functional implications. By modulating protein phosphorylation patterns, both single- and multisite combinations, downstream signaling pathways may be differentially activated and lead to biased signaling. Consistent with this idea, it has been shown that biased signaling arises with different ligand types and doses, as well

as with the relative abundance of receptors and their signaling partners (Chen et al., 2009; Wilson et al., 2012; Freed et al., 2017; Wolf-Yadlin et al., 2006).

Interestingly, dual Y1068/Y1173 phosphorylation occurred about two-times more frequently than expected if these sites were independent of each other, suggesting positive correlation between the sites. Mechanistically, this enrichment could be a result of either long-lived receptor interactions or repeated dimerization events. If dimer lifetimes are sufficiently long, then phosphorylation of multiple sites could happen in a single dimerization event, suggesting that phosphorylation occurs in a semi-processive manner. Alternatively, if a receptor undergoes many dimerization and dissociation events, then these repeated interactions could result in the phosphorylation of a unique tyrosine in each encounter. This would be similar to quasi-processive phosphorylation as described for ERK (Aoki et al., 2011). Experimental evidence exists to support each of these mechanisms. Both long-lived and transient EGFR dimerization has been observed on living cells, with dimer lifetimes dependent on ligand occupancy (Low-Nam et al., 2011). Recent work from the Lemmon group has shown that high and low affinity ligands induce distinct dimer structures, where low affinity ligands lead to less stable dimers and differential signaling outcomes (Freed et al., 2017). Oncogenic signaling by EGFR mutants has been shown to be driven by enhanced dimerization and increased catalytic activity that could amplify multi-phosphorylation (Kim et al., 2012; Shan et al., 2012; Valley et al., 2015; Zhang et al., 2006).

Interplay between receptors and the membrane environment has also been shown to affect the efficiency of EGFR encounters (Low-Nam et al., 2011; Chung et al., 2010). Therefore, the frequency of dimerization and the duration of dimer lifetimes may serve as a kinetic proofreading mechanism, regulating the EGFR phosphorylation patterns and dictating cellular outcome. Additionally, adaptor protein binding and phosphatase activity likely play roles in phosphorylation extent. For example, Capuani et al have shown that Grb2 and Cbl can protect Y1045/Y1068 from dephosphorylation (Sigismund et al., 2013). These mechanisms are not mutually exclusive and may be more or less relevant depending on the cellular contexts. Therefore, it would be expected that phospho-EGFR patterns will be modulated by differences in ligand dose, ligand-dependent dimer lifetimes, membrane architecture and adaptor protein abundance. The contributions that these mechanisms have in EGFR and other signaling pathways, remain unclear. The unique datasets provided by SiMPull, combined with other experimental and computational modeling tools, should prove useful in addressing these types of questions.

**Acknowledgements**

This work was supported by the New Mexico Spatiotemporal Modeling Center (NIH P50GM085273), Mexican Secretariat of Public Education SEP (E.S.-C.), National Science Foundation CAREER MCB-0845062 (D.S.L.), and the UNM Comprehensive Cancer Center (NIH P30CA118100). We thank Dr. Bridget Wilson and Dr. William Hlavacek for useful discussions, and Shayna

Lucero for assistance with cell culture. We thank Dr. Donna Arndt-Jovin for the CHO-ACP-EGFR cell line and suggesting the ACP-EGFR comparison experiment. We gratefully acknowledge use of the University of New Mexico Comprehensive Cancer Center fluorescence microscopy and flow cytometry facilities, as well as the NIH P30CA118100 support for these cores. The authors declare no competing financial interests.

**Author contributions**

E.S.-C. performed all experiments and data analysis; D.S.L and E.S.-C. designed experiments and wrote the manuscript; E.S.-C. and K.A.L. developed analysis methods and algorithms. All authors discussed the results and commented on the manuscript.

**Competing financial interests**

The authors declare no competing financial interests.

## 2.4 Methods

### 2.4.1 Cell lines and reagents

CHO cells expressing GFP-tagged (Brock et al., 1999; Lidke et al., 2004) or ACP-tagged EGFR (provided by Dr. Donna Arndt-Jovin) were cultured in DMEM supplemented with 10% FBS, penicillin–streptomycin and 2 mM L-

glutamine (Thermo Fisher Scientific). ACP-tagged EGFR was as described in (Valley et al., 2015; Ziomkiewicz et al., 2013) with the exception that a shortened 16 aa sequence was introduced at the EGFR N-terminus (George, 2006). EGF, Protease and Phosphatase Inhibitor Cocktail, Alexa Fluor 647 NHS Ester, and NeutrAvidin were purchased from Thermo Fisher Scientific. CoA 488 and ACP Synthase were purchased from New England Biolabs. N-(2-aminoethyl)-3-aminopropyltrimethoxysilane was purchased from United Chemical Technologies (#A0700). Sodium bicarbonate and sodium borohydride were purchased from EMD Millipore (#SX0320-1, #SX0380-3). mPEG-Succinimidyl Valerate (MPEG-SVA-5000-5g) and biotin-PEG-Succinimidyl Valerate (Biotin-PEG-SVA-5000-500mg) were from Laysan Bio. Biotinylated anti-EGFR antibody (E101) was obtained from Leinco Technologies. Antibodies in carrier-free buffer were purchased from Cell Signaling Technologies: EGFR pY1068 (clone 1H12, 2236BF) and EGFR pY1173 (clone 53A5, 4407BF). Monoclonal antibody pre-labeled with AF647 to detect pan-tyrosine phosphorylation (PY99 antibody, sc-7020 AF647) was purchased from Santa Cruz Biotechnology. Mix-n-Stain CF555 and CF640R antibody labeling kits were purchased from Biotium Inc. Paraformaldehyde and glutaraldehyde were purchased from Electron Microscopy Sciences.

### 2.4.2 Labeling of antibodies

Carrier-free antibodies (50 µg at 0.5-1 mg/mL per reaction) were labeled using Mix-n-Stain antibody labeling kits following the manufacturer's instructions.

Briefly, the labeling reaction was carried out for 30 min at room temperature and antibodies were centrifuged using the ultrafiltration vial provided to remove the unconjugated dye. Antibodies were resuspended in PBS and stored at 4 °C. The labeling efficiency achieved was between 2.7-4.4 dyes/antibody.

### 2.4.3 Cell treatment and lysate preparation

CHO-EGFR-GFP cells were plated overnight in 60 mm tissue culture dishes at 800,000 cells/dish and CHO-ACP-EGFR cells in 24-well plates at 50,000 cells/well. For ACP labeling, CHO-ACP-EGFR cells were washed with serum-free DMEM medium (SFM), incubated with ACP labeling solution (SFM, 10 mM $MgCl_2$, 4 μM CoA 488 and 1 μM ACP) for 20 minutes at 37°C and washed three times with SFM previous to stimulation. Cells were washed in Tyrode's solution (135 mM NaCl, 10 mM KCl, 0.4 mM $MgCl_2$, 1 mM $CaCl_2$, 10 mM HEPES, 20 mM glucose, 0.1% BSA, pH 7.2) and treated with 25 nM EGF or Tyrode's solution alone (resting cells) at 37°C. At the indicated time points, cells were placed on ice, washed one time with cold PBS followed by addition of lysis buffer (1% IGEPAL CA-630, 150 mM NaCl, 50 mM Tris pH 7.2) containing Protease and Phosphatase Inhibitors. Cell lysates were collected using cell scrapers (Greiner Bio-One North America, #541070), transferred to fresh tubes on ice and vortexed every 5 min for a total of 20 min. Lysates were centrifuged at 16,000× g for 20 min at 4 °C and the supernatant was transferred to a new tube and stored at -80 °C. For experiments involving treatment of cells with phosphatase inhibitors, cells were pre-treated for 15 min with a Tyrode's solution

containing 1 mM pervanadate (PV) followed by incubation for 5 min in a solution with 50 nM EGF and 1 mM PV. A stock solution of 30 mM PV was prepared before each experiment by mixing equimolar concentrations of hydrogen peroxide and activated sodium orthovanadate that was incubated in the dark for at least 15 min before use.

### 2.4.4 Fabrication of hydrophobic arrays and surface functionalization

Coverglasses (24x60mm, #1.5; Electron Microscopy Sciences, #63793) were Piranha-cleaned(Labit et al., 2008) and placed in a coverglass holder (Fisher Scientific, #08-817). Coverglasses were sequentially sonicated in Methanol and Acetone for 10 min each, and in 1M KOH for 20 min using a bath sonicator (Branson Ultrasonics, B1200R-1). These solutions were stored in polypropylene 50 mL tubes (VWR, #89401-564) and reused up to five times. Coverglasses were rinsed with Milli-Q water two times, dried by quickly passing them multiple times over the flame of a Bunsen burner using metal tweezers and placed in a dry coverglass holder. A solution containing 76 mL of methanol, 4 mL of acetic acid and 0.8 mL of aminosilane (N-(2-aminoethyl)-3-aminopropyltrimethoxysilane) was prepared in an Erlenmeyer flask, immediately poured into the coverglass holder and incubated at room temperature for 10 min in the dark, followed by 2 min sonication and another 10 min incubation in the dark. Coverglasses were next washed with methanol for 2 min, rinsed and washed for 2 min with water, and dried in the dark.

Treated coverglasses were placed on top of a parafilm-covered coverglass containing a guide pattern, which was used as reference to draw the Sample Array with a hydrophobic barrier pen (Vector Laboratories, #H-4000). Ink was allowed to dry for at least 5 min before coverglasses were placed in a humidified chamber (empty tip rack with 50 mL of water; USA Scientific #1111-2820). For surface functionalization, 50 mg of mPEG-Succinimidyl Valerate, 1.3 mg of biotin-PEG-Succinimidyl Valerate and 200 µL of freshly prepared 10 mM sodium bicarbonate were mixed thoroughly by pipetting up and down, centrifuged for 1 min at 10,000 g at room temperature and immediately applied to the SiMPull array (10-13 uL per region). After incubating for 3-4 hours in the dark inside the humidified boxes, arrays were washed by sequential 30 sec submersions into three water-filled 250 mL glass beakers. Coverglasses were dried with nitrogen gas, stored in pairs (back to back) inside 50 mL tubes, which were filled with nitrogen gas before closing and sealing with Parafilm. Coverglasses were stored in the dark at -20°C for up to a week before use.

### 2.4.5 Labeling and quantification of surface receptors

CHO-EGFR-GFP cells grown in 24-well plates were placed on ice and washed 3 times with cold PBS. AF647-NHS Ester was dissolved at the indicated concentrations in PBS. Cells were incubated with this solution for 30 min at 4°C with gentle agitation, washed 3 times with cold PBS and subjected to cell lysis. The percent of receptors labeled with AF647 across different dye concentrations was assessed with SiMPull. To estimate the percent of receptors at the cell

surface the AF647-labeling curve was fitted to a biexponential decay curve in its increasing form using the 'fit' function in MATLAB: $y = C1\ (1 - e^{-ax}) + C2\ (1 - e^{-bx})$, where *y* is the % of AF647-labeled receptors, *x* is the concentration of reactive AF647-NHS ester used, and $a>0$, $b>0$, *C1* and *C2* are coefficients to be fitted. The sum of the coefficients *C1* and *C2* represent the asymptote of the curve and an approximation of the fraction of receptors at the cell surface.

### 2.4.6 Single-Molecule Pulldown and phospho-site labeling

T50 (10 mM Tris pH 8.0, 50 mM NaCl) and T50-BSA (T50 with 0.1 mg/mL BSA) solutions were prepared and stored for up to a month at room temperature. SiMPull arrays were equilibrated at room temperature and placed on a TC100 plate lined with Parafilm. Each region of the SiMPull array was treated with 10-15 µL of a 10 mg/mL sodium borohydride ($NaBH_4$)/PBS solution for 4 min at room temperature and washed 3 times with PBS. SiMPull regions were then incubated with a 0.2 mg/mL NeutrAvidin/T50 solution for 5 min and washed three times with T50, followed by incubation with a 2 µg/mL biotinylated anti-EGFR/T50-BSA solution for 10 min and washed three times with T50-BSA.

The plate containing the SiMPull array(s) was kept on ice during sample preparation. Lysates were diluted in cold T50-BSA with Protease and Phosphatase Inhibitors (T50-BSA/PPI), vortexed at medium speed, and added to the SiMPull array. After 10 min incubation, the lysates were removed and the SiMPull regions washed 4 times with cold T50-BSA/PPI. To determine appropriate dilution factor, the density of pulldown receptors as a function of

lysate concentration was first assessed to achieve a pulldown density 0.04-0.08/μm$^2$. Antibodies were diluted in cold T50-BSA/PPI, incubated for 1 hr, washed 6 times with cold T50-BSA for a total of 6-8 minutes, and washed twice with cold PBS. Immediately after, antibodies were fixed for 10 min with a 4% PFA/0.1% GA solution (paraformaldehyde/glutaraldehyde) and washed 2 times with 10 mM Tris (pH 7.4)/PBS for a total of 10 min to inactivate fixatives. For 3-color SiMPull experiments the same antibody incubation and fixation procedure was performed for the second antibody. Tris solution was replaced by T50-BSA and the SiMPull array was equilibrated to room temperature before proceeding to imaging.

### 2.4.7 SiMPull imaging

Imaging of SiMPull samples was performed using an inverted microscope (Olympus America, model IX71) equipped with a 150×/1.45 NA oil-immersion objective for Total Internal Reflection Fluorescence Microscopy (Olympus America, UAPON 150XOTIRF) and a three-dimensional piezostage (Mad City Labs, Nano-LPS100). Excitation of CF640R- or AF647-labeled antibodies was done using a 642-nm laser (Thorlabs, HL63133DG), CF555-labeled antibodies using a 561-nm laser (Coherent Inc, Sapphire 561-100 CW CDRH), and of GFP- and CoA 488-tagged receptors using a 488-nm laser (Spectra Physics, Cyan 100mW). All lasers were set in total internal reflection configuration, and laser powers were adjusted to prevent photobleaching of the sample at the timescale of the image exposure time (300 msec). Sample illumination and emission were

filtered using a quad-band dichroic and emission filter set (Semrock, LF405/488/561/635-A-000). Emission light was separated into four channels using a quad-view multichannel imaging system (Photometrics, model QV2) equipped with the appropriate dichroics (Chroma, 495 DCLP, 565 DCLP, 660 DCLP) and emission filters (Semrock, 685/40 nm, 600/37 nm, 525/45 nm). Emission light was collected with an electron-multiplying charge-coupled device (EMCCD) camera (Andor Technology, DU-897E-C50-#BV) with EM gain set to 200.

Each channel was 256 x 256 pixels, with a pixel size of 106.7 nm. Photobleaching and bleed through were prevented by controlling the laser shutters and microscope stage through a MATLAB script to sequentially excite and acquire the different fluorophores (642-nm laser first, 488-nm laser last). A minimum of 20 regions of interest were acquired per condition. For quantification of step photobleaching of EGFR-GFP molecules, a 100 frame time series (300 msec exposure time) was acquired after imaging of the other two channels.

### 2.4.8 Quantification of Receptor Phosphorylation

All image processing was performed using MATLAB together with the MATLAB toolbox for image-processing DIPImage (Delft University of Technology) (Hendriks et al., 1999) and all software is available upon request. The location of emitters in each channel was calculated using graphics processor unit (GPU) computing as previously described (Smith et al., 2010). Fits in the GFP channel were filtered based on the quality of the fit to the point spread

function to reduce the chances of detecting multiple receptors in close proximity as a single molecule. Image registration was performed as previously described (Schwartz et al., 2017). In this work, the root mean square error for image registration was <10 nm. For visualization purposes, Gaussian blob representations of the fluorophore localizations were generated. A receptor was considered to be phosphorylated when the localization centers of the receptor and labeled antibody were at a distance <106.7 nm (within 1 pixel).

Phosphorylation percentages were calculated as $100*(N_{Phos})/(N_{GFP}-N_{BG})$ where $N_{Phos}$ is the number of receptors identified as phosphorylated, $N_{GFP}$ is the number of observed single molecules in the GFP channel and $N_{BG}$ is the non-specific background rate in the GFP channel.

The number of GFP localizations was calculated by subtracting background spots and accounting only for surface receptors as follows: $N_{GFP} = (N_{LOC}-N_{BG})*SR$, where $N_{LOC}$ is the total number of emitters localized, $N_{BG}$ is the expected number of background emitters in the area imaged, and *SR* (surface ratio) is the fraction of receptors located at the cell surface. The density of background emitters was quantified for each SiMPull array and used for background correction of samples in that array. For 3-color SiMPull experiments where steric hindrance between sequentially incubated antibodies was observed (i.e. pY1068-pY1173 detection), estimations of dual phosphorylation were corrected to account for this hindrance as explained in Figure 2.7.

### 2.4.9 Statistical Analysis

Based on the consideration that the phosphorylation state of each receptor analyzed has the properties of a Bernoulli trial, standard errors (SE) of phosphorylation measurements were calculated as for sample proportions in a binomial distribution: $SE = p(1-p)/n$, where $p$ is the fraction of receptors phosphorylated and $n$ is the total number of receptors. The condition np>10 (with the exception of Figure 2.8E, np>5) and np(1-p)>10 was ensured to be met to allow this approximation to be adequate. Two-sample Z-test (two-tailed) was used to estimate p-values (LeBlanc, 2004).

### 2.4.10 Step-photobleaching Analysis

For step-photobleaching analysis of multi-phosphorylated receptors, the average fluorescence intensity of the area (200x200 nm) surrounding each of these EGFR-GFP molecules was quantified and plotted for the duration of the time series. Intensity plots were manually analyzed and the number of photobleaching steps was quantified. For a small fraction of the emitters, the number of molecules could not be reliably counted because either they photobleached too quickly (<2 frames) or did not photobleach during the duration of the movie, and therefore were excluded from the analysis.

# Chapter 3 : Building a rule-based model of the initial events of EGFR signaling

## 3.1 Introduction

A number of rule-based models for EGFR signaling have been developed in recent years, each varying in complexity and the level of mechanistic details included (Blinov et al., 2006; Creamer et al., 2012; Kozer et al., 2013b). The refinement of these models was mostly done using techniques that provide ensemble or average quantitative measurements. The kind of information that the improved SiMPull technique (described in Chapter 2) provides was not available in the past, and therefore this model will be the first one whose refinement is performed using this type of data. As described in following chapters, coupling of this type of modeling and experimental data is providing new insights about the kinetics of the molecular events involved in EGFR activation.

The following sections present a detailed description of the development of a model that simulate the EGF-dependent activation of EGFR, phosphorylation of tyrosines 1068 and 1173 in the cytoplasmic tail of the receptor and the recruitment of adaptor proteins Grb2 and Shc1 to these sites (Figure 3.1). This model was developed using the BioNetGen language (BNGL) for rule-based modeling (Faeder et al., 2009). The model consists of 9 reaction rules and 12 parameters. BioNetGen software was used to generate the reaction network from these rules. The resulting reaction network consists of 39 molecular species and

183 unidirectional reactions. The network was simulated with the deterministic simulation engine used by BioNetGen.

In the next section, two example rules are presented to briefly describe the basic nomenclature in BioNetGen language. A more detailed description of BioNetGen language can be found in (Faeder et al., 2009). This is followed by a list of the reaction rules used for this model together with a brief description. Lastly, some of the key parameter values used in this model are enlisted, and the source(s) of these values, if available. A complete list of the parameter values and the file encoding the complete model can be found in Appendix B.

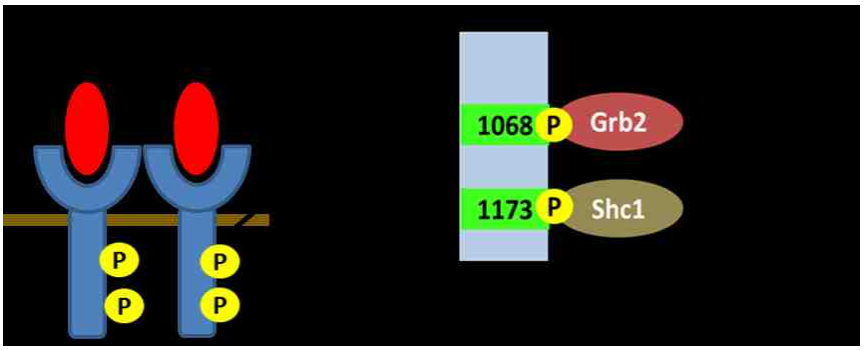


**Figure 3.1.** Graphical description of model. EGF-bound receptors can dimerize and phosphorylated each other. This model includes the phosphorylation of Y1068 and Y1173, and the recruitment of Grb2 and Shc1 to those respective sites.

## 3.2 Model development, Results and Discussion

### 3.2.1 Example reaction rules

**Example Rule 1.** Reversible interaction between A and B

```
A(domA)+B(domB)<->A(domA!1).B(domB!1) kp,km
```

The proteins 'A' and 'B' can interact through their domains 'domA' and 'domB', and form a complex. Components of a complex are separated by a dot and the molecular bonds are indicated with a '!' sign and a number. For example, if more than one molecular bond is present in a molecular complex then the bonds would be labeled as '!1', '!2', '!3', and so on. Note that 'domA' and 'domB' in the reactants side do not have a '!', which means that they have to be free or unbound in order for the association reaction to occur. The forward ('kp' or k+) and reverse ('km' or k-) reaction rates are indicated after the rule and separated by a comma. For unidirectional reactions, only one reaction rate is specified.

**Example Rule 2.** Phosphorylation of protein A when it is bound (to protein B)

```
A(domA!+,Y1~0)->(domA!+,Y1~P) kp
```

Phosphorylation of 'Y1' can only happen if its domain 'domA' is bound to other protein ('domA!+'). In this case, protein 'B' can be thought as a kinase that phosphorylates protein 'A' in 'Y1'. The expression '!+' in 'domA!+' indicates that 'domA' needs to be bound, regardless of to which protein it is bound to. Alternatively, the whole complex can be explicitly specified as indicated below:

```
A(domA!1,Y1~0).B(domB!1)-> A(domA!1,Y1~P).B(domB!1) kp
```

Only domains that are relevant to the reaction are included. For example, in the first example rule the site 'Y1' was not included, because the interaction happens regardless of the state of 'Y1'. BioNetGen software will generate a reaction for each possible molecular species.

### 3.2.2 Reaction rules

**R1.** EGF reversibly binds EGFR

```
EGF(EGFL)+EGFR(I_III)<->EGF(EGFL!1).EGFR(I_III!1) kp_EGF,km_EGF
```

In this simple rule EGF binds to domains I and III on EGFR with a rate 'kp_EGF', and dissociates with a rate 'km_EGF'.

