A Modeling Architecture for Data Analytics in Systems Biology : Simulating the ErbB Signaling Pathway

Thesis submitted to

Cochin University of Science and Technology

For the award of degree of

Doctor of Philosophy

in the Faculty of Technology

By

Arya A. Das

(Reg No. 4599)

Computational Modeling and Simulation Group

Environmental Technology Division

CSIR-National Institute for interdisciplinary Science and technology

Thiruvananthapuram 695019, India

November 2017

A Modeling Architecture for Data Analytics in Systems Biology : Simulating the ErbB Signaling Pathway

Ph. D. Thesis
November 2017

Author:

Arya A. Das
Senior Research Fellow (SRF)
Computational Modeling & Simulation Group
Environmental Technology Division
Phone: +91 471 2515381, Fax: +91 471 2491712
Email: aryaanildas@gmail.com

Thesis Supervisor:

Dr. Elizabeth Jacob
Chief Scientist
Computational Modeling & Simulation Group
Environmental Technology Division
Phone: +91 471 2515381, Fax: +91 471 2491712
Email: liz.csir@gmail.com

DECLARATION

I hereby declare that the work presented in this thesis entitled **"A Modeling Architecture for Data Analytics in Systems Biology : Simulating the ErbB Signaling Pathway"** is the original work done by me under the supervision of Dr. Elizabeth Jacob, Computational Modeling and Simulation Group, Environmental Technology Division of CSIR – National Institute for Interdisciplinary Science and Technology, Thiruvananthapuram, India. I also declare that this work did not form part of any dissertation submitted for the award of any degree, diploma, associateship, or any other title or recognition from any University/Institution.

Thiruvananthapuram

Date:

Arya A. Das

Council of Scientific and Industrial Research, Government of India
National Institute for Interdisciplinary Science and Technology
Pappanamcode, Thiruvananthapuram 695019, Kerala, India

Dr. Elizabeth Jacob
Chief Scientist
Computational Modeling & Simulation Group
Environmental Technology Division
Phone: +91 471 2515381, Fax: +91 471 2491712
Email: liz.csir@gmail.com

Thiruvananthapuram
November 2017

CERTIFICATE

I hereby certify that the work presented in this thesis entitled "**A Modeling Architecture for Data Analytics in Systems Biology : Simulating the ErbB Signaling Pathway**" is a *bona fide* record of the research carried out by **Ms. Arya A. Das** under my guidance and supervision.

All the relevant corrections, modifications and recommendations suggested by the audience and the Doctoral committee members during the pre-synopsis seminar have been incorporated in the thesis. I also certify that this work or no part of it has been submitted elsewhere for the award of any degree, diploma, associateship or any other title or recognition.

(Elizabeth Jacob)

Thesis Supervisor

Acknowledgement

I would like to thank all the people who helped me in completing this work. First and foremost to my research supervisor, Dr. Elizabeth Jacob, Chief Scientist, Computational Modeling and Simulation Section (CMS), Environmental Technology Division (ETD), CSIR - National Institute for Interdisciplinary Science and Technology (NIIST). She had always kept her doors open with a wide smile whenever I had queries regarding my research. Her knowledge, professional ethics, quest for perfection, work-freedom and frankness has helped me to complete this thesis in a successful manner. My research life would not have been much easier, had she not encouraged me to involve in other extra-curricular activities at NIIST.

I also thank Dr. Savithri S., an all-time problem-solver, for all her help, motherly hugs and for showing me what it is to be bold and optimistic.

It is now my privilege to acknowledge Dr. A. Ajayaghosh and Dr. Suresh Das (present and former directors of NIIST respectively), for providing me the necessary research facilities at the institute. I also thank Dr. Ajith Haridas (Former Head, ETD) for his help and support.

I would also like to acknowledge the Council for Scientific and Industrial Research (CSIR), New Delhi, for providing me financial assistance in the form of Senior Research Fellowship.

I am grateful to scientists, Dr. K. G. Raghu, Dr. P. Jayamurthy and Dr. K.V. Radhakrishnan for providing opportunities to collaborate with their works through docking studies. Special thanks to Dr. Shyni G.L., Dr. Dhanya R., Dr. Anusree S. S., Dr. Nishanth Kumar S, Amjesh, Syama H. P., Kavitha S. and Dhanya B.P. for helping me with the same.

I express heartfelt thanks to my doctoral committee external expert, Dr. Sumam Mary Idicula, Professor, Department of Computer Science Technology, CUSAT. I also extend my sincere thanks to Dr. G. Santhosh Kumar, Associate Professor, Department of Computer Science Technology, CUSAT, expert member of my Doctoral Committee for providing valuable comments and suggestions that helped in improving the quality of my work immensely. I express my sincere thanks to Rahul S., Akhil S. Karun and Jeen Mariam Mathews, my batch mates who have registered for Ph.D. in CUSAT.

I thank Dr. E. V. Soniya, Scientist, Rajiv Gandhi Centre for Biotechnology for being a member of my SRF assessment committee and for her valuable suggestions.