**R2.** Dimerization of EGF-bound receptors

```
EGFR(I_III!+,II~u)+EGFR(I_III!+,II~u)->
EGFR(I_III!+,II~b)+EGFR(I_III!+,II~b) kp_dim_L_L
```

EGFR that is bound to EGF ('I_III' domain occupied) and that has its dimerizing domain 'II' unbound ('II~u') can dimerize with another receptor of the same kind. The dimerization event changes the state of domain II from unbound to bound (II~u->II~b), which is an implicit way of representing dimerization. Originally the dimerization event was described explicitly, meaning that the two receptors actually formed a complex. Comparing simulation results from the models that considered dimerization implicitly and explicitly, both models behaved almost identically (not shown), although the model with explicit dimerization had a much higher number of molecular species and reactions. This reduced complexity of

the implicit dimers is because they have the same number of components as a monomer, while in explicit dimers there are twice as many components as in a monomer. Considering that complexity increases non-linearly with the number of components of a molecule, explicit dimers have many more possible states that implicit dimers (for the current model a factor of 7 increase). This increase in complexity becomes relevant when fitting the model to experimental results, which could take days or even weeks.

*Alternative versions tested*: Dimerization between unliganded-unliganded or liganded-unliganded receptors was initially considered. In this model transition of EGFR ectodomain between tethered and extended conformations was considered implicitly by defining a parameter that defined the probability of receptors to be in an extended conformation (dimerization competent). Comparing simulation results with the model that only allowed two EGF-bound receptors to dimerize showed that they behaved very similar, and that only varied slightly at low EGF doses (not shown). Therefore, the model was simplified and only allow two EGF-bound receptors to form dimers. If the model is intended to be used for studying processes at low EGF doses, like ERK activation at sub-nanomolar EGF concentrations, then it may be necessary to add these rules back into the model.

**R3.** Dissociation of EGFR dimer

```
EGFR(II~b)->EGFR(II~u) km_dim_L_L
```

Similarly to EGFR dimerization, dissociation of EGFR dimers was considered implicitly, giving equal results as when considered explicitly. The dissociation event changes the state of domain II from bound to unbound (II~b->II~u).

**R4 and R5.** EGFR autophosphorylation

```
EGFR(II~b,Y1068~0)->EGFR(II~b,Y1068~P) kphos1068
EGFR(II~b,Y1173~0)->EGFR(II~b,Y1173~P) kphos1173
```

Receptors that are in a dimer (II~b) can be phosphorylated with rate 'kphos'. The state of the tyrosine residue changes from unphosphorylated to phosphorylated (Y~0->Y~P).

Asymmetric arrangement of the kinase domains in an EGFR dimer was not considered for the results presented in this dissertation, but an updated version of the model including asymmetric phosphorylation have just been created, and will be used for the peer-reviewed publication of this work. Although the parameters for phosphorylation rate needed to be modified for this updated model, the simulation results from this model are very similar to the ones obtained with the model used for this dissertation.

**R6 and R7.** Dephosphorylation of pY sites

```
EGFR(Y1068~P)->EGFR(Y1068~0) kdephos1068
EGFR(Y1173~P)->EGFR(Y1173~0) kdephos1173
```

Sites that are phosphorylated (Y~P) and unbound (absence of '!+' sign) can be dephosphorylated with a rate 'kdephos'. Dephosphorylation is considered implicitly as being constitutively active (constant dephosphorylation rate).

**R8.** Binding of Grb2 to pY1068

```
GRB2(SH2)+EGFR(Y1068~P)<-> GRB2(SH2!1).EGFR(Y1068~P!1)
kp_GE,km_GE
```

The SH2 domain from Grb2 binds to phosphorylated tyrosine 1068 in EGFR. Both domains/sites must be free in order to bind.

Even though Grb2 can also bind directly to other sites, like pY1086, or indirectly through Shc1, these interactions were not considered in this model.

**R9.** Binding of Shc1 to pY1173

```
SHC1(PTB)+EGFR(Y1173~P)<-> SHC1(PTB!1).EGFR(Y1173~P!1)
kp_SE,km_SE
```

The PTB domain from Shc1 binds to phosphorylated Y1173 in EGFR.

### 3.2.3 Model parameters

Model parameters presented here were subsequently transformed to units of molecules/cell and to units of seconds.

Avogadro constant **(NA)**

Value: 6.02214 $x10^{23}$ /mol

Used to convert concentrations in molarity to concentrations in molecules per cell.

Cytoplasmic volume **(Vc)**

Value: 1 picoliter

Source: Cytoplasmic volume estimated by Fujioka et al. for HeLa cells (Fujioka et al., 2006).

Grb2 and Shc1 concentration **(GRB2_total and SHC1_total)**

Value: $1.0x10^4$ - $1.0x10^6$ copies/cell

Source: Abundance of these proteins was allowed to vary in the range specified above for the fitting process. Naïve model was set to have values of $1.0x10^5$ for both proteins.

Number of EGFR **(EGFR_total)**

Value: $6.0x10^5$ receptors/cell

Source: Value estimated previously in our laboratory for CHO EGFR-GFP cells using flow cytometry.

Association rate for cytoplasmic interactions **(kon)**

Value: $5.0x10^6$ /M/s

Source: Assumed to be $5.0x10^6$ /M/s for all interactions occurring at the cytoplasm (including receptor-adaptor protein interactions).

Association rate for EGF-EGFR interaction **(kon_EGF)**

Value: $8.0x10^6$ /M/s

Source: This rate was set to this value so that EGFR phosphorylation kinetics occurs similarly to kinetics measured at high temporal resolution using mass spectrometry (Reddy et al., 2016).

Dissociation constant (Kd) for EGF-EGFR interaction **(Kd_EGF)**

Value: 1 nanomolar ($1.0x10^{-9}$ M)

Source: A dissociation constant of 1 nM was used for this interaction, which it is close to previously estimated values (Björkelund et al., 2011). Dissociation rate for this interaction was obtained using the values for **kon_EGF** and **Kd_EGF**.

Dissociation rate for EGF-bound EGFR dimers **(Vc)**

Value: 0.273/s

Source: Dissociation rate estimated in A431 cells by Low-Nam et al. using Single Particle Tracking (Low-Nam et al., 2011).

Dissociation constant (Kd) for EGF-bound EGFR dimers **(KD_dim)**

Value: 60,000 molecules/cell

Source: This parameter was set so that EGFR_total/KD_dim >> 1.

Dissociation constant (Kd) for Grb2-pY1068 EGFR interaction **(Kd_GE)**

Value: 600 nM

Source: As estimated by Morimatsu *et al*. using single-molecule microscopy (Morimatsu, M., Takagi, H., Ota, K. G., Iwamoto, R., Yanagida, T., & Sako, 2007).

Dissociation constant (Kd) for Shc1-pY1173 EGFR interacton **(Kd_SE)**

Value: 600 nM

Source: Assumed to be the same as for Grb2-pY1068. This assumption was based on affinity measurements for protein domains binding to phospho-peptides of these sites showing similar affinities for Grb2-pY1068 and Shc1-pY1173 interactions (Kaushansky et al., 2008; Hause et al., 2012).

Phosphorylation and dephosphorylation rates **(kphos and kdephos)**

Value: 0.5-5.0 /s

Source: Allowed to vary in this range. Congruent with rates estimated by (Kleiman et al., 2011).

# Chapter 4 : Insights on EGFR signaling by integrating computational modeling and single molecule data

## 4.1 Introduction

This Chapter focuses on the integration between the experimental and computational modeling tools developed in Chapters 2 and 3. Chapter 2 describes a series of modifications to the Single Molecule Pull-down (SiMPull) technique to quantify the fraction of site-specific phosphorylation in EGFR. Then, Chapter 3 describes the development of a rule-based model for EGFR signaling. Building on those developments, SiMPull data was used to refine the rule-based model, and vice versa, the rule-based model was used to generate predictions that can be tested experimentally. The overall goal of this experiment-modeling integration is to gain a quantitative understanding of the dynamic behavior of the different processes involved in signaling.

One particular observation that the model could not initially reproduce was the differential phosphorylation in tyrosine 1068 and 1173 of EGFR. In these results, Y1173 consistently had higher phosphorylation levels than Y1068 at different EGF doses and times of stimulation (see Figure 2.5 in Chapter 2). To explore the possible origins of this behavior the rule-based model was used. The model suggested that adaptor proteins are able to protect the phospho-sites to which they bind from dephosphorylation, and that differences in adaptor protein abundances could give rise to differential phosphorylation. Particularly, the model predicted that an increase in the abundance of Grb2 would result in a higher

percentage of receptors phosphorylated at sites to which Grb2 binds. In agreement with this prediction, overexpression of Grb2 caused a dramatic increase in the phosphorylation levels of a Grb2-binding site in EGFR (Y1068), but not in a site which Grb2 does not bind (Y1173). Preliminary results suggest that these observations using protein overexpression may translate to cells naturally expressing different levels of adaptor proteins. Results show that adaptor protein abundances are able to alter the phosphorylation levels of their binding partners resulting in biased phosphorylation *in vivo*.

## 4.2 Results

### 4.2.1 Model can fit experimental data with only allowing adaptor protein abundances to vary

In the previous chapter, a model of EGFR signaling was described, and while many of the parameter values have been estimated experimentally, some others have not. Also, it is important to consider that measurements are approximations of the actual values and that these values likely change between different conditions, including in which cell type the measurements were performed. Therefore, the model parameters need to be adjusted to allow the model to be able to reproduce experimental data. Simulation results from the unmodified or naïve model show similar levels of phosphorylated tyrosine 1068 and 1173 (pY1068 and pY1173), while the experimental data suggests that phosphorylation at Y1173 is higher than for Y1068 (Figure 4.1).

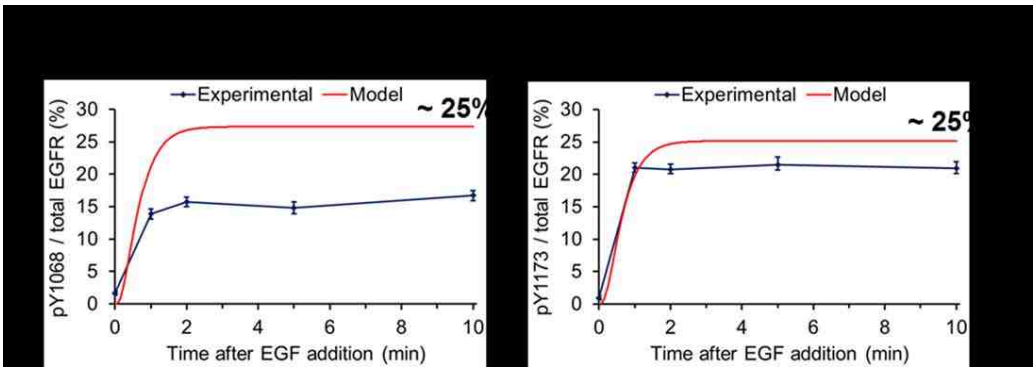


**Figure 4.1.** Comparison of naïve model and experimental data. Percentages of EGFR phosphorylation at tyrosines 1068 **(a)** and 1173 **(b)** after stimulation with 25 nM of EGF obtained experimentally using SiMPull (blue) or predicted by the naïve model (red). Error bars represent mean +/- S.E.M.

There are different parameter values that can be adjusted to reproduce the biased phosphorylation observed experimentally. Probably the most obvious one would be to allow the phosphorylation rate in one site (Y1173) to be higher than in the other. Another possibility is to have different dephosphorylation rates for each site. Phosphorylation rates measured *in vitro* as well as dephosphorylation rates measured *in vivo* were very similar for both sites, and therefore could not account for the biased phosphorylation observed in our experiments (Kim et al., 2012; Kleiman et al., 2011). A third and less obvious possibility would be to allow the abundance of the adaptor proteins (i.e. Grb2 and Shc1) to vary. The reason changes in adaptor protein abundances affect phosphorylation levels in the model is because an adaptor protein bound to the phosphorylated site protects the site from dephosphorylation by means of steric

hindrance (Figure 4.2a). This hypothesis is supported by *in vitro* studies showing the ability of SH2 domains to protect phospho-sites from dephosphorylation (Rotin et al., 1992; Brunati et al., 1998). To test if variation in protein abundances alone could explain the data, abundances of Grb2 and Shc1 were allowed to differentially vary during the fitting process, while phosphorylation and dephosphorylation were varied equally for both sites. The ability of the model to match the experimental results based on differential abundance of adaptor proteins supports the feasibility of this mechanism being responsible for the biased phosphorylation observed (Figure 4.2b,c).

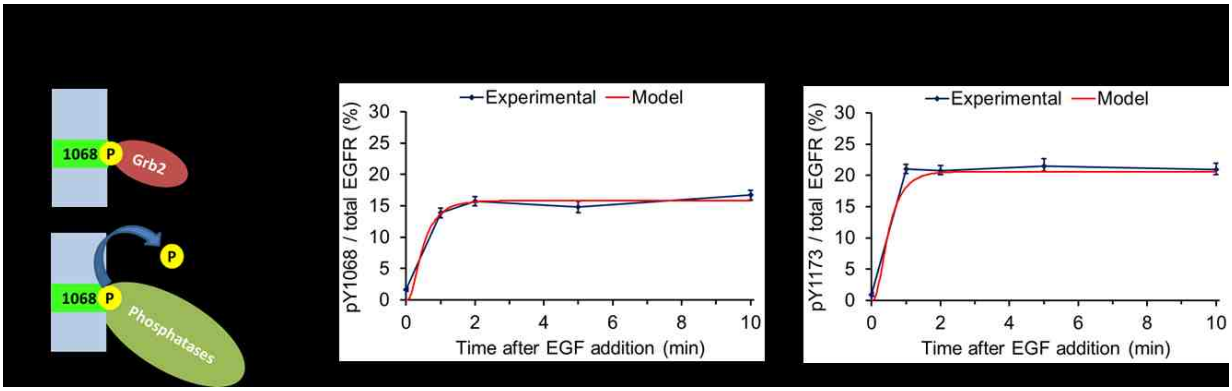


**Figure 4.2.** Fitting model to experimental data by varying adaptor protein abundances. **(a)** Graphical representation of mechanism of protection of phosphorylated sites from phosphatases by adaptor proteins. Percentages of EGFR phosphorylation at tyrosines 1068 **(b)** and 1173 **(c)** after stimulation with 25 nM of EGF obtained experimentally using SiMPull (blue) or predicted by the fitted model (red). Error bars represent mean +/- S.E.M.

### 4.2.2 Predicted influence of Grb2 overexpression in phosphorylation levels is observed experimentally

The fact that the model is able to simulate biased phosphorylation by having different values for the abundances in Grb2 and Shc1 does not supports the hypothesis that adaptor proteins are able to modulate phosphorylation levels. Therefore, a series of predictions of the expected effect that overexpression of Grb2 would have on the phosphorylation levels of the two sites were generated. The model predicts that Grb2 overexpression will lead to increased phosphorylation at Y1068, where Grb2 binds, and no change in phosphorylated Y1173, where Grb2 is not expected to bind (Figure 4.3a,b). To test this prediction, human Grb2-mCherry was overexpressed and its effect on EGFR phosphorylation was quantified. Consistent with model predictions, overexpressing Grb2 lead to a marked increase in the phosphorylation of Y1068, but only a slight increase in the phosphorylation of Y1173 (Figure 4.3c,d).

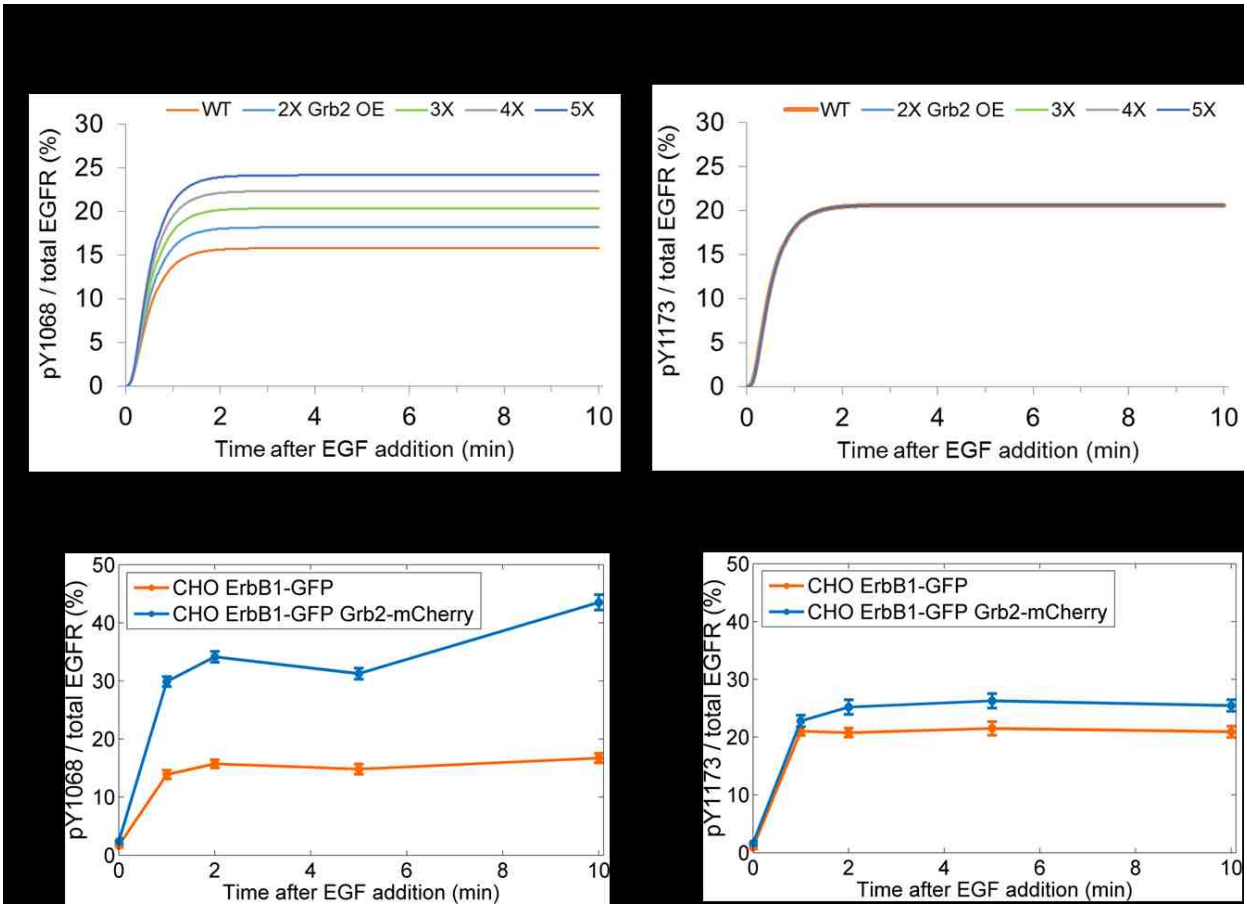


**Figure 4.3.** Predicted and observed phosphorylation kinetics in cells overexpressing Grb2. Predicted percentages of EGFR phosphorylation at tyrosines 1068 **(a)** and 1173 **(b)** after stimulation with 25 nM of EGF in cells with increasing overexpression (OE) of Grb2. Percentages of EGFR phosphorylation at tyrosines 1068 **(c)** and 1173 **(d)** after stimulation with 25 nM of EGF in cells expressing endogenous levels of Grb2 (orange, CHO ErbB1-GFP) or overexpressing Grb2-mCherry (blue). Measurements were done using SiMPull. Error bars represent mean +/- S.E.M.

### 4.2.3 Model predicts cell-specific phosphorylation patterns based on differences in adaptor protein abundances

Considering the observed ability of adaptor protein overexpression to affect the levels of EGFR phosphorylation in a site-specific manner, it can be hypothesized that cell types that naturally express different levels of these adaptor proteins would display different phosphorylation patterns. Using global and targeted proteomics two research groups recently obtained estimates for the abundance of different proteins in different normal and cancer cell lines (Shi et al., 2016; Kulak et al., 2014). These estimates include the protein copy numbers (per cell) for EGFR, Grb2 and Shc1 in the non-tumorigenic mammary epithelial HMEC and MCF10A cells, and in the cervical cancer HeLa cells (Figure 4.4a). Simulations were performed using these values and model predictions for the phosphorylation patterns/kinetics in these three cell lines were obtained (Figure 4.4b-d). In the HMEC cells, where the estimated abundances of both adaptor proteins are relatively low, the model predicts similar levels of phosphorylation at both tyrosine residues. For the MCF10A cells, the model predicts slightly higher phosphorylation at Y1173 given that its binding partner Shc1 is expressed in higher amounts than Grb2. The most evident difference in expression levels is found in HeLa cells, where it is estimated that there are ~600,000 molecules of Grb2 per cell, compared to ~100,000 molecules of Shc1 per cell. The model predicts that phosphorylation at Y1068 would be ~1.45 times higher than at Y1173 (Figure 4.4d). This prediction was tested experimentally in HeLa S3 cells, where phosphorylation at Y1068 was higher than at Y1173, in agreement with

the model prediction (Figure 4.4e). The difference in phosphorylation of these two sites was lower than what the model predicted, with Y1068 being ~1.11 times more phosphorylated than Y1173, compared to ~1.45 times higher pY1068 predicted by the model.

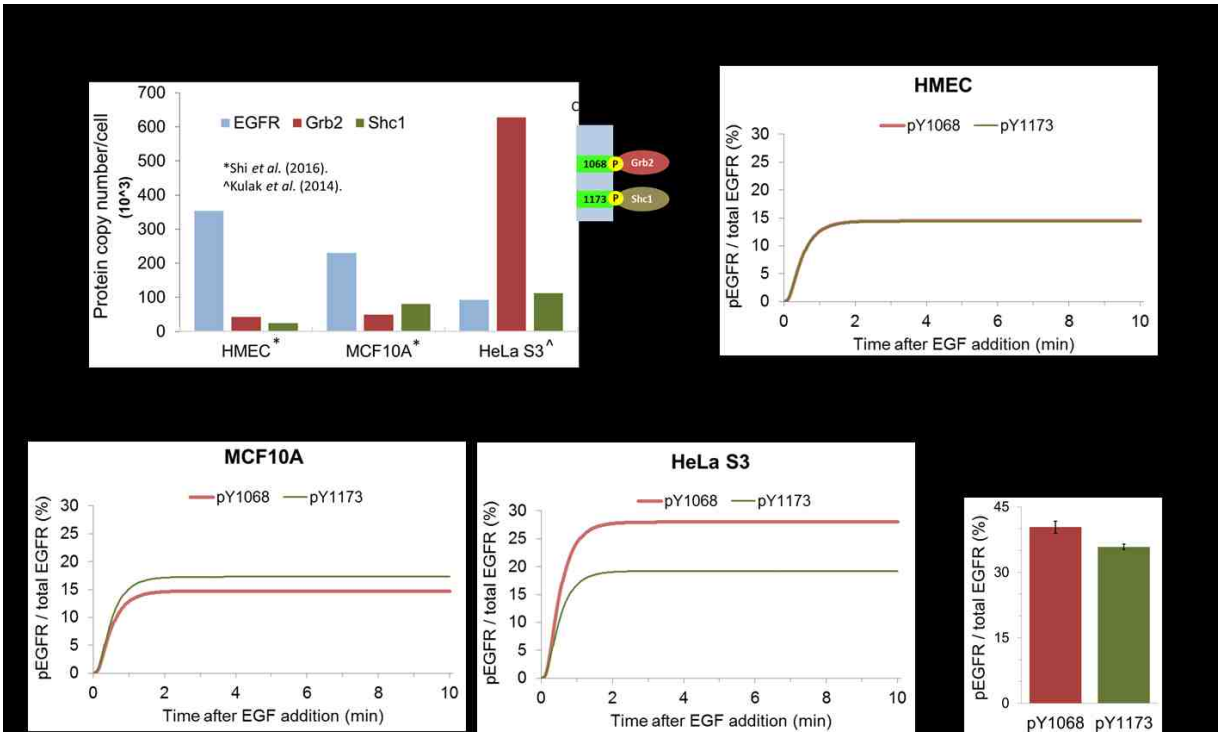


**Figure 4.4.** Phosphorylation patterns predicted for different cell types and observed for HeLa S3 cells. **(a)** Estimation of protein copy numbers for different cell lines estimated by Shi *et al*. (2016) and Kulak *et al*. (2014). **(b-d)** Predictions of phosphorylation kinetics for different cell lines. **(e)** Phosphorylation pattern in HeLa S3 cells obtained with SiMPull. Error bars represent mean +/- S.E.M.