I thank Adarsh, Deepak and Rajeev for helping me with the MATLAB coding. I am also grateful to Suneeth and Neenu for helping me in my work. I was also fortunate to work with the project students, Darsana and Pranav, who had also supported me in completing my projects. I thank Rahul and Shaban for patiently helping me in typesetting and editing my thesis.

Further, I thank my fellow labmates and friends -Abhilash, Vishnu, Hari, Amal, Rejitha, Vidya, Aneesh, Midhun, Vinay, Gaurab, Dhaneesh, Anitha, Sinu, Krishna, Dr. Paneerselvam and Bincia for all the support and good times shared. Words would fail to express my gratitude to my friends at NIIST. I would especially like to thank Divya S., for always being there to tell me exactly what I always wanted to hear and for being a reason for me to pursue PhD. I would also like to acknowledge my friends – Surjith, George, Anju, Hari, Vaishak, Della, Shilpa, Sumina, Saranya, Linda, Parvathy, Arun, Devi, Vijina, Renjith, Nimisha, Athul, Suyana, Vani and many others at NIIST for all their support. Thanks for tolerating all my idiosyncrasies and for doubling the happiness in my life.

I express my very profound gratitude to my in-laws, especially my mother-in-law and father-in-law for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis.

Words cannot express how grateful I am to my husband, Sandeep, for his open-mindedness, freedom and for the pride he took in all my achievements, steered me in completing this.

I would also like to thank my grandparents, my brother-Arjun and other family members for their boundless love and care. I am also indebted to Aswathy for helping me with the corrections.

I thank my father for always protecting me under his beautiful wings and for his wise counsel. I owe a lot to my mother, for all of the sacrifices that she had made on my behalf and for the sympathetic ear. It is their dream that builds my life.

I dedicate this work to late K. G. Sukumaran and late Prince, for inspiring me in pursuing research and teaching me the art of life.

Above all I thank almighty for making all this happen.

This accomplishment would not have been possible without any of you. All your prayers for me was what sustained me this far.

Thank you very much, everyone!

PREFACE

The goal of systems biology is to characterize the network components completely and to find quantitative estimates for the interactions of the components. Advances in high-throughput biotechnologies have led to increasing volumes of biochemical data. Data mining techniques and design-driven systems modeling approaches have led to the concept of reverse engineering biological systems. Biological systems can be reverse engineered by experimentation and computational modeling. The knowledge thus gained can be used to re-engineer a similar system or its variants under different conditions.

A modeling architecture has been developed that learns from the data given by reverse engineering, and uses it for re-engineering the system. Essentially, it is a pipeline architecture that takes the data and available knowledge through an analytics process to provide insights into complex cellular phenomena. For the purpose of demonstration, signal transduction by biomolecules inside the cell has been taken. The ErbB signaling pathway has been chosen as it is one of the most intensely studied networks and its success mainly due to the fact that there was prolific activity in both experiments and modeling for the past two decades.

The ErbB pathway is characterized by multiple extracellular ligands, the four trans-membrane ErbB receptors (ErbB1 or EGFR; and ErbB 2–4). These receptors are also known as HER (1–4). The other important members of the pathway are cytoplasmic adapters, scaffolds, enzymes and small molecules. Human epidermal growth factor receptors (HER) are essential receptor tyrosine kinases that mediate cell proliferation, differentiation, migration, adhesion, apoptosis and embryogenesis. This pathway has been intensely studied both experimentally and computationally, to understand the mechanisms underlying their oncogenic potential and to also exploit them as therapeutic targets. Binding of a ligand to the receptor initiates the signaling, leading to homo or hetero-dimerization of receptors, followed by the activation of signaling pathways like Erk and Akt. Reverse engineering of these pathways are generally supported by mathematical models which quantitatively estimate the parameters of the system by fitting the model to the experimental data. These parameters, together with the established knowledge about the pathway, have

been encapsulated into a set of if-then rules to re-engineer the system by the bottom-up approach of agent-based modeling (ABM). The ABM is a spatio-temporal simulation of the experiment where the actions and interactions between agents, result in the occurrence of various cellular events.

The mathematical model that predicts receptor (EGFR and HER2) activation and trafficking patterns to fit the experimental data, is taken as the reverse engineered system. Ligands and receptors (monomers and dimers) are considered as agents that reside indifferent cellular compartments. These agents interact quantitatively leading to receptor activation, trafficking and receptor dimerization which are validated with the results of the experiment. As agents undergo state change, it is also possible to calculate rate constants from the changing species concentration.

The model was expanded by including the HER3 receptors and HRG ligands linking receptor expression levels to dimerization and activation. The importance of HER3 expression in drug resistance and tumorigenesis has increased its focus as a potential drug target. The model quantitatively predicts the relative contributions of the HER dimers in four distinct cell lines and how the changes in the receptor phosphorylation levels alter activation patterns in the downstream cascades, mainly Erk and Akt kinases.

As a potential application, the model was extended to study the therapeutic efficacy that can be achieved from combining drugs. The potential of "Combine and Conquer" is a new challenge in the design of targeted therapies. However, the underlying molecular mechanisms on action of these drugs have remained enigmatic. The antitumor activity of two drugs-Trastuzumab and Pertuzumab, both humanized monoclonal antibodies, alone and in combination on cell lines, is investigated. Using the ABM model, by including Trastuzumab and Pertuzumab as drug agents, hypothesis-testing studies were done to rationalize the underlying mechanism of action and their observed synergetic effects. The model serves as a powerful technique for *in silico* clinical trials which can drastically reduce the lead time in the drug discovery pipeline.