## 4.3 Discussion

Using computational modeling and Single-molecule Pulldown, the ability of adaptor proteins to protect phospho-sites from dephosphorylation and alter phosphorylation levels *in vivo* was shown. As predicted by the model, it was shown that overexpression of the adaptor protein Grb2 leads to increased phosphorylation of an EGFR tyrosine residue where Grb2 binds (Y1068), while a site where Grb2 does not strongly bind is minimally affected (Figure 4.3). The model predicted different phosphorylation patterns for cell lines in which protein abundances were previously estimated (Figure 4.4) (Shi et al., 2016). For example, the model predicted that the cervical cancer HeLa S3 cells, which express ~6 times more Grb2 than Shc1, would have higher phosphorylation levels at the Grb2-binding site Y1068 than at the Shc1-binding site Y1173. Experimental testing of this prediction with SiMPull showed that phosphorylation levels are indeed higher at Y1068 than at Y1173. Nevertheless, the difference in phosphorylation between these sites was not as high as predicted by the model.

The ability of protein domains to protect the phosphorylated residues to which they bind from dephosphorylation has been demonstrated *in vitro* (Rotin et al., 1992; Batzer et al., 1994; Brunati et al., 1998). Even though this protection can happen in the conditions of *in vitro* experiments, in which high concentrations of the reactants are generally used, less is known about the relevance of this phospho-site protection in the context of living cells. A similar study was recently published (Jadwin et al., 2018). In this work, they studied the quantitative relation between Grb2 overexpression and enhancement of site-specific EGFR

phosphorylation (7 sites analyzed) obtaining similar results to the ones presented here. The evidence for the relevance of phospho-site protection at physiological Grb2 expression levels presented in this dissertation should be useful information additional to the one contributed by Jadwin *et al*. Also, while their computational model collapse all the tyrosine residues into one global tyrosine, our rule-based model has site-specific phosphorylation resolution.

The phosphorylation at Y1068 increased ~2-fold when Grb2 was overexpressed, but little change was observed in a site where Grb2 is not known to bind. These results support the idea that adaptor proteins are able to affect phosphorylation levels of the sites to which they bind. In order to get a better idea of the quantitative relation between Grb2 levels and protection of phospho-sites, the concentration of Grb2 in these cells will be measured.

As mentioned before, the model predicted that HeLa S3 cells would be ~1.45 times more phosphorylated at Y1068 than at Y1173, due to its high Grb2 expression levels. Quantification of phosphorylation levels using SiMPull showed that Y1068 phosphorylation was only ~1.11 times higher than Y1173 phosphorylation. This discrepancy may be explained by technical limitations in the quantification of protein abundances or phosphorylation levels, however there are other possible explanations.

It is possible that the ability of Grb2 to protect Y1068 from dephosphorylation is being overestimated, or the ability of Shc1 to protect Y1173 underestimated, or a combination of both. Also it is important to notice that phosphorylation levels in HeLa S3 cells are higher than in CHO cells, even when

expression levels of EGFR are expected to be lower in HeLa cells. A possible explanation for this could be that phosphatase activity is lower in these cells, in which case protection by adaptor proteins would be less apparent. To distinguish these possibilities, the phosphorylation patterns on CRISPR-engineered HeLa S3 clones lacking one or two of the copies of Grb2 or Shc1 will be measured. The level of decrease in Grb2 and Shc1 expression of these clones will be quantified using WB. These measurements together with the computational model will provide with valuable quantitative information about the effect that variations in adaptor protein abundances have on phosphorylation levels. This information could in turn be used in combination with other experiments to explore the relevance of this protection in downstream signaling. This may be of special relevance in cancer cells, which often have aberrant expression levels of adaptor proteins.

Even though the focus of this study is on the adaptor proteins Grb2 and Shc1, these observations likely translate to other proteins binding to post-translationally modified sites. Proteins with interaction lifetimes longer than those of Grb2 and Shc1 may protect their binding sites with a higher efficiency than that of these two proteins. In summary, this study has contributed to the understanding of the interplay between different factors that modulate phosphorylation patterns and kinetics.

## 4.4 Methods

### 4.4.1 Simulations and parameter estimation

A complete description of the model and its parameters can be found in Chapter 3 and Appendix B. Parameter values for adaptor protein abundances, phosphorylation rate and dephosphorylation rate were fitted to experimental data using the open-source software BioNetFit (Thomas et al., 2015). This software uses a genetic algorithm to find the best fit. Configuration files for the fitting process can be found in Appendix C. A list of the values selected by the fitting algorithm for each parameter is presented in Table 1.

**Table 1. List of parameter values selected by fitting algorithm (Chapter 4)**

| Parameter | Parameter Value (fit) |
|---|---|
| Grb2 concentration | 3.51 x$10^4$ molecules/cell |
| Shc1 concentration | 4.72 x$10^5$ molecules/cell |
| Phosphorylation rate | 0.62 /s |
| Dephosphorylation rate | 1.83 /s |

### 4.4.2 Single Molecule Pull-down experiments

For methods and reagents in SiMPull experiments please refer to Methods section in Chapter 2.

# Chapter 5 : A computational model of differential signaling induced by EGFR ligands

## 5.1 Introduction

In this chapter, a computational model that could explain differential signaling induced by EGFR ligands is described. Wilson et. al. observed that stimulating cells with saturating concentrations of different ligands lead to distinct cellular outcomes, with the low-affinity ligands epigen and epiregulin inducing greater cell proliferation than the high-affinity ligand EGF (Wilson et al., 2012). Using protein crystallography, FRET assays and Single Particle Tracking (performed by me and Dr. Diane Lidke), Freed et al. showed that the epigen and epiregulin induced a dimer structure that lacked some key interactions which are present in the EGF-induced dimers, leading to less stable EGFR dimers (Freed et al., 2017). These differences had an effect in receptor phosphorylation/degradation and in both ERK and Akt activation, with epigen/epiregulin stimulation leading to sustained signaling, and EGF-induced activation having a more transient behavior, with signaling being almost completely null by 90 minutes. The fact that these experiments were carried out at saturating ligand concentrations, and that another low affinity ligand (amphiregulin) showed transient EGFR phosphorylation similar to the one induced by EGF, supported the idea that the differential response elicited by these ligands was due to changes in receptor stability rather than in ligand-receptor affinity. The differences in signaling kinetics were also reflected in

cellular outcomes, with epigen and epiregulin leading to cell differentiation after long-term incubation with these ligands, and with EGF inducing cell proliferation.

The difference in signal downregulation observed for these ligands may be due to a combination of multiple factors, including differences in the ability to recruit protein phosphatases and differences in receptor trafficking. Ligand-dependent trafficking of EGFR has been previously reported; for example, TGF-α seems to preferentially induce EGFR recycling to the plasma membrane allowing for sustained ERK activation, whereas EGF promote receptors to be directed to the lysosomal pathway for degradation (Francavilla et al., 2016; Roepstorff et al., 2009).

The fate of EGFR after activation is also dependent on ligand concentration. For example, previous work suggests that low doses of EGF induce receptor internalization almost exclusively through clathrin-mediated endocytosis (CME), leading to relatively low levels of receptor degradation and high recycling (Sigismund et al., 2008). At high doses of EGF, a fraction of the receptors internalize also through CME, but another fraction internalizes through a type of non-clathrin mediated endocytosis that promotes receptor degradation. This type of endocytosis was shown to be dependent on ubiquitination of EGFR (Sigismund et al., 2005). One of the main proteins responsible of EGFR ubiquitination is the E3 ubiquitin ligase Cbl, which can be recruited directly to phosphorylated Y1045 in EGFR or indirectly through Grb2, with which Cbl forms a complex (Waterman, 2002). EGF dose-response curves of EGFR ubiquitination showed that receptor ubiquitination levels are low when EGF concentration is

below a threshold (3-10 ng/mL EGF in HeLa cells), and high after this concentration threshold is surpassed (Sigismund et al., 2013). On the other hand, receptor phosphorylation showed a more gradual increase as a function of EGF concentration. This behavior seems to happen at the plasma membrane, as it was unaffected by inhibition of receptor endocytosis by dynamin knockdown. Their results supported a mechanism in which Cbl-Grb2 complex is recruited in a cooperative fashion to receptors with dual phosphorylation at the Cbl and Grb2 recruiting sites (pY1045-pY1068). This hypothesis was further supported by testing a computational model of Cbl-Grb2 cooperative recruitment and ubiquitination based on quantitative measurements (Capuani et al., 2015). Consistent with previous results, receptor ubiquitination levels directly related to endocytosis of the receptor through non-clathrin mediated endocytosis, leading to degradation.

Using this information, it can be hypothesized that if dimer lifetimes induced by epiregulin or epigen are not long enough to result in significant simultaneous phosphorylation at Y1045 and Y1068, then only a few Cbl-Grb2 complexes will be recruited. This, in turn, would lead to lower receptor ubiquitination and signal downregulation as compared to EGF stimulation. To help guide experimental efforts in testing this hypothesis, the computational model described in Chapter 3 was extended to include cooperative recruitment of Cbl and receptor ubiquitination, and to account for differential dimer stability induced by EGFR ligands. The model predicts that, in cells with low to moderate levels of EGFR, epiregulin will induce notably less ubiquitination than EGF and

explain the different signaling outputs observed. In combination with experiments, this model is expected to improve the quantitative understanding of the processes involved in differential signaling by different EGFR ligands.

## 5.2 Results

### 5.2.1 Model reproduces Y1068-Y1173 dual phosphorylation measured experimentally

Before proceeding to extend the model to account for multi-phosphorylation at the tyrosine pair relevant for Cbl recruitment and ubiquitination (Y1045-Y1068), the performance of this model was tested for its ability to reproduce phosphorylation levels of the Y1068-Y1173 pair, for which SiMPull data was already available. Up to this point the data that have been used for training and testing the model come from 2-color SiMPull experiments, in which the phosphorylation state of a single specific tyrosine residue is determined. Therefore, the model does not provide information about simultaneous phosphorylation at multiple sites. Chapter 2 describes an improved protocol to allow detection of receptor multi-phosphorylation using 3-color SiMPull. In those experiments, the receptors were co-labeled with spectrally distinct antibodies for pY1068 and pY1173, and the percent of receptors with simultaneous phosphorylation at these two sites was estimated. The model predicted that ~6.8% would be phosphorylated in both sites (Figure 5.1a). The percentage of dual phosphorylation estimated experimentally was similar to that predicted by the model (~6.8% vs ~7.2%, Figure 5.1). These results show that

the model is capable of making congruent predictions about dual phosphorylation when provided with single-site phosphorylation data of two sites.

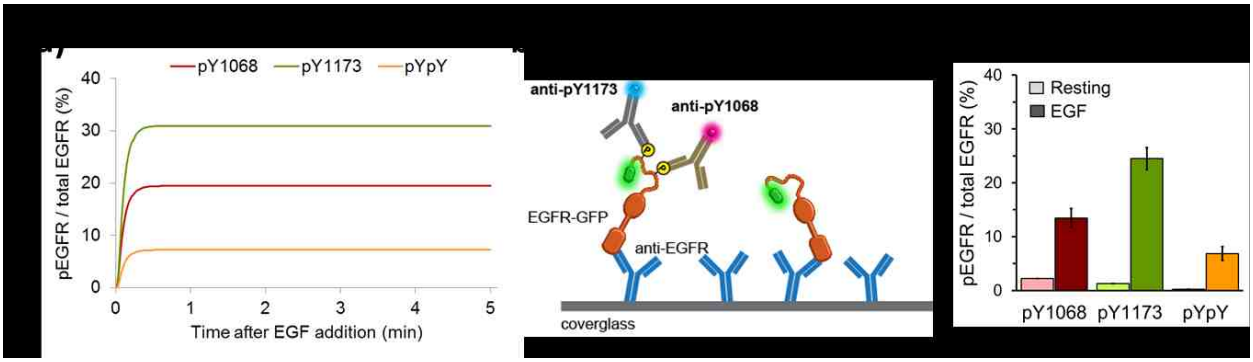


**Figure 5.1.** Predicted and observed dual Y1068-Y1173 phosphorylation. **(a)** Prediction of EGFR phosphorylation at sites Y1068, Y1173, and dual phosphorylation at these sites (pYpY). **(b)** Graphical representation of 3-color SiMPull for the analysis of individual and simultaneous phosphorylation at Y1068 and Y1173. **(c)** 3-color SiMPull results for EGFR from CHO EGFR-GFP cells stimulated with 25 nM of EGF for 5 minutes. Error bars represent mean +/- S.E.M.

### 5.2.2 Extended model for ubiquitination is able to reproduce experimental behavior

The model was extended to include some of the events that impact EGFR ubiquitination levels at the plasma membrane: phosphorylation of tyrosine 1045, direct recruitment of Cbl to pY1045, indirect recruitment of Cbl through the Grb2 and ubiquitination of EGFR by Cbl. The model also considered the cooperative recruitment of the Cbl-Grb2 complex to receptors phosphorylated at both Y1045

and Y1068 (Sigismund et al., 2013; Capuani et al., 2015). The current idea behind this observed cooperativity is that, when a Cbl-Grb2 complex is bound to one site in a dually phosphorylated receptor, the chances of (re)binding to the other site increases due to the increased proximity or local concentration. This, in turn, results in a significant increase in the time that the complex remains bound to these receptors.

To capture this behavior in the model, the association constants of Cbl and Grb2 to the phosphorylated receptor are multiplied by a cooperativity constant ($\boldsymbol{kc}$) in the reactions where the complex is already bound to the receptor and the other tyrosine residue is phosphorylated and unoccupied (Figure 5.2). In a similar way, if the Cbl-Grb2 complex breaks while both Cbl and Grb2 are bound directly to the receptor the association constant for Cbl-Grb2 is multiplied by $\boldsymbol{kc}$. Grb2 can also be recruited to tyrosine 1086 and contribute to the cooperative behavior. Comparison of a model containing one or two Grb2-binding sites showed that both models could fit the experimental data equally well. Thus, only one Grb2 binding site was considered in the model. The model also considers that the number of Cbl molecules available for binding to the receptor is generally low (~5,000-10,000 molecules/cell) compared to other adaptor proteins like Grb2 (~$1.0 \times 10^5$ to $1.0 \times 10^6$ molecules/cell) (Shi et al., 2016; Capuani et al., 2015).

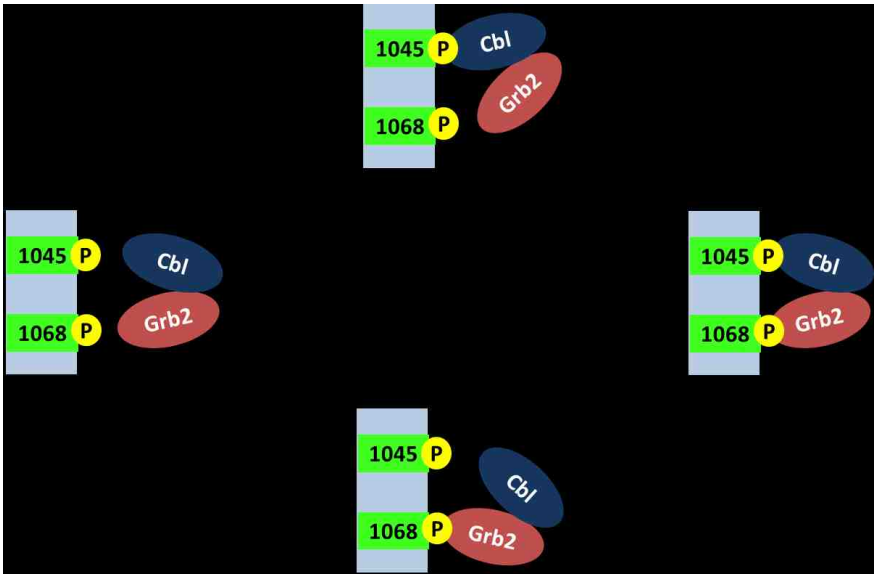


**Figure 5.2.** Graphical description of reactions included in extended model to consider cooperative Cbl recruitment. $\boldsymbol{kc}$ represents the cooperativity constant to account for increased local concentration. For simplification, the dissociation of the Cbl-Grb2 complex while bound to EGFR is not depicted, but in the actual model these reactions were included.

The model was fitted to the dose-response curves of WT receptor ubiquitination and phosphorylation in response to EGF in HeLa cells reported by Sigismund *et al*. (2013). The dataset for ubiquitination of an EGFR mutant lacking the sites for Grb2 recruitment (Y->F mutations) but having Y1045 (Y1045+) was also included for fitting the model. All parameters shared by the previous and extended models were kept at the same value and the parameters unique to the extended model were estimated (see Methods). As observed in Figure 5.3, the model was able to reproduce all three experimental dataset fairly well, with ubiquitination increase happening at lower EGF doses than

phosphorylation, and with ubiquitination in the Y1045+ mutant receptor being drastically lower than in the WT receptor (Figure 5.3a and Figure 5.3b, respectively). Notice that the ubiquitination levels in the model are low (~2%, solid red line in Figure 5.3b). The model was unable to yield a good fit for higher ubiquitination levels.

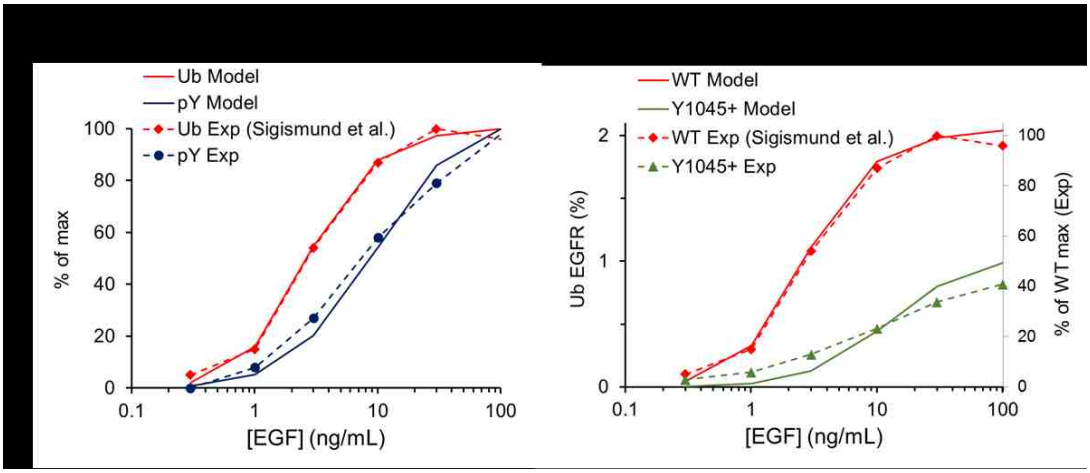


**Figure 5.3** Fitting model to experimental data. Comparison of model simulations (solid lines) to the experimental data (dashed lines) used for the fitting. **(a)** Dose-response curve for receptor phosphorylation at tyrosine 1068 (pY) and ubiquitination (Ub) in cells expressing WT EGFR. **(b)** Ubiquitination curve for WT and mutant receptor devoid of the sites for Grb2 recruitment but containing the site for direct recruitment of Cbl (Y1045+). Data obtained from Sigismund *et al*. (2013).

To validate the model, knockdown (KD) of EGFR was simulated and the ubiquitination dose-response curve from simulation results were compared to the experimental data obtained by Capuani et al. (2015) (Figure 5.4a). Consistent

with experimental results, model simulations showed that decreasing the number of EGFR molecules would result in a shift to the right of the ubiquitination curve. The reason for this behavior is that the number of receptors with dual phosphorylation at Y1045-Y1068 in the EGFR KD cells is much lower than in WT cells, and needs higher EGF concentrations to induce an enough receptors to be dually phosphorylated (and recruit Cbl efficiently) (Figure 5.4b).

When looking at the fraction of receptors ubiquitinated instead of the values normalized to the maximum in Figure 5.4c, it can be observed that the model predicts the ubiquitination percentage to be higher in cells with fewer receptors (EGFR KD, $50x10^3$ receptors) than in WT cells ($250x10^3$ receptors). This is consistent with the results of Capuani *et al.* (2016) and the idea that Cbl is present at a limiting concentration. To additionally validate the model, the effect of Cbl overexpression in receptor ubiquitination observed in this model and experimentally were compared (Figure 5.4d). Consistent with experimental results, an increase in the concentration of Cbl resulted in higher ubiquitination levels but did not result in a shift of the curve to the right or left. It is important to mention that, while this model reproduced most datasets, it was not able to completely match the dual phosphorylation curve at Y1045-Y1068 measured using co-IP WB by Sigismund *et al*. (2013) (Figure 5.4e).