The proposed modeling architecture can give new insights using the existing knowledge found in experimental literature and mathematical/ computational models. Re-engineering models, built using the results of reverse engineering, open up the possibility of harnessing the power pack of data which now lies scattered in literature. Virtual experiments could then become more realistic when empowered with the findings of empirical cell biology and modeling studies. The reverse-engineering-re-engineering pipeline architecture is a promising tool for deep data analytics in integrative systems biology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABM	Agent Based Model
AKT	Protein kinase B
BAD	Bcl2-associated death promoter
DAG	Diacylglycerol
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ErbB	Lineage of proteins (ErbB1-4). ErbB derived its name from avail viral erythroblastosis oncogene to which the receptors are related
ERK	Extracellular signal-Regulated Kinases
Grb2	Growth factor receptor-bound protein 2
GTP	Guanosine Triphosphate
HER	Human Epidermal growth factor Receptor
HME	Human Mammary Epithelial Cells
HRG	Heregulin
IP3	Inositol 1,4,5-trisphosphate
MEK/MAPK	Mitogen-activated protein kinase kinase
min	minutes
mTOR	Mammalian target of rapamycin
ODE	Ordinary Differential Equation
PI(3,4,5)P3/PIP3	Phosphatidylinositol 3,4,5-trisphosphate

PI(4,5)P2/PIP2	Phosphatidylinositol 4,5-bisphosphate
PI3K	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PKC	Protein serine/threonine kinase C
PLCɣ1	Phospholipase Cɣ1
Raf	**R**apidly **A**ccelerated **F**ibrosarcoma, a family of three serine/threonine-specific protein kinases
Ras	A low-molecular weight guanosine-nucleotide-binding protein
SOS	Son of Sevenless
STAT	Signal Transducer and Activator of Transcription

LIST OF ABBREVIATIONS, SYMBOLS AND PARAMETERS USED IN THE MODEL

123Model	A mathematical model developed by Shankaran, et al., 2013 with HER1-3
12Model	A mathematical model developed by Shankaran, et al., 2008 with HER1 and HER2
17L	Low HER2 expresser cell line
24H	Medium HER2 expresser cell line
A11H	High HER2 expresser
B5	High HER3 expresser cell line
D20	HER1,HER2 and HER3 expresser cell line
EE	Early Endosome
$\mathbf{k_d}$	Degradation rate
$\mathbf{k_e}$	Internalization rate for dimers
$\mathbf{k_f}$	Dimerization rate
$\mathbf{k_{ri}/k_{rs}}$	Dissociation rate for internal compartment/surface
$\mathbf{k_t}$	Internalization rate for monomers
LE	Late Endosome
mRPt	Total mass of phosphorylated receptors
mRt	Total mass of receptors
pf11	Phosphorylation factor of HER1-HER1
pf12/ pf21	Phosphorylation factor of HER1-HER2
pf22	Phosphorylation factor of HER2-HER2
pRPi	Phosphorylated receptor levels in EE.

pRPt	Total receptor phosphorylated (PM+EE)
Par	Parental cell line
PM	Plasma Membrane

1

Introduction

Scientists, have spent much time, looking through the lens of reductionism to decipher the working of complex systems. This is achieved by breaking the problem into component parts, and then studying the behaviour of each part. A major limitation of reductionism is the assumption that the parts are independent and can be studied in isolation. This assumption fails when the inter-connections between the parts are significant and cannot be ignored. For e.g., each gene in an organism can be sequenced separately, but gene expression is the result of complex regulatory networks of genes and the interactions of their protein products. This fostered the development of the systems theory where the focus of the lens shifted to the connections and the information/energy/material flow between the parts. Contrary to the reductionist approach, the emergent system behaviour (the whole) need not be the sum of the parts.

However, the system-level approach also has its drawbacks as it requires reasonable understanding of the parts, to succeed. The inter-connections within the system may be too many, too complex or too transient to measure and generation of such data is an unwieldy exercise. This is especially true of biological systems where emergent properties cannot be predicted without some knowledge of the parts to begin with. The theme of this thesis is to use a data driven approach to bring together the complementary paradigms of reductionism and holism in the field of systems biology. The key idea is generic and so can be applied to other areas as well.

Systems biology has mainly pursued the reductionistic path relying on well-designed experiments and computational tools like mathematical models. In such endeavours, large volumes of biochemical data are being generated due to the advances in high-throughput bio-technologies. This data which is assuming the proportions of "big data", offers a great challenge for data scientists to predict the dynamics of biological processes.

1.1 Objectives

The objective of this work is to create a platform for the study of emergent behaviour of complex biological networks by drawing upon the outcomes of reductionistic techniques.

The specific objectives are:

1. Design of a reverse engineering - re-engineering pipeline architecture for systems biology data analytics.

   Data mining and design-driven systems modeling approaches have been used for reverse engineering biological systems by the reductionist principle. Reverse engineering attempts to derive mechanistic insights into the complex processes involved in signaling, gene regulatory and metabolic networks by the interplay of data generated from wet lab experiments and modeling studies. The knowledge thus gained can be used to re-engineer a similar system or its variants.

2. Implementation of the modeling architecture on a simple model of signal transduction in the cell by activation and trafficking of EGFR and HER2 receptors.

   A mathematical model that predicts receptor (EGFR and HER2) activation and trafficking is chosen from literature. This is an effort of reverse engineering by which the parameters for dimer formation affinities and trafficking rates were estimated so that they fit to the experimental data. The data from this model-based analysis is used to re-engineer the system by emergent modeling.

3. Modeling the EGFR, HER2 and HER3 dynamics and their effect on the downstream Erk and Akt pathways.

   The starting point is again, a reverse engineered system by the same authors as in 2 above. The model is extended to include the third receptor of the ErbB family, namely, ErbB3 in cells co-expressing EGFR, HER2 and HER3. As a consequence, Erk and Akt pathways are activated. The experiment measures the additional effects of the HER3 receptor and its partnering ligand on activation of dimers and downstream proteins that lead to proliferative

outcomes. The kinetic model takes into account the interactions of the additional receptor, giving rise to almost a three-fold increase in the number of parameters to be estimated.

4. Application: Modeling the effect of combination drugs on HER targeted therapies for cancer.

   Two or more drugs that individually produce overtly similar effects can display greatly enhanced effects when given in combination. Drug combinations are now being used in cancer therapy and there is growing interest in them being used for cancer prevention. Rigorous trials are required to prove the efficacy of such drugs. The mechanism of action of the drugs and their dose combinations are important in determining the existence of synergy. The modeling pipeline takes the current state of knowledge on the mechanism of action of two popular drugs used for ErbB2 positive cancers, to investigate the emergent behaviour of the drug-receptor interactions in a system. This application can be of great value to Systems Pharmacology where *in silico* experimentation may be the best way to evaluate the effect of such therapies in real time.

1.2 Outline of the thesis

A modeling pipeline has been developed, combining two formalisms of systems modeling, where the outcomes of reverse engineering can drive a re-engineering model. The pipeline architecture serves as a generic paradigm as there is no 'one size fits all' modeling solution. Combining modeling formalisms in this manner for data analytics can prove to be a powerful tool for new insights and predictive modeling in systems biology.

Chapters 2 and 3 familiarize the reader with the concepts and terminology related to the research presented in later chapters. Being an interdisciplinary work, basic understanding of the biology of signal transduction and also computing methodologies used for modeling, are required. Chapter2 is about the reverse engineering paradigm and how it has been applied to study the ErbB signal transduction pathway. The basic biology of the ErbB signal transduction and the

computational modeling schemes for studying the pathway have been introduced. The link between ErbB network dysregulation and cancer has been reviewed.

Chapter 3 introduces the concept of re-engineering and how it can benefit from the efforts of reverse engineering with the emergent principles of Agent-Based Modeling (**ABM**). This is followed by an overview of ABM and their applications in diverse fields ranging from economics to ecology. The online software tools available for the design and development of ABMs have also been enumerated.

The objectives listed in Section 1.1 are the subject of Chapters 4-6. Chapter 4 covers objectives 1 and 2. The first objective, namely, the modeling pipeline connecting reverse engineering with re-engineering, is the main focus of Chapter 4. The re-engineering part is implemented using an ABM. For the purpose of demonstration, the reverse engineered model for EGFR-HER2 receptor activation and trafficking in signal transduction has been re-engineered by an ABM (objective 2).

In Chapter 5, the model is further expanded to study the receptor dynamics and downstream signaling (objective 3).

Chapter 6 is a practical application of the computational pipeline (objective 4). The ABM for ErbB signal transduction has been used for predictive modeling of the effect of drug combinations on HER targeted therapies for cancer.

Chapter 7 concludes with a discussion of the research pipeline developed – its applications, issues, computational challenges, future perspectives and how it fares as a data analytics tool to bring new life to old data generated from biological experiments and models.

1.3 Contributions of the thesis

The major contributions of this thesis are:

1) The conceptual model of a data-driven pipeline architecture that makes the outcomes of reverse engineering, available for re-engineering.

2) To develop models that can simulate virtual experiments designed using the findings of empirical cell biology and modeling studies.

3) A modeling framework for *in silico* drug studies that can test the molecular mechanisms involved in the action of drugs, with specific application to synergy effects of combination drugs in cancer therapy.

4) Applications can be developed for complex systems.

5) A paradigm for data science - Deriving emergence from reductionism.

2

Reverse engineering the ErbB signaling pathway

Reductionism is the practice of describing and analyzing a complex phenomenon in terms of its simple or fundamental constituents. Reducing the problem to sub-problems and integrating their solutions, can help understand the working of the system as a whole. At this point, it is important to understand the difference between complicated and complex systems. Simple problems can be solved by following a protocol or mastering a technique and ensures repeatablity. Complicated problems are large in scale, requiring coordination and specialized skills but once mastered, complicated systems have a high degree of predictability and repeatability. These systems can be understood by taking them apart and analyzing the sub-systems. Complex systems are based on relationships, their properties of self-organisation, interconnections and evolution. Biological systems are complex and non-linear, operating at a wide range of scales, starting from molecular to systems level. Modeling such systems was possible only by classical reductionist-analytical strategies.