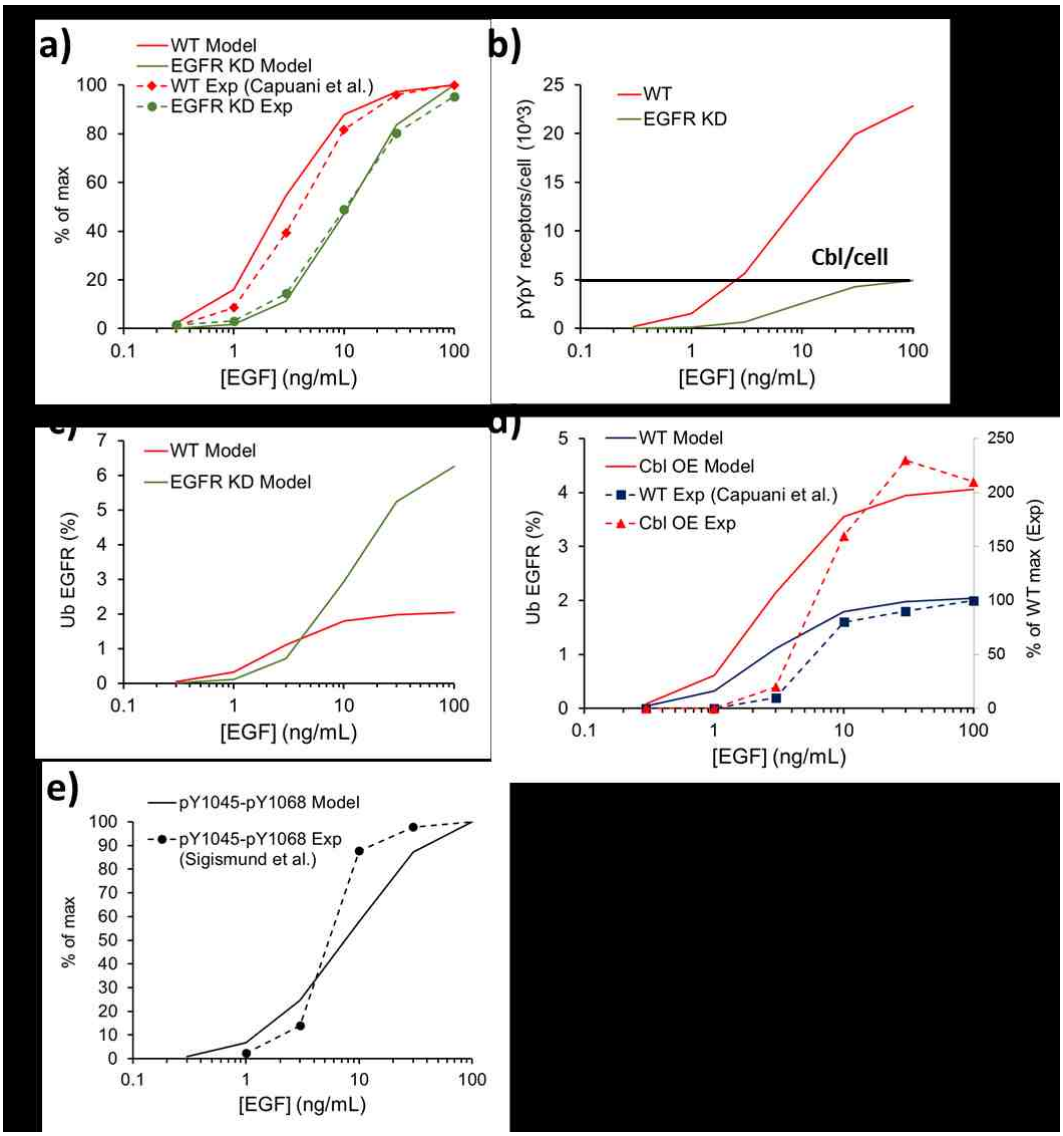


**Figure 5.4**. Model validation: Ubiquitination in cells with EGFR knockdown or Cbl overexpression. Solid lines represent results from model simulations and dashed lines from experimental results from the literature. **(a)** Knockdown (KD) of EGFR was simulated by a 5-fold decrease in the number of receptors per cell. **(b)** pYpY refers to pY1045-pY1068 pair. Dark line indicates the number of Cbl per cell in the model (5,000/cell). **(c)** Same as (a) but not normalized to maximum ubiquitination. **(d)** Cbl overexpression was simulated as a 2-fold increase in the

number of Cbl molecules per cell. **(e)** Comparison of dual phosphorylation in the model and in experimental results by Western blot.

### 5.2.3 Model predicts impaired ubiquitination in epiregulin-stimulated MCF7 cells

The validated model was used to predict the effects that differences in ligand-dependent dimer lifetimes would have in receptor ubiquitination. Even though *in vitro* and *in vivo* studies suggested an important decrease in the stability of epiregulin-induced dimers, dimer off-rates for this condition have not been reported (Freed et al., 2017). Single-Particle Tracking measurements of EGFR dimer off-rates previously performed by Low-Nam et al. in our laboratory suggest that the off-rates of singly EGF-bound dimers is ~9 larger (less stable) than that for doubly EGF-bound dimers in HeLa cells (Low-Nam et al., 2011). Taking this information into consideration, as a starting point a 4 times larger dissociation rate for epiregulin-induced dimers than for dimers bound to two molecules of EGF was used in the model. On rates (dimerization rates) were assumed to be the same for EGF- and epiregulin-bound receptors. To simplify direct comparison of ubiquitination as a function of fraction of ligand-bound receptor, ligand concentration and affinity were kept the same for both ligands.

Simulation results suggest that shorter-lived epiregulin dimers would not be able to induce as strong of ubiquitination as EGF dimers, even at saturating ligand concentrations in MCF7 cells, which express low levels of EGFR (Figure 5.5a). Reduced receptor ubiquitination could result in less EGFR downregulation

and a more sustained signal, as observed experimentally for epiregulin and epigen. The predicted ligand-dependent differences in receptor phosphorylation can be observed in Figure 5.5b.

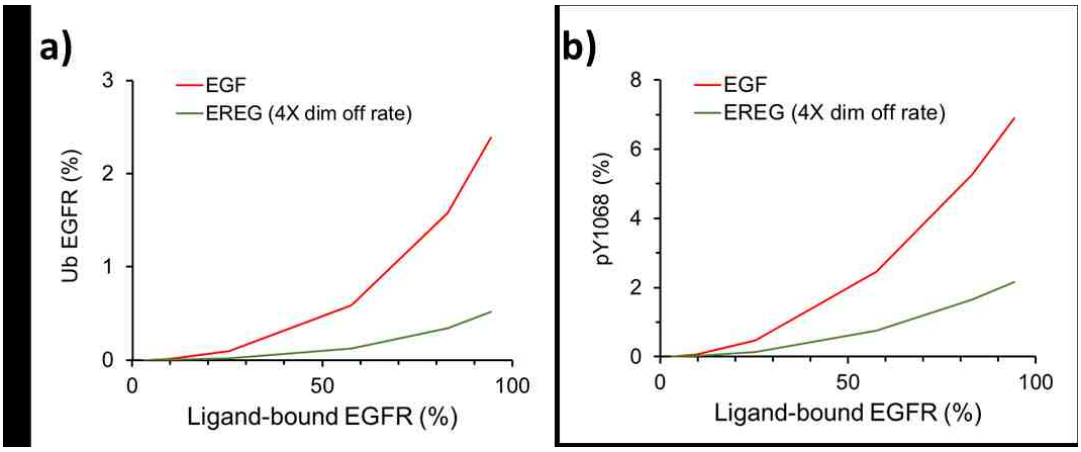


**Figure 5.5**. Predicted ubiquitination and Y1068 phosphorylation levels in MCF7 cells when stimulated with EGF vs epiregulin. Levels of EGFR were set to 5,000 receptors/cell to simulate low expression levels of EGFR observed in MCF7.

### 5.2.4 Using the model to help identify the mechanisms contributing to epiregulin-induced sustained signaling

Even if receptor ubiquitination is significantly lower in epiregulin-treated cells than in those stimulated with EGF, as predicted by the model, this would not rule out the possibility that other mechanisms are responsible for the prolonged EGFR signaling observed upon stimulation with high doses of epiregulin.

Therefore, the model was used to explore if there was a cellular context in which high doses of epiregulin were expected to achieve a similar level of receptor ubiquitination as doses of EGF that induce transient signaling.

Since expression levels of EGFR were shown to alter the dose-response curve of receptor ubiquitination (see Figure 5.4a and (Capuani et al., 2015)), the number of receptors per cell in the model was modulated and the predicted ubiquitination curves for EGF and epiregulin for these cellular contexts were simulated (Figure 5.6, top). The model predicts that as EGFR expression is increased the ubiquitination and phosphorylation response for EGF and epiregulin become more similar, especially at saturating ligand concentrations. If these predictions are confirmed and ubiquitination is responsible for differential signaling, it would expected both ligands to induce similar signaling kinetics (i.e. transient behavior). If, instead, epiregulin-induced signaling continues to be sustained even when having similar levels of receptor phosphorylation/ubiquitination as EGF-activated receptors, it may suggest that sustained signaling is likely originated by another distinct property induced by epiregulin (e.g. ligand-receptor stability in endosomes).

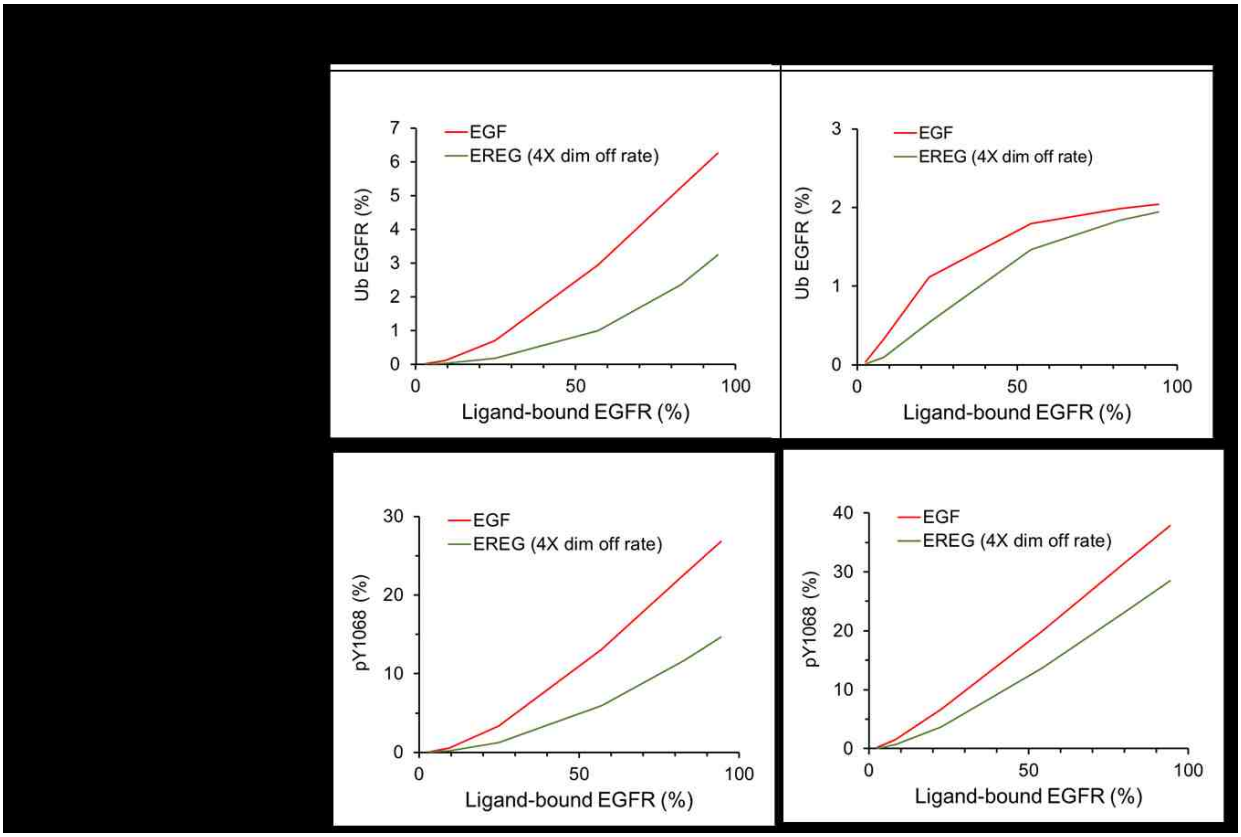


**Figure 5.6**. Predicted EGF- and epiregulin-induced ubiquitination and phosphorylation different EGFR expression levels. The model predicts that as EGFR expression levels increase, EGF vs EREG difference is less notorious.

## 5.3 Discussion

Previous experiments by Roepstorff et al. showed recycling of EGFR when stimulated with epiregulin (Roepstorff et al., 2009). Nevertheless, these experiments were carried out using a ligand concentration of 100 nM, which compared to the concentration necessary to elicit half maximum receptor phosphorylation in cells (20 μM) would be considered a non-saturating concentration (Freed et al., 2017). Therefore, it is unknown how much receptor

recycling and degradation is induced by high doses of ligands such as epigen and epiregulin.

Our model predicted that only a low fraction of receptors (2-6%) are ubiquitinated at the plasma membrane level at early time points (e.g. 2 min) (Figure 5.3b, and top of Figure 5.6). A set of parameters values could not be found that resulted in higher ubiquitination levels and still fit adequately the experimental results. An explanation may be that in the model the stability of the Cbl-Grb2 complex to dually phosphorylated receptors needed to be relatively high (half-life > 20 seconds) in order to fit to the observed ubiquitination curves. The slow unbinding of this complex would limit the number of different receptors to which it can bind at short timescales. SiMPull assay is going to be used to test whether indeed only a small fraction of receptors is ubiquitinated at this time point.

It is possible that only a low fraction of ubiquitinated receptors is needed to induce non-clathrin mediated endocytosis (NCE) of both ubiquitinated and non-ubiquitinated receptors. It is also possible that Cbl is able to ubiquitinate not only the receptor to which it binds but also receptors to which it dimerizes, or nearby receptors in the same microdomains at the plasma membrane. It was estimated that ~22% of receptors are ubiquitinated in at least one site of EGFR after 5 minutes of stimulation with a relatively high dose of EGF (20 ng/mL) (Huang et al., 2013). Considering that Cbl will continue to induce receptor ubiquitination after being endocytosed (Umebayashi et al., 2008), it is feasible that only a small

fraction of receptors is ubiquitinated at the membrane, as represented in our model.

Our model suggested that at higher receptor concentrations than the ones tested experimentally (~$5x10^3$ - $50x10^3$ receptors/cell) by Freed *et al.* (2017), there would be a dose of epiregulin that would induce similar levels of cooperative recruitment of Cbl and receptor ubiquitination as those induced by EGF at some other dose (Figure 5.6, top). It may be possible that sustained signaling kinetics by epiregulin will still be observed in these conditions. If that is the case, one possible explanation would be that the nature of the ubiquitination induced by epiregulin and EGF may be different (e.g. different levels of mono- vs poly-ubiquitination), even if the total ubiquitination measured is similar. Another possibility may lie in differences in receptor trafficking induced by factors other than ubiquitination, such as stability of the ligand-receptor complex inside endosomes. This could happen either by changes in endosomal pH or by ligand unbinding at endosomes, as suggested for the high-affinity ligand for EGFR TGF-α and the low-affinity ligand for EGFR amphiregulin, respectively (Ebner and Derynck, 1991; Roepstorff et al., 2009).

Quantitative measurements will be performed to test the predictions made by the computational model. This will contribute to gaining mechanistic insights about the processes involved in differential signaling induced by EGFR ligands.

## 5.4 Methods

### 5.4.1 Simulations and parameter estimation

Parameter values used in the fitted model of the previous chapter were used for the extended model presented here. A complete list of parameters and reaction rules used for this model can be found in Appendix B. A list of the parameters fitted using BioNetFit for this model, and the values selected by the fitting algorithm, can be found in Table 2. Configuration files for the fitting process can be found in Appendix C.

**Table 2. List of parameter values selected by fitting algorithm (Chapter 5)**

| Parameter | Parameter Value (fit) |
|---|---|
| $K_d$ Cbl-pY1045* | 2.51 $x10^{-7}$ M |
| $K_d$ Cbl-Grb2 | 7.24 $x10^{-7}$ M |
| Cooperativity constant | 2.59 $x10^{8}$ |
| Ubiquitination rate | 5.26 $x10^{-2}$ /s |
| Deubiquitination rate | 1.71 $x10^{-2}$ /s |

* $K_d$ Cbl-pY1045 means dissociation constant (in molar units) between Cbl and pY1045 in EGFR.

# Chapter 6 : Implications and Future Studies

## 6.1 Implications

### 6.1.1 Significance of improvements made on SiMPull technique

The improvements made on the SiMPull technique allows for the quantification of the heterogeneity in the activation states of single molecules. The majority of the proteins involved in signaling have multiple sites of phosphorylation and other post-translational modifications. Therefore, these improvements should prove useful in the study of other signaling players. Additionally, some of the improvements made for quantification of receptor phosphorylation, like reduction of background fluorescence, can be useful for when doing traditional SiMPull where the intention is to quantify the heterogeneity in composition of protein complexes.

### 6.1.2 Significance of understanding role of adaptor protein abundances in biased phosphorylation

By studying the quantitative relation between adaptor proteins and phosphorylation, the quantitative understanding of the different factors affecting phosphorylation levels has been improved. This information can be included in future computational models to have a more realistic representation of the processes modulating phosphorylation. The understanding of the phospho-site protection mechanism described could also help interpreting results in which

variations in protein abundances are involved, as the possibility of protection from phosphatases is not commonly considered. For example, if the phosphorylation levels of two sites are correlated, one hypothesis would be that phosphorylation of one site promotes the phosphorylation of the second, or vice versa, as hypothesized by Coba *et al*. when observing correlation between phosphorylated sites of a neuronal signaling pathway (Coba et al., 2009). But it can also be that the correlation is due to an adaptor protein protecting two nearby sites from dephosphorylation. Additionally, understanding the relation of adaptor protein abundances and phosphorylation could help understand what happens in cancer cells that overexpress certain adaptor proteins.

### 6.1.3 Significance of rule-based model for EGFR activation and ubiquitination by different ligands

By testing the predictions generated by the model described, it is expected that a better mechanistic understanding of differential signaling induced by EGFR ligands will be gained. Understanding how dimer lifetimes translates into different signaling outcomes can be helpful in understanding differences in signaling observed with other ligands, or in other membrane receptors whose binding lifetimes are different to those of EGFR. The rule-based model created in this work can be adapted to represent activation of these receptors. Additionally, this knowledge can be useful in designing intelligent modulators of signaling protein interactions to promote signal downregulation in diseases such as cancer.

## 6.2 Future Studies

### 6.2.1 Effects of adaptor protein abundances in downstream signaling and cellular outcomes

To gain a deeper quantitative understanding of the relationship between protein abundances and phosphorylation levels, our team is generating a series of HeLa S3 clones in which one or two copies of the genes encoding for Grb2 and Shc1 have been knocked out using CRISPR engineering.

Cancer cells commonly have altered expression of signaling proteins. Therefore it would be interesting to know what the effects of these alterations in signaling are. For example, the ERK pathway has been shown to be activated at very low doses of EGF, therefore it would be interesting to know if Grb2 overexpression leads to ERK activation happening at even lower doses, or if it makes the activation stronger. When looking at the effect of protein overexpression, it is difficult to distinguish how much of the observed effect is due to protein availability itself and how much due to site protection. To evaluate these contributions, a model where sites are not protected by adaptor proteins could be created.

### 6.2.2 Experimental testing of model predictions for differential ubiquitination induced by EGFR ligands

Measurements of ubiquitination levels in MCF7 (~5,000 receptors/cell) or T47D (~50,000 receptors/cell) cells in response to saturating concentrations of

EGF and EREG will be performed, in order to see if, as predicted by our model, differential ubiquitination is observed. These measurements could be performed first by Western blot and if a difference is observed, then SiMPull experiments for ubiquitination can be performed. This would provide the percent of receptors ubiquitinated and would be very useful to compare with the current model, which predicts low ubiquitination percentages (2-6%). Another experiment that would provide useful information to refine the model would be to measure phosphorylation at Y1045 with SiMPull. The anti-pY1045 antibody is available in the lab and it will be labeled for use in SiMPull experiments in the near future. Also, it is likely dual phosphorylation at Y1045-Y1068 with SiMPull will be performed. Considering the close proximity of these sites, there may be problems with steric hindrance between these antibodies, in which case, Fab fragments could be generated to decrease chances of steric hindrance.

Additionally, it is going to be tested if phosphorylation and ubiquitination induced by the two ligands become more similar in cells expressing higher levels of receptors, as predicted by the model. If that is the case, then it would be interesting to check if the signaling kinetics is also similar (i.e. transient), supporting the role for ubiquitination in biased signaling. If that is not the case, other mechanisms may be involved, for example differences in ligand-receptor stability in endosomes. EGF has been shown to still induce strong phosphorylation in early endosomes (Francavilla et al., 2016), but if EREG stability decreases in endosomes it could be expected to have lower phosphorylation in early endosomes, and likely increased receptor recycling. This

can be tested by using confocal microscopy by labeling for markers of early endosomes (e.g. Rab5) and EGFR phosphorylation, and also labeling for recycling markers (e.g. Rab11).

### 6.2.3 Study phosphorylation and downstream signaling of EGFR mutant L858R

The L858R EGFR mutant, involved in Non-Small Cell Lung Cancer, has been shown to have increased kinase activity and increased dimer stability (Zhang et al., 2006; Valley et al., 2015). It would be interesting to explore what are the effects of these altered kinetic parameters in biased phosphorylation and in receptor multi-phosphorylation. Preliminary results using 3-color SiMPull show that even though the levels of phosphorylation increase, the biased phosphorylation still remains (Figure 6.1).

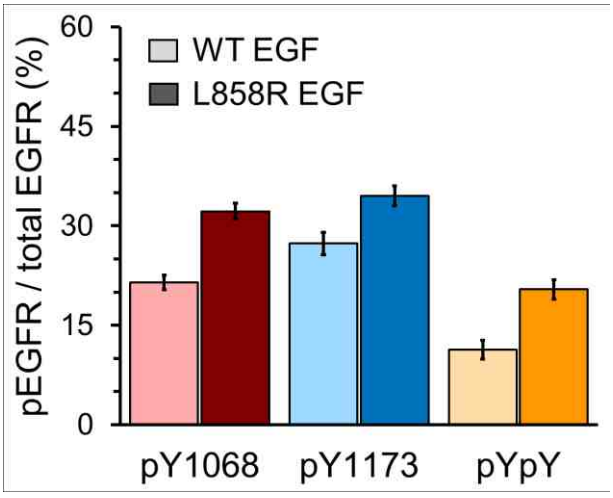


**Figure 6.1**. Single- and multi-phosphorylation in WT vs L858R mutant receptor. Receptors expressed in CHO cells. Receptors were labeled using an anti-pY

(pan-pY) antibody. An anti-EGFR antibody will be used in future experiments to label the receptors. Error bars represent mean +/- S.E.M.

It would also be relevant to understand how dual phosphorylation at Y1045-Y1068 and receptor ubiquitination is affected in the mutant receptor. Shtiegman et al. showed that Cbl recruitment to L858R EGFR and receptor ubiquitination were decreased compared to WT, even when Y1045 and Y1068 were more phosphorylated (Shtiegman et al., 2007). Their results suggest that by favored heterodimerization with HER2 this mutant evades Cbl recruitment and ubiquitination. It would be interesting to know if receptor ubiquitination is increased in this mutant in the absence of HER2, and explore how these changes affect downstream signaling and cellular outcomes.