Reverse engineering, basically, employs the reductionist principle. Section 2.1 introduces the applications of reverse engineering with special reference to ErbB signaling pathway. Sections 2.2 and 2.3 describe the basic biology of ErbB and reviews the computational modeling attempts, made to reverse engineer various aspects of its signaling dynamics. Section 2.4 throws light on the cancer connection of the ErbB network.

2.1 Reverse Engineering

Reverse engineering, is the technique of extracting knowledge or design information from a system or a man-made product and reproducing it or creating a new product, based on the extracted information.

2.1.1 Reverse engineering of hardware and software

While reverse engineering, an engineer works backwards from a product, disassembling it, in order to understand the processes involved in making it. It is generally used to get the know-how and copy a competitor's product. With the knowledge thus gained, one can, not only re-engineer the product, but can also make alterations and adaptations or even design new products.

Hardware reverse engineering involves, taking apart a device to see how it works. For example, if a processor manufacturer wants to see how a competitor's processor works, they can purchase it, disassemble and then make a processor similar to it. When machine parts fail and replacements are no longer available, reverse engineering is used to reproduce the part from a sample of the broken part. Using 3D scanning technologies, the 3D model of the part can be created for casting the part.

Software reverse engineering involves reversing a program's machine code to get back the source code when a bug has to be fixed, the performance has to be improved or to identify a virus in the program.

2.1.2 Reverse engineering of biological systems

Biological systems are extremely complex and have emergent properties that cannot be explained, or predicted, by studying their individual parts. The reductionist approach, although successful in the early days of molecular biology, underestimates this complexity and therefore has an increasingly detrimental influence on many areas of biomedical research, including drug discovery and vaccine development (Van Regenmortel, 2004).

Complexity in biological systems is due to many reasons. The cross-talk that exists among system components, and between these components and their (micro) environments is untractable. It is further augmented by the heterogeneity of system components and how macroscopic behaviour in biological systems is an outcome of a vast range of interactions at multiple scales, operating in the absence of a central organizing structure (Kaul and Ventikos, 2015).

The large volume of biochemical data from high-throughput biotechnologies have impacted the progress in reverse engineering biological systems (Quo, et al., 2012). Taking the biological data as input, these algorithms, when simulated recapitulates the dynamics of experimental *in vivo* outcomes (Lobo, et al., 2016). Reverse engineering has been successfully used for systems analysis in many contexts, from discovering transcriptional control in largely uncharacterized organisms (Bonneau, et al., 2007) to understanding the complex regulatory networks underlying development and evolution of multi-cellular organisms (Crombach, et al., 2012). They are also studied in microbial metabolics, signaling and regulatory networks and pattern-forming developmental processes in animals (Becker, et al., 2013). Sahin, et al. (2009) constructed a literature-based protein network and combined computational simulations and experimental testing of simulation results to define potential therapeutic strategies for *de novo* Trastuzumab resistant breast cancer. A combination of reverse and forward engineering was used to reconstruct ErbB network and simulate perturbations to study drug responses. The deregulated protein interactions in ErbB network was reverse engineered using Boolean models and forward engineering allowed to analyze the stable states of the reconstructed system (von der Heyde, et al., 2014).

In systems biology, reverse-engineering is characterized by the tight coupling of experiments and computational modeling, where dynamical models of regulatory or biochemical reaction networks are fit to quantitative data (Villaverde and Banga, 2014). A model-based analysis incorporates the basic assumptions and hypotheses of the system under study. The model is tested by repeatedly altering its parameters and fitting to experimental data. A successful fit will yield a unique set of parameter estimates that cause the model to reproduce the data accurately (Becker, et al., 2013).

Some of the tools for reverse engineering in systems biology are bayesian statistics, inverse problems, machine learning, nonlinear physics, (bio) chemical kinetics, control theory and optimization (Villaverde and Banga, 2014).

2.2 Basic Biology of the ErbB Signaling Pathway

Cellular signaling is the process of transmission of molecular signals initiated by cell-surface receptors to the interior of the cell. Signal transduction cascades are molecular circuits that detect, amplify and integrate external signals to generate responses such as changes in gene expression, enzyme activity or ion-channel activity. Signaling is effected by a network of chemical reactions between molecular species. ErbB, Wnt, notch and hedgehog are examples of signal transduction pathways in cells.

A receptor is a protein molecule that receives chemical signals from outside a cell. When such chemical signals bind to a receptor, they cause some form of response, e.g. a change in the electrical activity of a cell. In this sense, a receptor is a protein-molecule that recognizes and responds to endogenous chemical signals.

The ErbB/HER receptor tyrosine kinase family consisting of four receptors are activated following ligand binding (Zhang, et al., 2009). HER family receptors are expressed in epithelial, mesenchymal, neuronal lineages, endothelial and cardiac cells. The receptors activate a multi-layered network, mediating crucial pathways that lead to cell proliferation, differentiation, migration and altered metabolism. Furthermore, HER family receptors help in development and maintenance of various integrative body systems such as cardiovascular and nervous system (Mujoo, et al., 2014). Increased expression and mutation of members of ErbB family is associated with several types of cancers (Roskoski, 2014).