## APPENDIX A: MATLAB scripts for analysis of SiMPull data

This script was used to localize and process SiMPull data.

```
% Script to overlay fitting in 2 channels
close all;
clear all
clear dat


% Instructions: copy this script inside a 'scripts' folder located in
the
% same location as the files to be analyzed. Copy correct channel
% registration file name. Set threshold for the 3 channels (488 top-
right,
% 561 top-left, and 642 bottom-left; use
% thresholdTest.m to assess your decision). Use joinSequentialCh to
join
% the 3 channels in single images (it creates "joined_" files). Use
formatFilenames.m to create
% list of files ("joined_" files) and 4D-dipimage to assess frames that
should be excluded for
% analysis (bad frames).

% This version (v2_03) allows definin the min P value to filter fits
for
% each individual channel. Therefore you can filter out only irregular
GFP
% molecules that are probably more than one receptors in close
proximity

% This version (v2_02) saves the coordinates of the triple overlay GFP
% molecules in the variable 'coordTripleOv'.

% This version (v2_01) got significantly modified in order to be able
to
% use the new Registration Class that Mark wrote. In order to analyze
the
% fiducial (now acquired using nanogrid) you should use Matlab 2015a or
% later versions.

% This version (v1_04) don't sum the frame selected and the previous
one,
% only the frame specified, given that now we are using an automated
% process for the data acquisition and the shutters from the lasers and
the
% camera are coordinated.

% This version (v1_05) take first image in from right channel and
second
% from left channel because for some reason the green laser (right ch)
started
```

```
% turning on from the first frame acquired (frame 0), when normally is
from
% the second frame (frame 1)

% This version (v1_06) is the v1_05 adapted to get the fields of view
from
% the new QuadView setting. These fields of view are of 256x256 (double
% than before) and we image only 9 fields of view instead of 36.

% This version (v1_06_3color) is the v1_06 adapted to analyze and
overlay
% the 3 channels (488 top-right, 561 top-left, and 642 bottom-left)

% This version (v1_07_3color) fixed some errors
% v1_06_3color had in quantifying % of 561 spots with overlapping 642,
and
% included quantification for overlapping of spots of the three
channels
% (triple overlap)

% This version (v1_08_3color) is the v1_07_3color adapted to analyze
and overlay
% the 3 channels (488 top-right, 561 top-left, and 642 bottom-left)
% acquired with the new laser and instrumentation class. Here the 3rd
% dimension indicates the number of frame or field of view, and the 4th
% indicates the 3 different channels/lasers (642 first, 561 second and
488
% last, acquired sequentially to prevent bleed through and bleaching).

% Version v1_09_3color is v1_08_3color but with higher P value
threshold to
% prevent considering very bright crap that fluoresces in 2 or 3
channels
% and that would be otherwise considered phospohrylated receptors. (P
value
% changed to   instead of 0.0)


intperframe_488= 700; % Intensity to use for thresholding per frame in
the respective channels.
intperframe_561= 600;
intperframe_642= 200;
minPValue_488= 1e-99;%1e-99;2e-1 % Minimum P value for filtering
fits(the bigger the more fits will reject)
minPValue_561= 0;%1e-99; % Minimum P value for filtering fits(the
bigger the more fits will reject)
minPValue_642= 0;%1e-99; % Minimum P value for filtering fits(the
bigger the more fits will reject)

% define files
cd .. % change current directory to one level up
filedir= pwd; % use current directory as filedir
% filedir='E:\Emanuel\DATA\SiMPull\150803 CHO EGFR-GFP A647-NHS-ester
surface';
paramsfile=[filedir '\scripts\SPTparams.mat'];
```

```
% Create folder to save files for tracking
savedir=fullfile(filedir,'totrack');
if ~exist(savedir, 'dir')
  mkdir(savedir);
end

% Create folder to save files to save figures
savedirFigs=fullfile(filedir,'Figures');
if ~exist(savedirFigs, 'dir')
  mkdir(savedirFigs);
end

% Defined location and name of Registration Analysis files
regAnal561file=[filedir '\fiducial\RegAnal_A488_561.mat'];
regAnal642file=[filedir '\fiducial\RegAnal_A488_642.mat'];



% Set file info


% Small sample of files for dissertation

dat(1).filename='joined_CHO-ErbB1-GFP_5min-10nM-EGF_anti-EGFR-
AF555_pY1068-CF640R-2018-2-1-17-45-13';
badFrames{1}= [];

dat(2).filename='joined_CHO-ErbB1-GFP_5min-10nM-EGF_anti-EGFR-
AF555_pY1068-CF640R-2018-2-1-17-47-29';
badFrames{2}= [1];

dat(3).filename='joined_CHO-ErbB1-GFP_5min-10nM-EGF_anti-EGFR-
AF555_pY1068-CF640R-2018-2-1-17-47-51';
badFrames{3}= [];

dat(4).filename='joined_CHO-ErbB1-GFP_5min-10nM-EGF_anti-EGFR-
AF555_pY1068-CF640R-2018-2-1-17-48-11';
badFrames{4}= [1:3];

dat(5).filename='joined_CHO-ErbB1-GFP_5min-10nM-EGF_anti-EGFR-
AF555_pY1068-CF640R-2018-2-1-17-48-50';
badFrames{5}= [1];

dat(6).filename='joined_CHO-ErbB1-GFP_5min-10nM-EGF_anti-EGFR-
AF555_pY1068-CF640R-2018-2-1-17-49-48';
badFrames{6}= [1 2];




for iiDat=1:length(dat)
```

```
    goodFrames{iiDat}= 1:3; % Initialize good frames (it doesn't take
consider first and last frame (0 and 4) because they are used for
focusing)
    goodFrames{iiDat}([badFrames{iiDat}])= []; % Remove bad frames
    for iiFrame=1:length(goodFrames{iiDat})
        dat(iiDat).frms{iiFrame}=goodFrames{iiDat}(iiFrame); % create
structure with frames to be analyzed
    end
end




% Load Registration Analysis files and get transformation function
% Transform 561 based on 488
load(regAnal561file)
SXS = RA.SensorXSplit;      % channel 2 x-coordinates should be > SXS
% Construct the mapping from the right channel to the left.
% Note that mapping occurs in absolute coordinates.
% Optimal map:
M561 = RA.getOptimalMapPixels();
% Maps by algorithm:
%     (1) Null
%     (2) GlobalAffine
%     (3) LocalAffine
%     (4) SmoothAffine
%     (5) LWM
%     (6) Polynomial
%     (7) NRS
%M = RA.maps(5).mapFunctionPixels;

% Transform 642 based on 488load(regAnal561file)
load(regAnal642file)
SXS = RA.SensorXSplit;      % channel 2 x-coordinates should be > SXS
M642 = RA.getOptimalMapPixels();



totFrms= 0; % Initialize total number of frames to be analyzed
% Calculate total number of frames to be analyzed
for ii=1:size(dat,2)
    totFrms= totFrms+size(dat(ii).frms(:),1);
end

% Before looping through initiate define column identifiers and
'valuecell'
% to store the numbers of found fits

% define column identifiers
colIdent= {'FileName', '488-fits','561-fits','642-fits',...
'488-561-overlaps','488-642-overlaps','561-642-overlaps','Triple-
overlap',...
'% 488 w/561','Average %','% 488 w/642','Average %','% 488 w/561-
642','Average %',...
```

```
'% 488-561 w/642','Average %','% 488-642 w/561','Average %'};


valuecell= cell(totFrms+1,numel(colIdent));
valuecell(1,:)= colIdent;
absFrameNum= 0; % Absolute frame number (to correctly place info in
cell array 'valuecell')
blnk=dip_image(zeros(256,256)); % Blank image used in the process of
building gaussian blob images

% loop through each movie defined above
for ii=1:size(dat,2)
     load(fullfile(filedir,dat(ii).filename));
     % because sequence var gets re-written need to rename, Now
     % I'm using dataset instead of sequence so this renaming may
     % be unnecesary in the future
    for jj=1:length(dat(ii).frms)
        clear sptObj642 sptObj488 test currfrms image_out c svfilename
TrackXY642 TrackXY488 tots488 tots642

        absFrameNum= absFrameNum+1; % Absolute frame number (to
correctly place info in cell array 'valuecell')

        % process image, crop and save
        currfrms=dat(ii).frms{jj}; % variable for the current frames to
analyze
        fullsequence=squeeze(datasetJoined(:,:,:));
        test_488=sum(fullsequence(:,:,currfrms),[],3); % do a sum
projection of all frames
        test_488=squeeze(test_488);

        fullsequence=squeeze(datasetJoined(:,:,:));
        test_561=sum(fullsequence(:,:,currfrms),[],3); % do a sum
projection of all frames
        test_561=squeeze(test_561);

        fullsequence=squeeze(datasetJoined(:,:,:));
        test_642=sum(fullsequence(:,:,currfrms),[],3); % do a sum
projection of all frames
        test_642=squeeze(test_642);

        svfilename=[dat(ii).filename '_' mat2str(currfrms(1)-1) '-'
mat2str(currfrms(size(currfrms,2)))]; % create filestring with filename
and frame range

        dipsetpref('DefaultFigureWidth', 512,'DefaultFigureHeight',
512); % set default width of window to 512 to display next image
properly
        h_raw=dipshow(test_642); % use to show the image
%         saveas(h_raw,fullfile(savedirFigs,[svfilename
'_raw.png']),'png')
        saveas(h_raw,fullfile(savedirFigs,[svfilename
'_raw.fig']),'fig')
```

```
        % create sequence of 5 frames- just repeated image so that SPT
code can be
        % used for fitting
        sequence=repmat(test_488(256:end,0:255),[1 1 5]);
        save(fullfile(savedir,[svfilename '_488.mat']),'sequence'); %
save 488 of image

        sequence=repmat(test_561(0:255,0:255),[1 1 5]);
        save(fullfile(savedir,[svfilename '_561.mat']),'sequence'); %
save 561 of image

        sequence=repmat(test_642(0:255,256:511),[1 1 5]);
        save(fullfile(savedir,[svfilename '_642.mat']),'sequence'); %
save 642 of image




        %use spt to find positions of 488 channel
        params=load(paramsfile);
        sptObj488=SPT;
        sptObj488.DataFile=fullfile(savedir,[svfilename '_488.mat'])
        sptObj488.setParams(params);

sptObj488.ParamsFindBoxCenters.MinPhotons=size(currfrms,2)*intperframe_
488;
%         sptObj488.ParamsFindBoxCenters.maxPixelRegionSize=1; % change
        sptObj488.findBoxCenters
        % To prevent error if there are no spots in the image
        if size(sptObj488.BoxCenters,1) == 0
            TrackXY488 = [];
        else

        %
%         sptObj488.ParamsFilterFits.MinPhotons=.0001;
        sptObj488.ParamsFilterFits.MinPValue= minPValue_488;
        sptObj488.filterFits;

         %to view overlay of fits and raw data (note that fits are
shifted already)
        %         sptObj488.plotFitResults %plot results to compare
        %          sptObj488.SaveBaseName=fullfile(savedir,[svfilename
        %          '_488.mat']); --don't actually need this
        sptObj488.saveFile(0);

        % create gaussian blob series of 488 data (no shifting)
        TrackXY488=[sptObj488.FitResults(1).xCoord(:,1)
sptObj488.FitResults(1).yCoord(:,1)]; %must be Nx2 matrix
        end
        clear blnktot
        blnk=dip_image(zeros(256,256));
        if size(TrackXY488,1) == 1  % if statement added to avoid
creating a 2D instead of a 3D dip image when there are no spots
            blnktot= dip_image(zeros(256,256,2));
```

```
            blnktot= blnktot(:,:,1);
        else
            blnktot=dip_image(zeros(256,256,size(TrackXY488,1)));
        end
        for kk=1:size(TrackXY488,1)
            blnktot(:,:,kk-
1)=gaussianblob(blnk,TrackXY488(kk,:),1,5000,'spatial',3);
        end
        tots488=sum(blnktot,[],3);




        %use SPT to find positions of 561 channel
        params=load(paramsfile);
        sptObj561=SPT;
        sptObj561.DataFile=fullfile(savedir,[svfilename '_561.mat'])
        sptObj561.setParams(params);

sptObj561.ParamsFindBoxCenters.MinPhotons=size(currfrms,2)*intperframe_
561; %have intensity value proportional to # of frames summed
%         sptObj561.ParamsFindBoxCenters.maxPixelRegionSize=1; % change
        sptObj561.findBoxCenters
        % To prevent error if there are no spots in the image (below is
        % more code to prevent same error)
        if size(sptObj561.BoxCenters,1) == 0
            ShiftedTrackXY561 = [];
            TrackXY561 = [];

        else
%         sptObj561.ParamsFilterFits.MinPhotons=.0001;
        sptObj561.ParamsFilterFits.MinPValue= minPValue_561;
        sptObj561.filterFits;
        %         sptObj561.plotFitResults %plot results to compare
        %         sptObj561.SaveBaseName=fullfile(savedir,[svfilename
'_561.mat']); --don't actually need this
        sptObj561.saveFile(0); % save .spt files

        % Shift tracks from 561 channel
        TrackXY561=[sptObj561.FitResults(1).xCoord(:,1)
sptObj561.FitResults(1).yCoord(:,1)]; %must be Nx2 matrix
        % Adding SXS pixels shifts channel 2 x-coordinates from
relative coordinates
        % with channel 2 in isolation back to the absolute coordinates
used to
        % construct the map with the channels side-by-side.  The
mapping then
        % transforms channel 2 into the coordinate space of channel 1.
        [ ShiftedTrackXY561 ] = M561([TrackXY561(:, 1) + SXS,
TrackXY561(:, 2)]);
        save(fullfile(savedir,[svfilename '_561shiftedvals.mat']),
'ShiftedTrackXY561'); %save shifted values
        end

        % create gaussian blob series of shifted data
```

```
        % Code modified to prevent error when 1 spot
        clear blnktot
        blnk=dip_image(zeros(256,256));
        if size(ShiftedTrackXY561,1) == 1  % if statement added to
avoid creating a 2D instead of a 3D dip image when there are no spots
            blnktot= dip_image(zeros(256,256,2));
            blnktot= blnktot(:,:,1);
        else

blnktot=dip_image(zeros(256,256,size(ShiftedTrackXY561,1)));
        end

        for kk=1:size(ShiftedTrackXY561,1)
            blnktot(:,:,kk-
1)=gaussianblob(blnk,ShiftedTrackXY561(kk,:),1,5000,'spatial',3);
        end
        tots561=sum(blnktot,[],3);
%          joinchannels('rgb',c(0:255,:),squeeze(tots561)*10) %--
uncomment




        %use SPT to find positions of 642 channel
        params=load(paramsfile);
        sptObj642=SPT;
        sptObj642.DataFile=fullfile(savedir,[svfilename '_642.mat'])
        sptObj642.setParams(params);

sptObj642.ParamsFindBoxCenters.MinPhotons=size(currfrms,2)*intperframe_
642; %have intensity value proportional to # of frames summed
%         sptObj642.ParamsFindBoxCenters.maxPixelRegionSize=1; % change
        sptObj642.findBoxCenters
        % To prevent error if there are no spots in the image (below is
        % more code to prevent same error)
        if size(sptObj642.BoxCenters,1) == 0
            ShiftedTrackXY642 = [];
            TrackXY642 = [];

        else
%         sptObj642.ParamsFilterFits.MinPhotons=.0001;
        sptObj642.ParamsFilterFits.MinPValue= minPValue_642;
        sptObj642.filterFits;
        %         sptObj642.plotFitResults %plot results to compare
        %         sptObj642.SaveBaseName=fullfile(savedir,[svfilename
'_642.mat']); --don't actually need this
        sptObj642.saveFile(0); % save .spt files

        % Shift tracks from 642 channel
        TrackXY642=[sptObj642.FitResults(1).xCoord(:,1)
sptObj642.FitResults(1).yCoord(:,1)]; %must be Nx2 matrix
        % Adding SXS pixels shifts channel 2 x-coordinates from
relative coordinates
        % with channel 2 in isolation back to the absolute coordinates
used to
```

```
        % construct the map with the channels side-by-side.  The
mapping then
        % transforms channel 2 into the coordinate space of channel 1.
        [ ShiftedTrackXY642 ] = M642([TrackXY642(:, 1) + SXS,
TrackXY642(:, 2)]);
        save(fullfile(savedir,[svfilename '_642shiftedvals.mat']),
'ShiftedTrackXY642'); %save shifted values
        end

        % create gaussian blob series of shifted data
        % Code modified to prevent error when 1 spot
        clear blnktot
        blnk=dip_image(zeros(256,256));
        if size(ShiftedTrackXY642,1) == 1  % if statement added to
avoid creating a 2D instead of a 3D dip image when there are no spots
            blnktot= dip_image(zeros(256,256,2));
            blnktot= blnktot(:,:,1);
        else

blnktot=dip_image(zeros(256,256,size(ShiftedTrackXY642,1)));
        end

        for kk=1:size(ShiftedTrackXY642,1)
            blnktot(:,:,kk-
1)=gaussianblob(blnk,ShiftedTrackXY642(kk,:),1,5000,'spatial',3);
        end
        tots642=sum(blnktot,[],3);
%          joinchannels('rgb',c(0:255,:),squeeze(tots642)*10) %--
uncomment




        % To prevent error if there are no spots in any of the channels
        % when trying to overlay shifted blobs

        if size(TrackXY488,1)==0
            tots488= dip_image(zeros(256,256,2));
            tots488= tots488(:,:,1);
        end


        if size(ShiftedTrackXY561,1)==0
            tots561= dip_image(zeros(256,256,2));
            tots561= tots561(:,:,1);
        end


        if size(ShiftedTrackXY642,1)==0
            tots642= dip_image(zeros(256,256,2));  % if statement added
to avoid creating a 2D instead of a 3D dip image when there are no
spots
            tots642= tots642(:,:,1);
        end
```

```
        % overlay shifted blobs on raw data
        ovlayimage=joinchannels('rgb',tots642,tots488,tots561);
        h_raw=dipshow(ovlayimage); % use to show the image

        % save shifted images
        saveas(h_raw,fullfile(savedirFigs,[svfilename
'_overlayedchannels.fig']),'fig')
        saveas(h_raw,fullfile(savedirFigs,[svfilename
'_overlayedchannels.png']),'png')
%         close(h)% close image after saving


        % Calculate number of spots in 561 channel overlapping with
spots in 488 channel
        clear blnktot

        if size(ShiftedTrackXY561,1) <= 1  % "if" statement added to
avoid creating a 2D instead of a 3D dip image when there are only 1 or
no spots
            blnktot= dip_image(zeros(256,256,2));
            blnktot= blnktot(:,:,1);
        else

blnktot=dip_image(zeros(256,256,size(ShiftedTrackXY561,1)));
        end
        numoverlap561w488 = 0; % number of overlapping blobs
        pixallow = 1.0; % number of pixels allowed to be shifted
between overlapping blobs to consider them as being the same molecule

        for mm=1:size(ShiftedTrackXY561,1)
            for nn=1:size(TrackXY488,1)
                if abs(ShiftedTrackXY561(mm,:)-TrackXY488(nn,:)) <
pixallow
                    numoverlap561w488 = numoverlap561w488+1;
                    blnktot(:,:,mm-
1)=gaussianblob(blnk,ShiftedTrackXY561(mm,:),1,5000,'spatial',3);
                end
            end
        end
        tots561ov488=sum(blnktot,[],3); % image with total overlapping
blobs in left channel



        % Calculate number of spots in 642 channel overlapping with
spots in 488 channel
        clear blnktot

        if size(ShiftedTrackXY642,1) <= 1  % "if" statement added to
avoid creating a 2D instead of a 3D dip image when there are only 1 or
no spots
            blnktot= dip_image(zeros(256,256,2));
            blnktot= blnktot(:,:,1);
```

```
        else

blnktot=dip_image(zeros(256,256,size(ShiftedTrackXY642,1)));
        end
        numoverlap642w488 = 0; % number of overlapping blobs
        pixallow = 1.0; % number of pixels allowed to be shifted
between overlapping blobs to consider them as being the same molecule

        for mm=1:size(ShiftedTrackXY642,1)
            for nn=1:size(TrackXY488,1)
                if abs(ShiftedTrackXY642(mm,:)-TrackXY488(nn,:)) <
pixallow
                    numoverlap642w488 = numoverlap642w488+1;
                    blnktot(:,:,mm-
1)=gaussianblob(blnk,ShiftedTrackXY642(mm,:),1,5000,'spatial',3);
                end
            end
        end
        tots642ov488=sum(blnktot,[],3); % image with total overlapping
blobs in left channel



        % Calculate number of spots in 642 channel overlapping with
spots in 561 channel
        clear blnktot

        if size(ShiftedTrackXY642,1) <= 1  % "if" statement added to
avoid creating a 2D instead of a 3D dip image when there are only 1 or
no spots
            blnktot= dip_image(zeros(256,256,2));
            blnktot= blnktot(:,:,1);
        else

blnktot=dip_image(zeros(256,256,size(ShiftedTrackXY642,1)));
        end
        numoverlap642w561 = 0; % number of overlapping blobs
        pixallow = 1.0; % number of pixels allowed to be shifted
between overlapping blobs to consider them as being the same molecule

        for mm=1:size(ShiftedTrackXY642,1)
            for nn=1:size(ShiftedTrackXY561,1)
                if abs(ShiftedTrackXY642(mm,:)-ShiftedTrackXY561(nn,:))
< pixallow
                    numoverlap642w561 = numoverlap642w561+1;
                    blnktot(:,:,mm-
1)=gaussianblob(blnk,ShiftedTrackXY642(mm,:),1,5000,'spatial',3);
                end
            end
        end
        tots642ov561=sum(blnktot,[],3); % image with total overlapping
blobs in left channel
```

```
        % Calculate number of spots overlapping in the three channels
(488, 561 and 642)
        clear blnktot

        if size(ShiftedTrackXY642,1) <= 1  % "if" statement added to
avoid creating a 2D instead of a 3D dip image when there are only 1 or
no spots
            blnktot= dip_image(zeros(256,256,2));
            blnktot= blnktot(:,:,1);
        else

blnktot=dip_image(zeros(256,256,size(ShiftedTrackXY642,1)));
        end
        numTripleOverlap = 0; % number of overlapping blobs
        coordTripleOv = zeros(size(TrackXY488,1),2); % empty array to
place coordinates of 488 with triple overlap
        pixallow = 1.0; % number of pixels allowed to be shifted
between overlapping blobs to consider them as being the same molecule

        for mm=1:size(ShiftedTrackXY642,1)
            for nn=1:size(ShiftedTrackXY561,1)
                if abs(ShiftedTrackXY642(mm,:)-ShiftedTrackXY561(nn,:))
< pixallow
                    % If 642 spot overlaps with 561, then loop through
the
                    % 488 spots to check if it also overlaps with 488
spot
                    for nnTriple=1:size(TrackXY488,1)
                        if abs(ShiftedTrackXY642(mm,:)-
TrackXY488(nnTriple,:)) < pixallow
                            numTripleOverlap = numTripleOverlap+1;
                            coordTripleOv(numTripleOverlap,:)=
TrackXY488(nnTriple,:); % store number of tracks in 488 channel that
have triple overlap
                            blnktot(:,:,mm-
1)=gaussianblob(blnk,ShiftedTrackXY642(mm,:),1,5000,'spatial',3);
                        end
                    end

                end
            end
        end
        totsTripleOverlap=sum(blnktot,[],3); % image with total
overlapping blobs in left channel

        % Remove rows with all zeros and save coordinates
        coordTripleOv(all(coordTripleOv==0,2),:)=[];
        save(fullfile(savedir,[svfilename
'_488_tripleOverlapCoord.mat']), 'coordTripleOv');



        % Display image with only of those spots overlapping with the
ones
        % in 488 channel
```

```
        if isempty(tots642ov488)==0 || isempty(tots561ov488)==0

ovlayOnly=joinchannels('rgb',tots642ov488,tots488,tots561ov488); %
shows all green blobs and only those red (642) and blue (561) blobs
that overlay
        h_ovlayOnly=dipshow(ovlayOnly); % use to show the image
        saveas(h_ovlayOnly,fullfile(savedirFigs,[svfilename
'_overlayW488only.png']),'png')
        else
        end

%         % Display image with only of those 642 spots overlapping with
the ones
%         % in 561 channel
%         if isempty(tots642ov561)==0
%         ovlayOnly=joinchannels('rgb',tots642ov561,tots488,tots561); %
shows all green blobs and only those red blobs that overlay
%         h_ovlayOnly=dipshow(ovlayOnly); % use to show the image
%         saveas(h_ovlayOnly,fullfile(savedirFigs,[svfilename
'_overlaySpotsOnly.png']),'png')
%         else
%         end


        % For reference below are the current column identifiers for
the
        % cell array or spread sheet
%         colIdent= {'FileName', '488-fits','561-fits','642-fits',...
%         '488-561-overlaps','488-642-overlaps','561-642-
overlaps','Triple-overlap',...
%         '% 488 w/561','Average %','% 488 w/642','Average %','% 488
w/561-642','Average %',...
%         '% 488-561 w/642','Average %','% 488-642 w/561','Average %'};

        % Concatenate results to cell array (see column identifiers
above)
        % Add 1 (+1) to absFrameNum to count for column identifiers

valuecell(absFrameNum+1,:)={svfilename,size(TrackXY488,1),size(TrackXY5
61,1),size(TrackXY642,1)...

,numoverlap561w488,numoverlap642w488,numoverlap642w561,numTripleOverlap
,...

numoverlap561w488/size(TrackXY488,1)*100,'',numoverlap642w488/size(Trac
kXY488,1)*100,'',numTripleOverlap/size(TrackXY488,1)*100,'',...

numTripleOverlap/numoverlap561w488*100,'',numTripleOverlap/numoverlap64
2w488*100,''};

        % Create overlayed localizations and raw data
        locMask= newim(512,512); % Create figure to save mask with
localizations
        locMask(256:end,0:255)= squeeze(tots488>600); % Add mask to
corresponding areas/channel
        locMask(0:255,0:255)= squeeze(tots561>600);
```

```
        locMask(0:255,256:511)= squeeze(tots642>600);

        h_rawOverlay=dipshow(overlay(test_642,locMask,[15000,0,0])); %
use to show the image
        saveas(h_rawOverlay,fullfile(savedirFigs,[svfilename
'_rawOverlayLoc.fig']),'fig')


    end
    close all
end



%after loop write to excel and matlab files
xlswrite([filedir '\results.xls'],valuecell) % data saved in excel file
save([filedir '\results.mat'],'valuecell','coordTripleOv') % data saved
in matlab file

cd scripts % change current directory back to scripts folder
```

# APPENDIX B: BNGL files of rule-based models

## BNGL file for model in Chapter 4

```
begin model


# References
# 1. Hause et al., 2012. Plos ONE.
# 2. Engelmann BW et al., 2014. Mol Cell Proteomics.
# 3. Kulak NA et al. (2014) Nat Methods 11: 319-324.
# 4. Shankaran H. et al., 2012. Molecular BioSystems.
# for plasma membrane the ref. is Hendriks 2003 (Cancer Research) and for
endosomes (pH 6.0) from French 1995
# 5. Kholodenko B.N. et al., 1999. The Journal of Biological Chemistry
# 6. Blinov M.L. et al., 2006. BioSystems
# 7. Chook, Yuh Min, et al. "The Grb2-mSos1 complex binds phosphopeptides with
higher affinity than
#    Grb2." Journal of Biological Chemistry 271.48 (1996): 30472-30478.
# 8. Macdonald JL, Pike LJ (2008) Proc Natl Acad Sci USA 105: 112-117.
# 9. Macdonald-Obermann JL, Pike LJ (2009) J Biol Chem 284: 13570-13576.
# 10. Elleman TC et al. (2001) Biochemistry 40: 8930-8939.
# 11. Low-Nam ST et al. (2011) Nat Struct Mol Biol 18: 1244-1249.
# 12. Kleiman LB et al. (2011) Mol Cell 43: 723Ð737.
# 13. Kim Y et al. (2012) Biochemistry 51 (25). American Chemical Society:
5212–22.
# 14. Morimatsu, Miki, et al. "Multiple-state reactions between the epidermal
growth factor
#     receptor and Grb2 as observed by using single-molecule analysis."
Proceedings of the National
#     Academy of Sciences 104.46 (2007): 18013-18018.
# 15. Reddy, Raven J., et al. "Early signaling dynamics of the epidermal
growth
#     factor receptor." Proceedings of the National Academy of Sciences 113.11
(2016): 3114-3119.


begin parameters

GRB2_total__FREE__ 3.50998295e+04 # Vary between 1e4 and 1e6 copies per cell
SHC1_total__FREE__ 4.72130741e+05 # Vary between 1e4 and 1e6 copies per cell
kdephos1068__FREE__ 1.83276225e+00 # Vary between 0.01 and 10
kphos1068__FREE__ 6.23793050e-01 # Vary between 0.01 and 10



# Keep constant

kon__ 5.0e6 # Assumed to be 5.0e6 /M/s
```

```
kon_EGF__ 8.0e6 # This rate was set so EGFR phosphorylation kinetics occurs
similarly as observed by Reddy et. al (2016)
Kd_EGF__ 1.0e-9 # A typical value of 1 nM for the EGF dissociation constant
was used
ratio_kdephos__ 1.0 # Equal phosphoryaltion rates for pY1068 and pY1173 were
assumed
ratio_kphos__ 1.0 # Equal dephosphoryaltion rates for pY1068 and pY1173 were
assumed



# Avogadro constant
NA 6.02214e23 # [=] molecules per mol

# Fraction of cell to consider in a stochastic simulation
f 1 # [=] dimensionless, 0<=f<=1


# Cytoplasmic volume
#   A volume of 1 to 2 pL is typical for a mammalian cell.
Vc f*1.0e-12 # [=] L (1.0 pL)

# Number of cells per dish
numCells 1.0e7 # [=] cells per 60 mm^2 dish (10 million)

# Volume of media per dish
volMedia 1.0e-2 # [=] L (10 mL)

# Volume of extracellular fluid surrounding a cell
Vextra=f*volMedia/numCells


GRB2_total   GRB2_total__FREE__*f # [=] molecules per cell
EGFR_total 6.0e5*f # [=] molecules per cell (as estimated by flow cytometry
for this CHO EGFR-GFP cells)
SHC1_total   SHC1_total__FREE__*f # [=] molecules per cell



# Concentration of EGF
EGFconc 0 # [=] M
# EGFconc 25.0e-9 # [=] M

EGF_total=EGFconc*(NA*Vextra) # [=] molecules per cell



# A typical association rate constant for a protein-protein interaction
# A value of 1e7/M/s instead of 1e6/M/s (used originally) is closer to the on
rate estimated in Ref. 14 , which is ~1e7-1e8/M/s
# This value also allows off rates to be more similar to those estimated in
Ref. 14.
```

```
kon kon__ # [=] /M/s

kon_EGF kon_EGF__ # [=] /M/s

# Dissocation and association rate constants for EGF-EGFR interaction at the
plasma membrane
Kd_EGF Kd_EGF__*(NA*Vextra) # [=] molecules . 1.0e-9nM
kp_EGF=kon_EGF/(NA*Vextra)
km_EGF=Kd_EGF*kp_EGF


# Dissociation constant for EGFR dimerization
#   This parameter is set so that EGFR_total/KD_dim >> 1 when number of
receptors is high (e.g. 6.0e5)
KD_dim 6.0e5/10 # [=] molecules per cell


# Dissocation and association rate constants for interaction between two
liganded (EGF-bound) receptors
# '_pre' because off rates will be modified by a factor of 'offrate_f'. If
offrate_f=1 then
# 'km_dim_L_L_pre' and 'km_dim_L_L' are the same.
km_dim_L_L_pre 0.273 # [=] /s (Ref 11)
kp_dim_L_L=km_dim_L_L_pre/KD_dim # [=] /(molecule/cell)/s


# Increase off rates by a factor of 'offrate_f'
offrate_f 1.0
km_dim_L_L=km_dim_L_L_pre*offrate_f



# Kd for Grb2-SH2 domain binding to pY1068 EGFR
Kd_GE 0.6e-6*(NA*Vc) # [=] molecules   , as estimated by Morimatsu et al.
(2007)
kp_GE=kon/(NA*Vc)
km_GE=Kd_GE*kp_GE


# Kd for SHC1-PTB domain binding to pY1173 EGFR
Kd_SE 0.6e-6*(NA*Vc) # [=] molecules , assumed to have equal Kd as Grb2
kp_SE=kon/(NA*Vc)
km_SE=Kd_SE*kp_SE




# Generic (pseudo first-order) dephosphorylation rate constant
#     From Ref. 12 we have that dephosphorylation rate of pY-EGFR on cells
after using gefitinib is
#     0.05/s, which corresponds to a half-life of 15 s and represents a lower
#     bound (estimation without considering protection of sites by binding
proteins and other influencing factors)
#   When considering Shc binding in the paper they come with a model (M3) with
1/sec rate for dephosphorylation (and
```

```
#   2/sec for phosphorylation), which is the same dephos rate I had picked to
fit our data.
#   Note: keep in mind that they are using 10uM of gefitinib and according to
Ref. 13 that concentration would only
#  inhibit 20-40% of the in vitro kinase activity of EGFR at Y1068, Y1148 and
Y1173.
kdephos1068 kdephos1068__FREE__ # 0.15

ratio_kdephos 1.0
kdephos1173 kdephos1068*ratio_kdephos

# Generic (pseudo first-order) phosphorylation rate constant
#     Value was set to fit experimental behavior observed by SiMPull

kphos1068 kphos1068__FREE__ # 0.043

ratio_kphos 1.0
kphos1173 kphos1068*ratio_kphos


end parameters

begin molecule types

# Ligand, growth factor
EGF(EGFL)

# Receptor tyrosine kinase, Epidermal growth factor receptor
#   I_III: domains I and III in the ectodomain for EGF binding
#     II: domain II for dimerization through ectodomain. Dimerization reaction
will change state of receptor
#     from monomer (unbound) to dimer (bound), and dissociation reaction the
opposite.
EGFR(I_III,II~u~b,Y1068~0~P,Y1173~0~P)

# Grb2 adaptor protein. SH3 represents both N- and C-terminus SH3 domains
GRB2(SH2)

#   SHC-Y317 in the p52 isoform | Y427 in the p66 isoform
SHC1(PTB)

end molecule types


begin seed species

EGF(EGFL) EGF_total
EGFR(I_III,II~u,Y1068~0,Y1173~0) EGFR_total
GRB2(SH2) GRB2_total
SHC1(PTB) SHC1_total


end seed species

begin observables
```

```
Molecules EGF EGF()
Molecules EGFRtot EGFR()
Molecules Grb2tot GRB2()
Molecules Shc1tot SHC1()
Molecules EGFR_EGF EGFR(I_III!+)
Molecules monR EGFR(II~u)
Molecules dimR EGFR(II~b)
Molecules pY1068 EGFR(Y1068~P!?)
Molecules pY1173 EGFR(Y1173~P!?)
Molecules Grb2_EGFR GRB2(SH2!+)
Molecules Shc1_EGFR SHC1(PTB!+)

Molecules pY1068_pY1173 EGFR(Y1068~P!?,Y1173~P!?)

Molecules monR_pYpY EGFR(II~u,Y1068~P!?,Y1173~P!?)
Molecules dimR_pYpY EGFR(II~b,Y1068~P!?,Y1173~P!?)



end observables


begin functions

pY1068_percent() 100*pY1068/EGFRtot
pY1173_percent() 100*pY1173/EGFRtot
pYpY_per() 100*pY1068_pY1173/EGFRtot
random_pYpY_per() 100*(pY1068/EGFRtot)*(pY1173/EGFRtot)

monR_pYpY_per() 100*monR_pYpY/(monR+1) # +1 to avoid dividing by 0
dimR_pYpY_per() 100*dimR_pYpY/(dimR+1) # +1 to avoid dividing by 0



end functions


begin reaction rules

# EGF reversibly binds EGFR
EGF(EGFL)+EGFR(I_III)<->EGF(EGFL!1).EGFR(I_III!1) kp_EGF,km_EGF


# Dimerization of EGFR for:
# Two EGF-bound receptors. Transition from monomer (II~u) to dimer (II~b)
states.
# Simplification: dimerization only happens between two EGF-bound receptors
EGFR(I_III!+,II~u)+EGFR(I_III!+,II~u)->EGFR(I_III!+,II~b)+EGFR(I_III!+,II~b)
kp_dim_L_L



# Dissociation of EGFR dimer.
```

```
# Simplification: dimer dissociation occurs equally regardless of how many
receptors are EGF-bound (remember EGF can dissociate while in a dimer)
EGFR(II~b)->EGFR(II~u) km_dim_L_L



# EGFR autophosphorylation
#   Occurs only within a dimer
EGFR(II~b,Y1068~0)->EGFR(II~b,Y1068~P) kphos1068
EGFR(II~b,Y1173~0)->EGFR(II~b,Y1173~P) kphos1173

# Unregulated dephosphorylation of pTyr sites
#   (mediated by constitutively active phosphatases)
EGFR(Y1068~P)->EGFR(Y1068~0) kdephos1068
EGFR(Y1173~P)->EGFR(Y1173~0) kdephos1173


# Binding of Grb2 to pY1068 in EGFR
GRB2(SH2)+EGFR(Y1068~P)<-> GRB2(SH2!1).EGFR(Y1068~P!1) kp_GE,km_GE


# Binding of SHC1 to pY1173 in EGFR
SHC1(PTB)+EGFR(Y1173~P)<-> SHC1(PTB!1).EGFR(Y1173~P!1) kp_SE,km_SE



end reaction rules

end model

begin actions

generate_network({overwrite=>1})

# Save parameters and concentrations before parameter scan
#saveParameters()
#saveConcentrations("pre_scan")

# Perform a parameter scan for EGF ligand concentrations
#parameter_scan({suffix=>"dose_resp",parameter=>"EGFconc",par_scan_vals=>[0.05
e-9,0.1e-9,0.5e-9,1.0e-9,2.5e-9,5.0e-9,10.0e-9,50.0e-9],\
#
method=>"ode",t_start=>0,t_end=>600,n_steps=>600,print_functions=>1})

parameter_scan({suffix=>"dose_resp",parameter=>"EGFconc",par_scan_vals=>[0.05e
-9,0.167e-9,0.5e-9,1.67e-9,5.0e-9,16.7e-9,50.0e-9],\

method=>"ode",t_start=>0,t_end=>120,n_steps=>3,print_functions=>1})



# Reset parameters and concentrations to those before parameter scan
#resetParameters()
#resetConcentrations("pre_scan")
```

```
# Equilibrate for 300 seconds
#simulate({suffix=>"equil",method=>"ode",t_start=>0,t_end=>300,n_steps=>300,pr
int_functions=>1})


# Add 25 nM EGF, and simulate for 300 seconds
setParameter("EGFconc","25.0e-9")
setConcentration("EGF(EGFL)","EGF_total")
simulate({suffix=>"EGF_25nM",method=>"ode",t_start=>0,t_end=>300,n_steps=>300,
print_functions=>1})

end actions
```