A ligand is a substance that forms a complex with a biomolecule to serve a biological purpose. In protein-ligand binding, the ligand is usually a molecule which produces a signal by binding to a site on a target protein. The binding typically results in a change of conformation of the target protein. Early signaling involves binding of an extracellular ligand that promotes homo- or hetero-dimerization between the ErbB family receptors. Receptor dimerization activates the intracellular tyrosine kinase domain of the receptors by auto and trans-tyrosine phosphorylation. On phosphorylation of multiple tyrosine residues, these domains act as binding sites for adaptor proteins that initiates the downstream signal transduction pathways (Yarden and Shilo, 2007). The ErbB signaling network consists of several

interconnected cascades like PI3K/Akt, MAPK, PLCγ1/PKC, Src and STAT which result in various biological responses, including apoptosis, migration, cellular proliferation, maturation, survival, adhesion, differentiation and angiogenesis.

While ligand binding and dimerization are taking place, regulation of receptor activity plays a crucial mechanism in cell signaling. Trafficking/endocytosis process remove receptors from the cell surface or shunting them to the degradative fates (Hendriks, et al., 2006). The trafficking process can be subdivided into receptor synthesis, internalisation, recycling and degradation. Endocytosis regulates cell signalling most simply by controlling the number of receptors available for activation in the plasma membrane (Section 2.2.3).

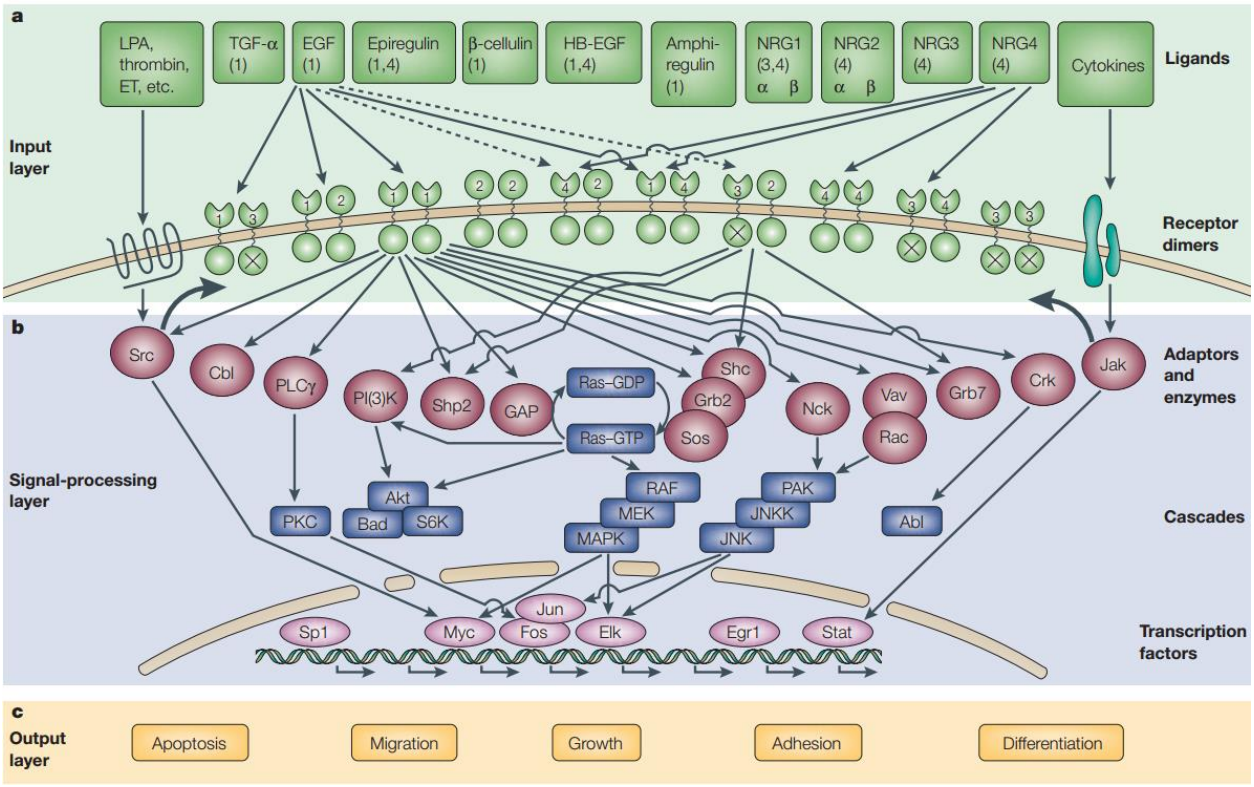


Figure 2.1 The ErbB Signalling Pathway (Reproduced from Yarden and Sliwkowski (2001))

2.2.1 Early Signaling: Receptor activation

The input layer (Figure 2.1) consisting of receptors and ligands constitutes early signaling.