## BNGL file for model in Chapter 5

```
begin model


# References
# 1. Hause et al., 2012. Plos ONE.
# 2. Engelmann BW et al., 2014. Mol Cell Proteomics.
# 3. Kulak NA et al. (2014) Nat Methods 11: 319-324.
# 4. Shankaran H. et al., 2012. Molecular BioSystems.
# for plasma membrane the ref. is Hendriks 2003 (Cancer Research) and for
endosomes (pH 6.0) from French 1995
# 5. Kholodenko B.N. et al., 1999. The Journal of Biological Chemistry
# 6. Blinov M.L. et al., 2006. BioSystems
# 7. Chook, Yuh Min, et al. "The Grb2-mSos1 complex binds phosphopeptides with
higher affinity than
#    Grb2." Journal of Biological Chemistry 271.48 (1996): 30472-30478.
# 8. Macdonald JL, Pike LJ (2008) Proc Natl Acad Sci USA 105: 112-117.
# 9. Macdonald-Obermann JL, Pike LJ (2009) J Biol Chem 284: 13570-13576.
# 10. Elleman TC et al. (2001) Biochemistry 40: 8930-8939.
# 11. Low-Nam ST et al. (2011) Nat Struct Mol Biol 18: 1244-1249.
# 12. Kleiman LB et al. (2011) Mol Cell 43: 723Ð737.
# 13. Kim Y et al. (2012) Biochemistry 51 (25). American Chemical Society:
5212–22.
# 14. Morimatsu M et al. (2007) PNAS  104 (46): 18013–18.
# 15. Tujin Shi, Mario Niepel,.., Peter K. Sorger, Wei-Jun Qian, H. Steven
Wiley.
#     Conservation of Protein Abundance Patterns Reveals the Regulatory
Architecture of the EGFR-MAPK Pathway.
#       Work in preparation to be submitted to Science Signaling.
# 16. Sun Q et al. (2010) PloS one 5, no. 9 (2010): e12819
# 17. Reddy, Raven J., et al. "Early signaling dynamics of the epidermal
growth
#     factor receptor." Proceedings of the National Academy of Sciences 113.11
(2016): 3114-3119.

begin parameters


alpha_Ub__FREE__ 1 # Vary between 20 and 100 for low Ub, or 1 and 4 for high
Ub
alpha_pY1045__FREE__ 1 # Vary between 1 and 100
kub__FREE__ 5.25555397e-02 # Vary between 0.001 and 0.1
kdeub__FREE__ 1.71408252e-02 # Vary between 0.001 and 0.1
Kd_CE__FREE__ 2.51475699e-07 # Vary between 0.1e-6 and 10.0e-6 M (0.1 uM to 10
uM)
Kd_CG__FREE__ 7.23610301e-07 # Vary between 0.1e-6 and 10.0e-6 M (0.1 uM to 10
uM)
kc__FREE__ 2.59365139e+08 # Vary between 5E5 5E8
```

```
# Keep constant

alpha_CE__ 1 # Vary between 1 and 2
alpha_pYpY__ 1 # Vary between 1 and 100


CBL_total__ 5e3 # 5,000 copies per cell (from Capuani 2015)
kon__ 5.0e6 # Assumed to be 5.0e6 /M/s
kon_EGF__ 0.8e7 # # This rate was set so EGFR phosphorylation kinetics occurs
similarly as observed by Reddy et. al (2016)
Kd_EGF__ 1.0e-9 # A typical value of 1 nM for the EGF dissociation constant
was used
Kd_GE__ 0.6e-6 # 600 nM as estimated by Morimatsu et al. (2007)
kdephos1068__ 1.83276225e+00 # Estimated by fitting to SiMPull data
kphos1068__ 6.23793050e-01 # Estimated by fitting to SiMPull data
ratio_kdephos__ 1.0 # Equal phosphoryaltion rates for pY1045 and pY1068 were
assumed
ratio_kphos__ 1.0 # Equal phosphoryaltion rates for pY1045 and pY1068 were
assumed



# Scaling factors for BioNetFit
alpha_Ub alpha_Ub__FREE__

alpha_CE alpha_CE__
alpha_pY1045 alpha_pY1045__FREE__
alpha_pYpY alpha_pYpY__


# Avogadro constant
NA 6.02214e23 # [=] molecules per mol

# Fraction of cell to consider in a stochastic simulation
f 1 # [=] dimensionless, 0<=f<=1


# Cytoplasmic volume
#   A volume of 1 to 2 pL is typical for a mammalian cell.
Vc f*1.0e-12 # [=] L (1.0 pL)

# Number of cells per dish
numCells 1.0e7 # [=] cells per 60 mm^2 dish (10 million)

# Volume of media per dish
volMedia 1.0e-2 # [=] L (10 mL)

# Volume of extracellular fluid surrounding a cell
Vextra=f*volMedia/numCells


GRB2_total   1.0e6*f # [=] molecules per cell
EGFR_total 2.5e5*f # [=] molecules per cell (3e5 from Capuani 2015)
# Abundance of Cbl in several cell lines have been estimated to be 9,000-
15,000 copies/cell (Refs. 3 and 15)
```

```
CBL_total    CBL_total__*f*1 # [=] molecules per cell. (5.2e3 from Capuani
2015)



# Concentration of EGF
EGFconc 0 # [=] M
# EGFconc 25.0e-9 # [=] M

EGF_total=EGFconc*(NA*Vextra) # [=] molecules per cell



# A typical association rate constant for a protein-protein interaction
# A value of 1e7/M/s instead of 1e6/M/s (used originally) is closer to the on
rate estimated in Ref. 14 , which is ~1e7-1e8/M/s
# This value also allows off rates to be more similar to those estimated in
Ref. 14.
kon kon__ # [=] /M/s

kon_EGF kon_EGF__ # [=] /M/s

# Dissocation and association rate constants for EGF-EGFR interaction at the
plasma membrane
Kd_EGF Kd_EGF__*(NA*Vextra) # [=] molecules . 1.0e-9nM
kp_EGF=kon_EGF/(NA*Vextra)
km_EGF=Kd_EGF*kp_EGF


# Dissociation constant for EGFR dimerization
#   This parameter is set so that EGFR_total/KD_dim >> 1 when number of
receptors is high (e.g. 6.0e5)
KD_dim 6.0e5/10 # [=] molecules per cell


# Dissocation and association rate constants for interaction between two
liganded (EGF-bound) receptors
# '_pre' because off rates will be modified by a factor of 'offrate_f'. If
offrate_f=1 then
# 'km_dim_L_L_pre' and 'km_dim_L_L' are the same.
km_dim_L_L_pre 0.273 # [=] /s (Ref 11)
kp_dim_L_L=km_dim_L_L_pre/KD_dim # [=] /(molecule/cell)/s


# Increase off rates by a factor of 'offrate_f'
offrate_f 1.0
km_dim_L_L=km_dim_L_L_pre*offrate_f



# Kd for Grb2-SH2 domain binding to pY1068 EGFR
Kd_GE Kd_GE__*(NA*Vc) # [=] molecules   , (2.6e-6*(NA*Vc)) from Ref. 1
kp_GE=kon/(NA*Vc)
km_GE=Kd_GE*kp_GE
```

```
# Kd for Cbl-TKB domain binding to pY1045 EGFR
Kd_CE=Kd_CE__FREE__*(NA*Vc) # [=] molecules , (1.0e-6*(NA*Vc)) from Ref. 16
kp_CE=kon/(NA*Vc)
km_CE=Kd_CE*kp_CE

# Kd for Cbl-PR region constitutive binding to the SH3 domains of Grb2
Kd_CG=Kd_CG__FREE__*(NA*Vc) # [=] molecules , from Ref. 16
kp_CG=kon/(NA*Vc)
km_CG=Kd_CG*kp_CG

# Cooperativity constant. In this case, association rates will be multiplied
by this constant to account for the
# increased local concentration, which in our context is the result of being
bound to the same molecular complex
kc kc__FREE__

# Association constants for reactions with cooperativity
# Grb2-SH2 domain binding to pY1068 or pY1086 EGFR when Grb2 is complexed with
a Cbl molecule bound to EGFR
kp_GE_c=kp_GE*kc
# Cbl-TKB domain binding to pY1045 EGFR when Cbl is complexed with a Grb2
molecule bound to EGFR (at pY1068 or pY1086)
kp_CE_c=kp_CE*kc
# Cbl-PR region binding to the SH3 domains of Grb2 when both molecules are
bound to the same receptor (Grb2 bound to pY1068 or pY1086)
kp_CG_c=kp_CG*kc



# Generic (pseudo first-order) dephosphorylation rate constant
#      From Ref. 12 we have that dephosphorylation rate of pY-EGFR on cells
after using gefitinib is
#      0.05/s, which corresponds to a half-life of 15 s and represents a lower
#      bound (estimation without considering protection of sites by binding
proteins and other influencing factors)
#   When considering Shc binding in the paper they come with a model (M3) with
1/sec rate for dephosphorylation (and
#   2/sec for phosphorylation), which is the same dephos rate I had picked to
fit our data.
#   Note: keep in mind that they are using 10uM of gefitinib and according to
Ref. 13 that concentration would only
#  inhibit 20-40% of the in vitro kinase activity of EGFR at Y1068, Y1148 and
Y1173.
kdephos1068 kdephos1068__ # 0.15

ratio_kdephos ratio_kdephos__
kdephos1045 kdephos1068*ratio_kdephos

# Generic (pseudo first-order) phosphorylation rate constant
#      Value was set to fit experimental behavior observed by SiMPull
```

```
kphos1068 kphos1068__ # 0.043

ratio_kphos ratio_kphos__
kphos1045 kphos1068*ratio_kphos


# Rate of ubiquitination and de-ubiquitination
kub kub__FREE__ # 0.03
kdeub kdeub__FREE__ # 0.01

maxValue 20 # Constant value to set max value of Y display in graph

end parameters

begin molecule types

# Ligand, growth factor
EGF(EGFL)

# Receptor tyrosine kinase, Epidermal growth factor receptor
#   I_III: domains I and III in the ectodomain for EGF binding
#      II: domain II for dimerization through ectodomain. Dimerization reaction
will change state of receptor
#      from monomer (unbound) to dimer (bound), and dissociation reaction the
opposite.
#      Lys_ub: Number of ubiquitin molecules covalently linked to lysines in
EGFR
EGFR(I_III,II~u~b,Y1045~0~P,Y1068~0~P,Lys_ub~0~1)

# Grb2 adaptor protein. SH3 represents both N- and C-terminus SH3 domains
GRB2(SH2,SH3)


# CBL E3 ubiquitin-protein ligase. Tyrosine-kinase binding motif (TKB)
contains a SH2-like domain near N-terminus.
# Poline-rich (PR) region closer to the C-terminus
CBL(TKB,PR)


end molecule types


begin seed species

EGF(EGFL) EGF_total
EGFR(I_III,II~u,Y1045~0,Y1068~0,Lys_ub~0) EGFR_total
GRB2(SH2,SH3) GRB2_total
CBL(TKB,PR) CBL_total


end seed species

begin observables
```

```
Molecules EGF EGF()
Molecules EGFRtot EGFR()
Molecules Grb2tot GRB2()
Molecules Cbltot CBL()
Molecules Cbl_EGFRpY1045 CBL(TKB!1).EGFR(Y1045~P!1)
Molecules Cbl_notEGFRpY1045 CBL(TKB)
Molecules EGFR_EGF EGFR(I_III!+)
Molecules monR EGFR(II~u)
Molecules dimR EGFR(II~b)
Molecules pY1045 EGFR(Y1045~P!?)
Molecules pY1068 EGFR(Y1068~P!?)
Molecules Grb2_EGFR GRB2(SH2!+)

Molecules Cbl_Grb2_comp CBL(PR!1).GRB2(SH3!1)
Molecules pY1045_pY1068 EGFR(Y1045~P!?,Y1068~P!?)

Molecules monR_pYpY EGFR(II~u,Y1045~P!?,Y1068~P!?)
Molecules dimR_pYpY EGFR(II~b,Y1045~P!?,Y1068~P!?)

Molecules EGFR_ub EGFR(Lys_ub~1)

#Molecules EGFRpY1068_Grb2_Cbl_loop
GRB2(SH2!3,SH3!1).CBL(PR!1,TKB!2).EGFR(Y1045~P!2,Y1068~P!3)
#Molecules EGFRpY1068_Grb2_Cbl_open1
GRB2(SH2,SH3!1).CBL(PR!1,TKB!2).EGFR(Y1045~P!2,Y1068~P)
#Molecules EGFRpY1068_Grb2_Cbl_open2
CBL(TKB,PR!1).GRB2(SH3!1,SH2!2).EGFR(Y1068~P!2,Y1045~P)
#Molecules EGFRpY1068_Grb2_Cbl_open3
CBL(PR,TKB!1).EGFR(Y1045~P!1,Y1068~P!2).GRB2(SH2!2,SH3)


end observables


begin functions

pY1045_percent() 100*pY1045/EGFRtot
pY1068_percent() 100*pY1068/EGFRtot
pYpY_per() 100*pY1045_pY1068/EGFRtot
random_pYpY_per() 100*(pY1068/EGFRtot)*(pY1045/EGFRtot)
Cbl_EGFR_per() 100*Cbl_EGFRpY1045/Cbltot

monR_pYpY_per() 100*monR_pYpY/(monR+1) # +1 to avoid dividing by 0
dimR_pYpY_per() 100*dimR_pYpY/(dimR+1) # +1 to avoid dividing by 0

BNF_Cbl_EGFR_per() alpha_CE*100*Cbl_EGFRpY1045/Cbltot
BNF_pY1045_per() alpha_pY1045*100*pY1045/EGFRtot
BNF_pYpY_per() alpha_pYpY*100*pY1045_pY1068/EGFRtot


# Percent of receptors ubiquitinated
Ubiq_EGFR_per() 100*EGFR_ub/EGFRtot
BNF_Ub1_EGFR_per() alpha_Ub*100*EGFR_ub/EGFRtot # Ub1 is for WT
BNF_Ub2_EGFR_per() alpha_Ub*2.5*100*EGFR_ub/EGFRtot # Ub2 is for Y1045+ mutant
(no-cooperativity) (multiply by 2.5 because max should be 40)
```

```
end functions


begin reaction rules

# EGF reversibly binds EGFR
EGF(EGFL)+EGFR(I_III)<->EGF(EGFL!1).EGFR(I_III!1) kp_EGF,km_EGF


# Dimerization of EGFR for:
# Two EGF-bound receptors. Transition from monomer (II~u) to dimer (II~b)
states.
# Simplification: dimerization only happens between two EGF-bound receptors
EGFR(I_III!+,II~u)+EGFR(I_III!+,II~u)->EGFR(I_III!+,II~b)+EGFR(I_III!+,II~b)
kp_dim_L_L



# Dissociation of EGFR dimer.
# Simplification: dimer dissociation occurs equally regardless of how many
receptors are EGF-bound (remember EGF can dissociate while in a dimer)
EGFR(II~b)->EGFR(II~u) km_dim_L_L



# EGFR autophosphorylation
#   Occurs only within a dimer
EGFR(II~b,Y1045~0)->EGFR(II~b,Y1045~P) kphos1045
EGFR(II~b,Y1068~0)->EGFR(II~b,Y1068~P) kphos1068

# Unregulated dephosphorylation of pTyr sites
#   (mediated by constitutively active phosphatases)
EGFR(Y1045~P)->EGFR(Y1045~0) kdephos1045
EGFR(Y1068~P)->EGFR(Y1068~0) kdephos1068


# Binding of Grb2 to pY1068 in EGFR when it is free in solution, either as
Grb2 or Cbl-Grb2 complex.
# In the model Grb2 cannot bind to both pY1068 and pY1086 simultaneously
# You need to specify each molecular complex specifically to avoid duplicate
rules
GRB2(SH2,SH3)+EGFR(Y1068~P)<-> GRB2(SH2!1,SH3).EGFR(Y1068~P!1) kp_GE,km_GE
GRB2(SH2,SH3!1).CBL(PR!1,TKB)+EGFR(Y1068~P)<->
GRB2(SH2!2,SH3!1).CBL(PR!1,TKB).EGFR(Y1068~P!2) kp_GE,km_GE



# Binding of CBL to pY1045 in EGFR when it is free in solution, either as Cbl
or Cbl-Grb2 complex
CBL(TKB,PR)+EGFR(Y1045~P) <-> CBL(TKB!1,PR).EGFR(Y1045~P!1) kp_CE,km_CE
CBL(TKB,PR!1).GRB2(SH2,SH3!1)+EGFR(Y1045~P) <->
CBL(TKB!2,PR!1).GRB2(SH2,SH3!1).EGFR(Y1045~P!2) kp_CE,km_CE
```

```
# Constitutive association of CBL proline-rich region to the SH3 domains of
Grb2
# When both CBL and GRB2 are free in solution
CBL(TKB,PR)+GRB2(SH2,SH3) <-> CBL(TKB,PR!1).GRB2(SH2,SH3!1) kp_CG,km_CG
# When either one of them is bound to pYEGFR and the other one free in
solution
CBL(TKB!+,PR)+GRB2(SH2,SH3) <-> CBL(TKB!+,PR!1).GRB2(SH2,SH3!1) kp_CG,km_CG
CBL(TKB,PR)+GRB2(SH2!+,SH3) <-> CBL(TKB,PR!1).GRB2(SH2!+,SH3!1) kp_CG,km_CG


# Association reactions having cooperativity. The following reactions occur
between molecules bound
# to the same molecular complexes, this increases the local concentration of
the reactants and therefore
# the association rates

# Cooperative binding of Grb2 to pY1068 and pY1086 in EGFR
# In the model Grb2 cannot bind to both pY1068 and pY1086 simultaneously
GRB2(SH2,SH3!1).CBL(PR!1,TKB!2).EGFR(Y1045~P!2,Y1068~P) <->
GRB2(SH2!3,SH3!1).CBL(PR!1,TKB!2).EGFR(Y1045~P!2,Y1068~P!3) kp_GE_c,km_GE

# Cooperative binding of CBL to pY1045 in EGFR
CBL(TKB,PR!1).GRB2(SH3!1,SH2!2).EGFR(Y1068~P!2,Y1045~P) <->
CBL(TKB!3,PR!1).GRB2(SH3!1,SH2!2).EGFR(Y1068~P!2,Y1045~P!3) kp_CE_c,km_CE

# Cooperative binding of CBL proline-rich region to the SH3 domains of Grb2
when both molecules are bount to EGFR
CBL(PR,TKB!1).EGFR(Y1045~P!1,Y1068~P!2).GRB2(SH2!2,SH3) <->
CBL(PR!3,TKB!1).EGFR(Y1045~P!1,Y1068~P!2).GRB2(SH2!2,SH3!3) kp_CG_c,km_CG



# Ubiquitination of EGFR when Cbl is bound to the receptor (directly or
through Grb2 bound to pY1068 or pY1086)
# Direct binding
EGFR(Y1045~P!+,Lys_ub~0)->EGFR(Y1045~P!+,Lys_ub~1) kub
# Indirect binding (pY1045 should be free and CBL should have its TKB free)
EGFR(Y1045,Y1068~P!1,Lys_ub~0).GRB2(SH2!1,SH3!2).CBL(TKB,PR!2)-
>EGFR(Y1045,Y1068~P!1,Lys_ub~1).GRB2(SH2!1,SH3!2).CBL(TKB,PR!2) kub

# De-ubiquitination of EGFR
EGFR(Lys_ub~1)->EGFR(Lys_ub~0) kdeub



end reaction rules

end model

begin actions

generate_network({overwrite=>1})

# Save parameters and concentrations before parameter scan
#saveParameters()
```

```
#saveConcentrations("pre_scan")


parameter_scan({suffix=>"dose_resp",parameter=>"EGFconc",par_scan_vals=>[0.05e
-9,0.167e-9,0.5e-9,1.67e-9,5.0e-9,16.7e-9,50.0e-9],\

method=>"ode",t_start=>0,t_end=>120,n_steps=>3,print_functions=>1})

# Eliminate Grb2 binding and therefore cooperative recruitment of Cbl
setParameter("kp_GE",0.0);
parameter_scan({suffix=>"dose_resp_no_coop",parameter=>"EGFconc",par_scan_vals
=>[0.05e-9,0.167e-9,0.5e-9,1.67e-9,5.0e-9,16.7e-9,50.0e-9],\

method=>"ode",t_start=>0,t_end=>120,n_steps=>3,print_functions=>1})


# Reset parameters and concentrations to those before parameter scan
#resetParameters()
#resetConcentrations("pre_scan")

# Equilibrate for 300 seconds
#simulate({suffix=>"equil",method=>"ode",t_start=>0,t_end=>300,n_steps=>300,pr
int_functions=>1})


## Add 25 nM EGF, and simulate for 300 seconds
#setParameter("EGFconc","25.0e-9")
#setConcentration("EGF(EGFL)","EGF_total")
#simulate({suffix=>"EGF_25nM",method=>"ode",t_start=>0,t_end=>30,n_steps=>50,p
rint_functions=>1})

end actions
```

# APPENDIX C: Configuration files for parameter estimation (.conf)

## Configuration file for model in Chapter 4

```
##############
### PATHS ###
##############

# The directory to which job output will be written
output_dir=/home/esc1987/Modeling/BioNetFit_v1.1_with_Simulators/output/

# The BioNetGen executable
# bng_command=Simulators/BNG2.pl
bng_command=
/home/esc1987/Modeling/BioNetFit_v1.1_with_Simulators/Simulators/BNG2.pl

# ESC added the following two lines because NFsim was not able to run simulations
# NFsim for Cygwin
#nfsim_dir=/home/esc1987/Modeling/NFsim_v1.11/bin
#nfsim_command=NFsim_x86_64-cygwin
#nfsim_command=NFsim_i686-cygwin

# The model file to be used in fitting simulations
model=/home/esc1987/Modeling/BioNetFit_v1.1_with_Simulators/CHO_EGFR/fit
_v1_16/180222_CHO_EGFR.bngl


# The experimental data to be fit
exp_file=/home/esc1987/Modeling/BioNetFit_v1.1_with_Simulators/CHO_EGFR/
fit_v1_16/dose_resp.exp
exp_file=/home/esc1987/Modeling/BioNetFit_v1.1_with_Simulators/CHO_EGFR/
fit_v1_16/EGF_25nM.exp
```

```
########################
### General Options ###
########################

# The job name
job_name=fit_v1_16_2

# Whether or not to generate plots for best-fit outputs
make_plots=0

# Number of simulations to run in parallel.
# Change parellel_count to the number of CPU cores on your machine for increased performance.
parallel_count=5

# Kill a job and continue without it if  process runs longer than walltime. Adjust if needed.
max_walltime=10:00

# Delete files that are no longer needed to save disk space
# Property not working
delete_old_files=1

# Ask if you want to overwrite existing existing job output. In this case disable ask, that way you can do batch fitting


########################
### Fitting Options ###
########################
```

# Which objective function to minimize in fitting. A complete list of objective functions is described in GenFit documentation.

# 1:  sum-of-squares function (i.e. nonlinear least squares fitting).

# 2 chi-square function (i.e. weighted nonlinear least squares fitting)

objfunc=1

# Do not divide by initial value (at t=0) of simulation results

divide_by_init= 0

# The maximum number of generations to run.

max_generations=75

# The number of unique parameter sets simulated in first generation.

first_gen_permutations=150

# The number of unique parameter sets simulated in a generation.

permutations=75

# Do bootstrapping

bootstrap= 100

bootstrap_chi= 30

# The mutation probability and mutation factor for free parameters.

mutate=default 0.2 0.2

# The free parameters. These are generated on a random log scale between numbers indicated.

loguniform_var=GRB2_total__FREE__ 1E4 1E6

loguniform_var=SHC1_total__FREE__ 1E4 1E6

loguniform_var=kdephos1068__FREE__ 0.5 5

loguniform_var=kphos1068__FREE__ 0.5 5

**Configuration file for model in Chapter 5**

```
##############
### PATHS ###
##############

# The directory to which job output will be written
output_dir=/home/esc1987/Modeling/BioNetFit_v1.1_with_Simulators/output/

# The BioNetGen executable
# bng_command=Simulators/BNG2.pl
bng_command=
/home/esc1987/Modeling/BioNetFit_v1.1_with_Simulators/Simulators/BNG2.pl

# ESC added the following two lines because NFsim was not able to run
simulations
# NFsim for Cygwin
#nfsim_dir=/home/esc1987/Modeling/NFsim_v1.11/bin
#nfsim_command=NFsim_x86_64-cygwin
#nfsim_command=NFsim_i686-cygwin

# The model file to be used in fitting simulations
model=/home/esc1987/Modeling/BioNetFit_v1.1_with_Simulators/CHO_EGFR/fit
_v1_17/180301_HeLa_EGFR_Cbl_Ub.bngl


# The experimental data to be fit
exp_file=/home/esc1987/Modeling/BioNetFit_v1.1_with_Simulators/CHO_EGFR/
fit_v1_17/dose_resp.exp
exp_file=/home/esc1987/Modeling/BioNetFit_v1.1_with_Simulators/CHO_EGFR/
fit_v1_17/dose_resp_no_coop.exp
```

```
########################
### General Options ###
########################

# The job name
job_name=fit_v1_17

# Whether or not to generate plots for best-fit outputs
make_plots=0

# Number of simulations to run in parallel.
# Change parellel_count to the number of CPU cores on your machine for increased performance.
parallel_count=6

# Kill a job and continue without it if  process runs longer than walltime. Adjust if needed.
max_walltime=10:00

# Delete files that are no longer needed to save disk space
# Property not working
delete_old_files=1

# Ask if you want to overwrite existing existing job output. In this case disable ask, that way you can do batch fitting


########################
### Fitting Options ###
########################

# Which objective function to minimize in fitting. A complete list of objective functions is described in GenFit documentation.
# 1:  sum-of-squares function (i.e. nonlinear least squares fitting).
```

# 2 chi-square function (i.e. weighted nonlinear least squares fitting)
objfunc=1

# The maximum number of generations to run.
max_generations=75

# The number of unique parameter sets simulated in first generation.
first_gen_permutations=150

# The number of unique parameter sets simulated in a generation.
permutations=75

# Do bootstrapping
bootstrap= 6
bootstrap_chi= 35

# Do not divide by initial value (at t=0) of simulation results
divide_by_init= 0

# The mutation probability and mutation factor for free parameters.
mutate=default 0.2 0.2

# The free parameters.
# These are scaling factors generated on a random scale between numbers indicated.
#random_var=alpha_pY1045__FREE__ 4 20
#random_var=alpha_Ub__FREE__ 20 100
random_var=alpha_pY1045__FREE__ 1 100
random_var=alpha_Ub__FREE__ 6 60

# These are generated on a random log scale between numbers indicated.
loguniform_var=Kd_CE__FREE__ 0.1E-6 10.0E-6
loguniform_var=Kd_CG__FREE__ 0.1E-6 10.0E-6
loguniform_var=kc__FREE__ 5E5 5E8
loguniform_var=kub__FREE__ 0.001 0.1
loguniform_var=kdeub__FREE__ 0.001 0.1

# REFERENCES


Andrews, N.L., K.A. Lidke, J.R. Pfeiffer, A.R. Burns, B.S. Wilson, J.M. Oliver, and D.S. Lidke. 2008. Actin restricts FcεRI diffusion and facilitates antigen-induced receptor immobilization. *Nat. Cell Biol.* 10:955–963. doi:10.1038/ncb1755.

Aoki, K., M. Yamada, K. Kunida, S. Yasuda, and M. Matsuda. 2011. Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proc. Natl. Acad. Sci.* 108:12675–12680. doi:10.1073/pnas.1104030108.

Batzer, A.G., D. Rotin, J.M. Ureña, E.Y. Skolnik, and J. Schlessinger. 1994. Hierarchy of binding sites for Grb2 and Shc on the epidermal growth factor receptor. *Mol. Cell. Biol.* 14:5192–201. doi:10.1128/MCB.14.8.5192.Updated.

Björkelund, H., L. Gedda, and K. Andersson. 2011. Comparing the Epidermal Growth Factor Interaction with Four Different Cell Lines: Intriguing Effects Imply Strong Dependency of Cellular Context. *PLoS One*. 6:e16536. doi:10.1371/journal.pone.0016536.

Blinov, M.L., J.R. Faeder, B. Goldstein, and W.S. Hlavacek. 2006. A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *BioSystems*. 83:136–151. doi:10.1016/j.biosystems.2005.06.014.

Brock, R., I.H.L. Hamelers, and T.M. Jovin. 1999. Comparison of fixation protocols for adherent cultured cells applied to a GFP fusion protein of the

epidermal growth factor receptor. *Cytometry*. 35:353–362. doi:10.1002/(SICI)1097-0320(19990401)35:4<353::AID-CYTO8>3.0.CO;2-M.

Brunati, A.M., L.A. Pinna, E. Bergantino, M. Ruzzene, P. Cirri, G. Ramponi, and A. Donella-Deana. 1998. Src homology-2 domains protect phosphotyrosyl residues against enzymatic dephosphorylation. *Biochem. Biophys. Res. Commun.* 243:700–5. doi:10.1006/bbrc.1998.8153.

Brunner, A.M., P. Lössl, F. Liu, R. Huguet, C. Mullen, M. Yamashita, V. Zabrouskov, A. Makarov, A.F.M. Altelaar, and A.J.R. Heck. 2015. Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization. *Anal. Chem.* 87:4152–4158. doi:10.1021/acs.analchem.5b00162.