2.2.1.1 Receptors

The EGFR family consists of four members that belong to the ErbB lineage of proteins (ErbB1-4). ErbB derives its name from avail viral erythroblastosis oncogene to which these receptors are related. The receptors are:

a) EGFR(Epidermal Growth Factor receptor)/ErbB1/HER1

b) ErbB2/HER2/Neu- Neu is an oncogene related to rat ErbB2 gene

c) ErbB3/HER3, and

d) ErbB4/HER4.

The ErbB family of protein kinases consist of (Figure 2.2):

a) An extracellular domain which is divided into 4 parts: Leucine rich domains I and III that participate in ligand binding, domains II and IV which contain cysteine residues that participate in disulphide bond formation. Domain II is involved in the formation of homo and heterodimers.

b) Single transmembrane segment of 19-25 amino acids.

c) An intracellular portion of about 550 amino acids containing a juxtamembrane segment, protein kinase domain and a carboxy terminal tail.

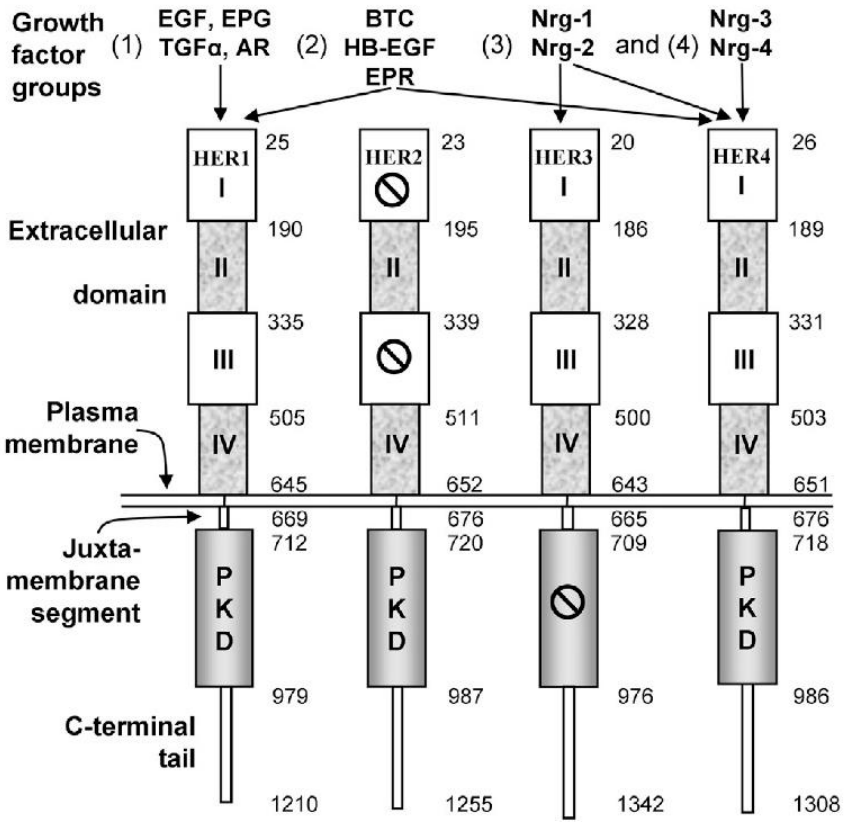


Figure 2.2 Organisation of Human ErbB/HER receptor. Courtesy Roskoski (2014)

The ErbB1/3/4 receptors employ only a ligand mediated dimerization mechanism. These receptors resemble each other and possess a tethered conformation with the dimerization arms buried within, restricting their movements, thus, auto inhibiting ligand binding and dimerization. In contrast, ErbB2 exist in open conformation making it a preferred dimerization partner for other ErbB family receptors. Upon binding of the ligand, the receptors are activated by exposing their dimerization interface by promoting a large domain rearrangement(Ferguson, et al., 2003). The active dimeric form thereby juxtaposes the intracellular kinase domain for mutual transphorylation on several tyrosine residues (Orton, et al., 2008). These sites act as docking sites for a plethora of cytoplasmic adaptor proteins, typically containing SH2 or PTB domain, which further stimulates the downstream cascades.

The inactive EGFR is usually present as a monomer in normal cells, but at higher levels of expression they form inactive pre-formed dimers. These dimers are primed to bind to ligands (Chung, et al., 2010). It is reported that a single ligand can

induce ErbB1 or ErbB4 homodimer formation and activation(Liu, et al., 2007). In spite of extended conformation of ErbB2, it does not form active dimers unless overexpressed. All receptors possess similar protein kinase domains. ErbB1/2/4 possess protein kinase activity while ErbB3 is catalytically impaired. Experimental studies suggest that ErbB3 is not kinase dead but undergoes autophosphorylation at a rate that is $1/1000^{th}$ that of EGFR. In ErbB2-ErbB3 heterodimer, both receptors get phosphorylated by ErbB2 catalysing the phosphorylation of ErbB3 (Roskoski, 2014).