Capuani, F., A. Conte, E. Argenzio, L. Marchetti, C. Priami, S. Polo, P.P. Di Fiore, S. Sigismund, and A. Ciliberto. 2015. Quantitative analysis reveals how EGFR activation and downregulation are coupled in normal but not in cancer cells. *Nat. Commun.* 6:7999. doi:10.1038/ncomms8999.

Chen, W.W., B. Schoeberl, P.J. Jasper, M. Niepel, U.B. Nielsen, D.A. Lauffenburger, and P.K. Sorger. 2009. Input–output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol. Syst. Biol.* 5:239. doi:10.1038/msb.2008.74.

Chung, I., R. Akita, R. Vandlen, D. Toomre, J. Schlessinger, and I. Mellman. 2010. Spatial control of EGF receptor activation by reversible dimerization

on living cells. *Nature*. 464:783–787. doi:10.1038/nature08827.

Coba, M.P., A.J. Pocklington, M.O. Collins, M. V. Kopanitsa, R.T. Uren, S. Swamy, M.D.R. Croning, J.S. Choudhary, and S.G.N. Grant. 2009. Neurotransmitters Drive Combinatorial Multistate Postsynaptic Density Networks. *Sci. Signal.* 2:ra19. doi:10.1126/scisignal.2000102.

Creamer, M.S., E.C. Stites, M. Aziz, J.A. Cahill, C. Tan, M.E. Berens, H. Han, K.J. Bussey, D.D. Von Hoff, W.S. Hlavacek, and R.G. Posner. 2012. Specification, annotation, visualization and simulation of a large rule-based model for ERBB receptor signaling. *BMC Syst. Biol.* 6:107. doi:10.1186/1752-0509-6-107.

Curran, T.G., Y. Zhang, D.J. Ma, J.N. Sarkaria, and F.M. White. 2015. MARQUIS: A multiplex method for absolute quantification of peptides and posttranslational modifications. *Nat. Commun.* 6:5924. doi:10.1038/ncomms6924.

Ebner, R., and R. Derynck. 1991. Epidermal growth factor and transforming growth factor-alpha: differential intracellular routing and processing of ligand-receptor complexes. *Cell Regul.* 2:599–612. doi:10.1091/MBC.2.8.599.

Eden, E.R., I.J. White, A. Tsapara, and C.E. Futter. 2010. Membrane contacts between endosomes and ER provide sites for PTP1B–epidermal growth factor receptor interaction. *Nat. Cell Biol.* 12:267. doi:10.1038/ncb2026.

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

doi:10.1007/978-1-59745-525-1_5.

Feinerman, O., J. Veiga, J.R. Dorfman, R.N. Germain, and G. Altan-Bonnet. 2008. Variability and Robustness in T Cell Activation from Regulated Heterogeneity in Protein Levels. *Science.* 321:1081–1084. doi:10.1126/science.1158013.

Fortian, A., and A. Sorkin. 2014. Live-cell fluorescence imaging reveals high stoichiometry of Grb2 binding to the EGF receptor sustained during endocytosis. *J. Cell Sci.* 127:432–444. doi:10.1242/jcs.137786.

Francavilla, C., M. Papetti, K.T.G. Rigbolt, A.-K. Pedersen, J.O. Sigurdsson, G. Cazzamali, G. Karemore, B. Blagoev, and J. V Olsen. 2016. Multilayered proteomics reveals molecular switches dictating ligand-dependent EGFR trafficking. *Nat. Struct. Mol. Biol.* 23:608–618. doi:10.1038/nsmb.3218.

Freed, D.M., N.J. Bessman, A. Kiyatkin, E. Salazar-Cavazos, P.O. Byrne, J.O. Moore, C.C. Valley, K.M. Ferguson, D.J. Leahy, D.S. Lidke, and M.A. Lemmon. 2017. EGFR Ligands Differentially Stabilize Receptor Dimers to Specify Signaling Kinetics. *Cell*. 171:683–695. doi:10.1016/j.cell.2017.09.017.

Fujioka, A., K. Terai, R.E. Itoh, K. Aoki, T. Nakamura, S. Kuroda, E. Nishida, and M. Matsuda. 2006. Dynamics of the Ras/ERK MAPK cascade as monitored by fluorescent probes. *J. Biol. Chem.* 281:8917–26. doi:10.1074/jbc.M509344200.

George, N. 2006. A new method for protein labeling with small molecules based

on acyl carrier protein. EPFL.

Gibson, S.K., J.H. Parkes, and P.A. Liebman. 2000. Phosphorylation modulates the affinity of light-activated rhodopsin for G protein and arrestin. *Biochemistry*. 39:5738–5749. doi:10.1021/bi991857f.

Hause, R.J., K.K. Leung, J.L. Barkinge, M.F. Ciaccio, C. pin Chuu, and R.B. Jones. 2012. Comprehensive Binary Interaction Mapping of SH2 Domains via Fluorescence Polarization Reveals Novel Functional Diversification of ErbB Receptors. *PLoS One*. 7:e44471. doi:10.1371/journal.pone.0044471.

Hendriks, C.L.L., L.J. van Vliet, B. Rieger, G.M.P. van Kempen, and M. van Ginkel. 1999. DIPimage: a scientific image processing toolbox for MATLAB. *Quant. Imaging Group, Fac. Appl. Sci. Delft Univ. Technol. Delft, Netherlands*.

Huang, F., X. Zeng, W. Kim, M. Balasubramani, A. Fortian, S.P. Gygi, N.A. Yates, and A. Sorkin. 2013. Lysine 63-linked polyubiquitination is required for EGF receptor degradation. *Proc. Natl. Acad. Sci. U. S. A.* 110:15722–7. doi:10.1073/pnas.1308014110.

Jadwin, J.A., T.G. Curran, A.T. Lafontaine, F.M. White, and B.J. Mayer. 2018. Src homology 2 domains enhance tyrosine phosphorylation in vivo by protecting binding sites in their target proteins from dephosphorylation. *J. Biol. Chem.* 293:623–637. doi:10.1074/jbc.M117.794412.

Jain, A., R. Liu, B. Ramani, E. Arauz, Y. Ishitsuka, K. Ragunathan, J. Park, J. Chen, Y.K. Xiang, and T. Ha. 2011. Probing cellular protein complexes using

single-molecule pull-down. *Nature*. 473:484–488. doi:10.1038/nature10016.

Jain, A., R. Liu, Y.K. Xiang, and T. Ha. 2012. Single-molecule pull-down for studying protein interactions. *Nat. Protoc.* 7:445–452. doi:10.1038/nprot.2011.452.

Jallal, B., J. Schlessinger, and A. Ullrich. 1992. Tyrosine phosphatase inhibition permits analysis of signal transduction complexes in p1850_HER2/neu_-overexpressing human tumor cells. *J.Biol.Chem.* 267:4357–4363.

Kaushansky, A., A. Gordus, B. Chang, J. Rush, and G. Macbeath. 2008. A quantitative study of the recruitment potential of all intracellular tyrosine residues on EGFR, FGFR1 and IGF1R. *Mol. Biosyst.* 4:643–653. doi:10.1039/b801018h.

Kholodenko, B.N., O. V. Demin, G. Moehren, and J.B. Hoek. 1999. Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.* 274:30169–30181. doi:10.1074/jbc.274.42.30169.

Kim, K.L., D. Kim, S. Lee, S.-J. Kim, J.E. Noh, J.-H. Kim, Y.C. Chae, J.-B. Lee, and S.H. Ryu. 2016. Pairwise detection of site-specific receptor phosphorylations using single-molecule blotting. *Nat. Commun.* 7:11107. doi:10.1038/ncomms11107.

Kim, Y., Z. Li, M. Apetri, B. Luo, J.E. Settleman, and K.S. Anderson. 2012. Temporal resolution of autophosphorylation for normal and oncogenic forms of EGFR and differential effects of gefitinib. *Biochemistry*. 51:5212–5222. doi:10.1021/bi300476v.

Kleiman, L.B., T. Maiwald, H. Conzelmann, D.A. Lauffenburger, and P.K. Sorger. 2011. Rapid Phospho-Turnover by Receptor Tyrosine Kinases Impacts Downstream Signaling and Drug Binding. *Mol. Cell*. 43:723–737. doi:10.1016/j.molcel.2011.07.014.

Kozer, N., D. Barua, C. Henderson, E.C. Nice, A.W. Burgess, W.S. Hlavacek, and A.H.A. Clayton. 2014. Recruitment of the adaptor protein Grb2 to EGFR tetramers. *Biochemistry*. 53:2594–2604. doi:10.1021/bi500182x.

Kozer, N., D. Barua, S. Orchard, E.C. Nice, A.W. Burgess, W.S. Hlavacek, and A.H.A. Clayton. 2013a. Exploring higher-order EGFR oligomerisation and phosphorylation—a combined experimental and theoretical approach. *Mol. Biosyst.* 9:1849. doi:10.1039/c3mb70073a.

Kozer, N., D. Barua, S. Orchard, E.C. Nice, A.W. Burgess, W.S. Hlavacek, and A.H.A. Clayton. 2013b. Exploring higher-order EGFR oligomerisation and phosphorylation—a combined experimental and theoretical approach. *Mol. Biosyst.* 9:1849. doi:10.1039/c3mb70073a.

Kulak, N.A., G. Pichler, I. Paron, N. Nagaraj, and M. Mann. 2014. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods*. 11:319–324. doi:10.1038/nmeth.2834.

Labit, H., A. Goldar, G. Guilbaud, C. Douarche, O. Hyrien, and K. Marheineke. 2008. A simple and optimized method of producing silanized surfaces for FISH and replication mapping on combed DNA fibers. *Biotechniques*.

45:649–658. doi:10.2144/000113002.

Lahav, G., N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A.J. Levine, M.B. Elowitz, and U. Alon. 2004. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat. Genet.* 36:147–150. doi:10.1038/ng1293.

Lau, E.K., M. Trester-Zedlitz, J.C. Trinidad, S.J. Kotowski, A.N. Krutchinsky, A.L. Burlingame, and M. von Zastrow. 2011. Quantitative Encoding of the Effect of a Partial Agonist on Individual Opioid Receptors by Multisite Phosphorylation and Threshold Detection. *Sci. Signal.* 4:ra52. doi:10.1126/scisignal.2001748.

LeBlanc, D.C. 2004. Statistics: concepts and applications for science. Vol. 2. Jones & Bartlett Learning, Burlington. 175-176 pp.

Lemmon, M.A., and J. Schlessinger. 2010. Cell signaling by receptor tyrosine kinases. *Cell*. 141:1117–1134. doi:10.1016/j.cell.2010.06.011.

Levkowitz, G., H. Waterman, S.A. Ettenberg, M. Katz, A.Y. Tsygankov, I. Alroy, S. Lavi, K. Iwai, Y. Reiss, A. Ciechanover, S. Lipkowitz, and Y. Yarden. 1999. Ubiquitin Ligase Activity and Tyrosine Phosphorylation Underlie Suppression of Growth Factor Signaling by c-Cbl/Sli-1. *Mol. Cell*. 4:1029–1040. doi:10.1016/S1097-2765(00)80231-2.

Lidke, D.S., P. Nagy, R. Heintzmann, D.J. Arndt-Jovin, J.N. Post, H.E. Grecco, E.A. Jares-Erijman, and T.M. Jovin. 2004. Quantum dot ligands provide new insights into erbB/HER receptor–mediated signal transduction. *Nat. Biotechnol.* 22:198–203. doi:10.1038/nbt929.

Low-Nam, S.T., K.A. Lidke, P.J. Cutler, R.C. Roovers, P.M.P. van Bergen en Henegouwen, B.S. Wilson, and D.S. Lidke. 2011. ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding. *Nat. Struct. Mol. Biol.* 18:1244–1249. doi:10.1038/nsmb.2135.

Morimatsu, M., Takagi, H., Ota, K. G., Iwamoto, R., Yanagida, T., & Sako, Y. 2007. Multiple-state reactions between the epidermal growth factor receptor and Grb2 as observed by using single-molecule analysis. *Proc. Natl. Acad. Sci.* 104:18013–18018.

Munsky, B., G. Neuert, and A. van Oudenaarden. 2012. Using Gene Expression Noise to Understand Gene Regulation. *Science (80-. ).* 336:183–187. doi:10.1126/science.1216379.

Oh, D., M. Ogiue-Ikeda, J.A. Jadwin, K. Machida, B.J. Mayer, and J. Yu. 2012. Fast rebinding increases dwell time of Src homology 2 (SH2)-containing proteins near the plasma membrane. *Proc. Natl. Acad. Sci.* 109:14024–14029. doi:10.1073/pnas.1203397109.

Olayioye, M. a, R.M. Neve, H. a Lane, and N.E. Hynes. 2000. The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J.* 19:3159–3167. doi:10.1093/emboj/19.13.3159.

Pinilla-Macua, I., S.C. Watkins, and A. Sorkin. 2016. Endocytosis separates EGF receptors from endogenous fluorescently labeled HRas and diminishes receptor signaling to MAP kinases in endosomes. *Proc. Natl. Acad. Sci. U. S. A.* 113:2122–7. doi:10.1073/pnas.1520301113.

Raiborg, C., and H. Stenmark. 2009. The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature*. 458:445–452. doi:10.1038/nature07961.

Reddy, R.J., A.S. Gajadhar, E.J. Swenson, D.A. Rothenberg, T.G. Curran, and F.M. White. 2016. Early signaling dynamics of the epidermal growth factor receptor. *Proc. Natl. Acad. Sci.* 113:3114–3119. doi:10.1073/pnas.1521288113.

Roepstorff, K., M.V. Grandal, L. Henriksen, S.L.J. Knudsen, M. Lerdrup, L. Grøvdal, B.M. Willumsen, and B. Van Deurs. 2009. Differential effects of EGFR ligands on endocytic sorting of the receptor. *Traffic*. 10:1115–1127. doi:10.1111/j.1600-0854.2009.00943.x.

Rotin, D., B. Margolis, M. Mohammadi, R.J. Daly, G. Daum, N. Li, E.H. Fischer, W.H. Burgess, A. Ullrich, and J. Schlessinger. 1992. SH2 domains prevent tyrosine dephosphorylation of the EGF receptor: identification of Tyr992 as the high-affinity binding site for SH2 domains of phospholipase C gamma. *EMBO J.* 11:559–67.

Salazar, C., and T. Höfer. 2009. Multisite protein phosphorylation - From molecular mechanisms to kinetic models. *FEBS J.* 276:3177–3198. doi:10.1111/j.1742-4658.2009.07027.x.

Schlessinger, J. 2002. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell*. 110:669–672. doi:10.1016/S0092-8674(02)00966-2.

Schulze, W.X., L. Deng, and M. Mann. 2005. Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol. Syst. Biol.* 1:2005.0008. doi:10.1038/msb4100012.

Schwartz, S.L., C. Cao, O. Pylypenko, A. Rak, and A. Wandinger-Ness. 2007. Rab GTPases at a glance. *J. Cell Sci.* 120:3905–10. doi:10.1242/jcs.015909.

Schwartz, S.L., C. Cleyrat, M.J. Olah, P.K. Relich, G.K. Phillips, W.S. Hlavacek, K.A. Lidke, B.S. Wilson, and D.S. Lidke. 2017. Differential Mast Cell Outcomes Are Sensitive to FcεRI-Syk Binding Kinetics. *Mol. Biol. Cell*. 28:3397–3414. doi:10.1091/mbc.E17-06-0350.

Shan, Y., M.P. Eastwood, X. Zhang, E.T. Kim, A. Arkhipov, R.O. Dror, J. Jumper, J. Kuriyan, and D.E. Shaw. 2012. Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell*. 149:860–870. doi:10.1016/j.cell.2012.02.063.

Shankaran, H., Y. Zhang, W.B. Chrisler, J.A. Ewald, H.S. Wiley, and H. Resat. 2012. Integrated experimental and model-based analysis reveals the spatial aspects of EGFR activation dynamics. *Mol. Biosyst.* 8:2868. doi:10.1039/c2mb25190f.

Shi, T., M. Niepel, J.E. McDermott, Y. Gao, C.D. Nicora, W.B. Chrisler, L.M. Markillie, V.A. Petyuk, R.D. Smith, K.D. Rodland, P.K. Sorger, W.-J. Qian, and H.S. Wiley. 2016. Conservation of protein abundance patterns reveals the regulatory architecture of the EGFR-MAPK pathway. *Sci. Signal.* 9:1–14.

doi:10.1126/scisignal.aaf0891.

Shtiegman, K., B.S. Kochupurakkal, Y. Zwang, G. Pines, A. Starr, A. Vexler, A. Citri, M. Katz, S. Lavi, Y. Ben-Basat, S. Benjamin, S. Corso, J. Gan, R.B. Yosef, S. Giordano, and Y. Yarden. 2007. Defective ubiquitinylation of EGFR mutants of lung cancer confers prolonged signaling. *Oncogene*. 26:6968–6978. doi:10.1038/sj.onc.1210503.

Sigismund, S., V. Algisi, G. Nappo, A. Conte, R. Pascolutti, A. Cuomo, T. Bonaldi, E. Argenzio, L.G.G.C. Verhoef, E. Maspero, F. Bianchi, F. Capuani, A. Ciliberto, S. Polo, and P.P. Di Fiore. 2013. Threshold-controlled ubiquitination of the EGFR directs receptor fate. *EMBO J.* 32:2140–2157. doi:10.1038/emboj.2013.149.

Sigismund, S., E. Argenzio, D. Tosoni, E. Cavallaro, S. Polo, and P.P. Di Fiore. 2008. Clathrin-Mediated Internalization Is Essential for Sustained EGFR Signaling but Dispensable for Degradation. *Dev. Cell*. 15:209–219. doi:10.1016/j.devcel.2008.06.012.

Sigismund, S., T. Woelk, C. Puri, E. Maspero, C. Tacchetti, P. Transidico, P.P. Di Fiore, and S. Polo. 2005. Clathrin-independent endocytosis of ubiquitinated cargos. *Proc. Natl. Acad. Sci.* 102:2760–2765. doi:10.1073/pnas.0409817102.

Smith, A.M., W. Xu, Y. Sun, J.R. Faeder, and G.E. Marai. 2012. RuleBender: integrated modeling, simulation and visualization for rule-based intracellular biochemistry. *BMC Bioinformatics*. 13:S3. doi:10.1186/1471-2105-13-S8-S3.

Smith, C.S., N. Joseph, B. Rieger, and K.A. Lidke. 2010. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat. Methods*. 7:373–375. doi:10.1038/nmeth.1449.

Sorkin, A., and L.K. Goh. 2009. Endocytosis and intracellular trafficking of ErbBs. *Exp. Cell Res.* 315:683–696. doi:10.1016/J.YEXCR.2008.07.029.

Spencer, S.L., S. Gaudet, J.G. Albeck, J.M. Burke, and P.K. Sorger. 2009. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*. 459:428–432. doi:10.1038/nature08012.

Steinkamp, M.P., S.T. Low-Nam, S. Yang, K.A. Lidke, D.S. Lidke, and B.S. Wilson. 2014. erbB3 Is an Active Tyrosine Kinase Capable of Homo- and Heterointeractions. *Mol. Cell. Biol.* 34:965–977. doi:10.1128/MCB.01605-13.

Swaney, D.L., C.D. Wenger, and J.J. Coon. 2010. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J. Proteome Res.* 9:1323–1329. doi:10.1021/pr900863u.

Tarcic, G., S.K. Boguslavsky, J. Wakim, T. Kiuchi, A. Liu, F. Reinitz, D. Nathanson, T. Takahashi, P.S. Mischel, T. Ng, and Y. Yarden. 2009. An Unbiased Screen Identifies DEP-1 Tumor Suppressor as a Phosphatase Controlling EGFR Endocytosis. *Curr. Biol.* 19:1788–1798. doi:10.1016/j.cub.2009.09.048.

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

32:798–800. doi:10.1093/bioinformatics/btv655.

Umebayashi, K., H. Stenmark, and T. Yoshimori. 2008. Ubc4/5 and c-Cbl continue to ubiquitinate EGF receptor after internalization to facilitate polyubiquitination and degradation. *Mol. Biol. Cell*. 19:3454–62. doi:10.1091/mbc.E07-10-0988.

Valley, C.C., D.J. Arndt-Jovin, N. Karedla, M.P. Steinkamp, A.I. Chizhik, W.S. Hlavacek, B.S. Wilson, K.A. Lidke, and D.S. Lidke. 2015. Enhanced dimerization drives ligand-independent activity of mutant epidermal growth factor receptor in lung cancer. *Mol. Biol. Cell*. 26:4087–4099. doi:10.1091/mbc.E15-05-0269.

Waterman, H. 2002. A mutant EGF-receptor defective in ubiquitylation and endocytosis unveils a role for Grb2 in negative signaling. *EMBO J.* 21:303–313. doi:10.1093/emboj/21.3.303.

Wilson, K.J., C. Mill, S. Lambert, J. Buchman, T.R. Wilson, V. Hernandez-Gordillo, R.M. Gallo, L.M.C. Ades, J. Settleman, and D.J. Riese. 2012. EGFR ligands exhibit functional differences in models of paracrine and autocrine signaling. *Growth Factors*. 30:107–16. doi:10.3109/08977194.2011.649918.

Wolf-Yadlin, A., N. Kumar, Y. Zhang, S. Hautaniemi, M. Zaman, H.-D. Kim, V. Grantcharova, D. a Lauffenburger, and F.M. White. 2006. Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol. Syst. Biol.* 2:54. doi:10.1038/msb4100094.

Yao, Z., K. Darowski, N. St-Denis, V. Wong, F. Offensperger, A. Villedieu, S. Amin, R. Malty, H. Aoki, H. Guo, Y. Xu, C. Iorio, M. Kotlyar, A. Emili, I. Jurisica, B.G. Neel, M. Babu, A.-C. Gingras, and I. Stagljar. 2017. A Global Analysis of the Receptor Tyrosine Kinase-Protein Phosphatase Interactome. *Mol. Cell*. 65:347–360. doi:10.1016/j.molcel.2016.12.004.

Yarden, Y., and M.X. Sliwkowski. 2001. Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* 2:127–137. doi:10.1038/35052073.

Yudushkin, I.A., A. Schleifenbaum, A. Kinkhabwala, B.G. Neel, C. Schultz, and P.I.H. Bastiaens. 2007. Live-cell imaging of enzyme-substrate interaction reveals spatial regulation of PTP1B. *Science*. 315:115–9. doi:10.1126/science.1134966.

Zhang, X., J. Gureasko, K. Shen, P.A. Cole, and J. Kuriyan. 2006. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell*. 125:1137–1149. doi:10.1016/j.cell.2006.05.013.

Ziomkiewicz, I., A. Loman, R. Klement, C. Fritsch, A.S. Klymchenko, G. Bunt, T.M. Jovin, and D.J. Arndt-Jovin. 2013. Dynamic conformational transitions of the EGF receptor in living mammalian cells determined by FRET and fluorescence lifetime imaging microscopy. *Cytom. Part A*. 83:794–805. doi:10.1002/cyto.a.22311.