2.2.1.2 Ligands

Ligands of the HER family receptors are divided into three groups based on the receptors they bind to:

1. HER1: Epidermal growth factor (EGF), amphiregulin (AR), and transforming growth factor-α (TGF-α).
2. HER1 and HER4: Betacellulin (BTC), heparin-binding EGF (HB-EGF), and epiregulin (EPR).
3. HER3 and HER4: Neuregulins (NRG, also known as Neu differentiation factors, NDFs, or heregulins, HRG) which include two subgroups based on their capacity to bind HER3 and HER4 (NRG1 and NRG2) or only HER4 (NRG3 and NRG4).

2.2.2 Receptor Trafficking

Endocytosis regulates cell signalling by controlling the number of receptors available on the plasma membrane for activation. Clathrin–dependent and independent endocytosis of EGFR represents the first step in receptor downregulation. Many signalling receptors are also modified by ubiquitylation thus serving as clathrin-coated pit targeting signals. Endocytosis in some receptors is stimulated by ligand-induced activation. The nature of ligands also play significant role in fate of EGFR following endocytosis, ubiquitylation, and degradation. EGF has high affinity for EGFR by remaining persistently bound to the receptor, efficiently inducing its degradation. Expression of new receptor requires RNA and protein synthesis. Internalised receptors are sorted to recycling and endosomal degradation pathways (Figure2.3). An important feature of the endosomal sorting is its efficiency in targeting these receptors to late endosomes and lysosomes for degradation. Ligand-

induced ubiquitylation has a key role in the lysosomal targeting and downregulation of receptors (Sorkin and Goh, 2009).

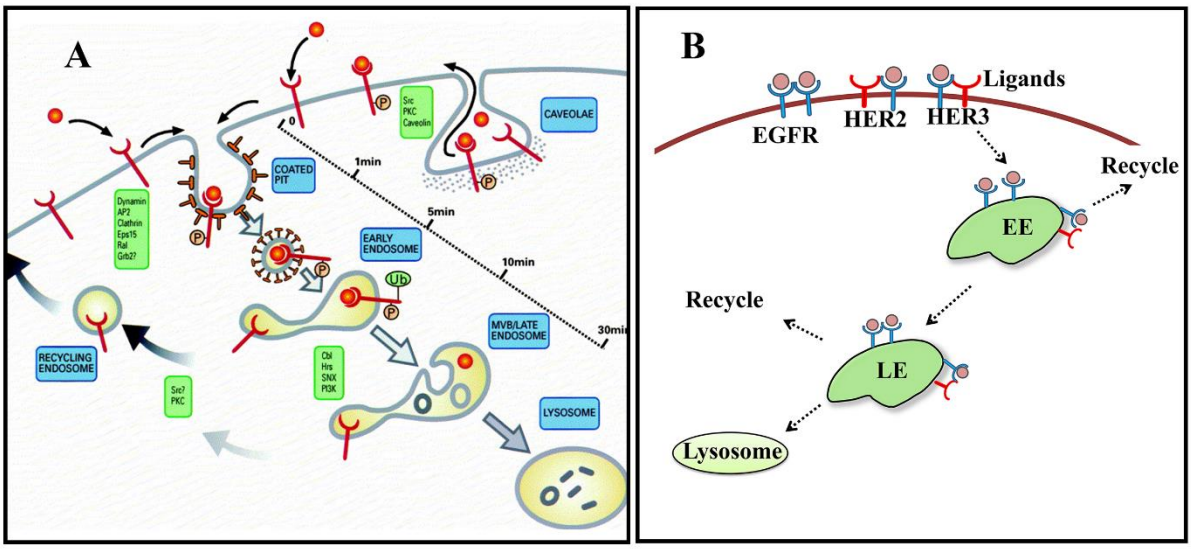


Figure 2.3 Receptor trafficking. A. The biological model B. A Simplistic view

2.2.3 Downstream Signaling

The ErbB signaling comprises several cascades that are interconnected and overlapping. These cascades integrate signals from external stimuli translating them into intracellular signals. These includes PI3K/Akt, Ras/MEK/ERK1/2, STAT and phospholipase C (PLCɣ). Our focus in this study is on pathways PI3K/Akt and Ras/MEK/ERK1/2 (Figure 2.4).

2.2.3.1 PI3K/Akt

PI3K/Akt pathway plays a major role in cell survival. PI3K binds mostly to the phosphotyrosine sites of ErbB3 receptor, leading to its activation (Thorpe, et al., 2015). Following receptor activation, membrane-bound phosphatidylinositol 4,5-bisphosphate (PI(4,5)P2) gets phosphorylated to generate phosphatidylinositol 3,4,5-trisphosphate (PI(3,4,5)P3). PIP3 translocates the serine/threonine kinase Akt (also known as Protein kinase B, PKB) to the membrane by binding to it with a higher affinity. Activated Akt catalyses the activation of serine/threonine kinase mTOR (mammalian target of rapamycin) which acts as a substrate for many cellular processes for cell survival. Akt also inhibits Bcl2-associated death promoter (BAD);

which helps in pro-apoptosis. PTEN negatively regulates PI3K/Akt pathway by hydrolysing PIP3 to PIP2. The ErbB3/PI3K/Akt pathway is a major cause of treatment failure in cancer therapy because of its role in therapeutic resistance. HER3 is frequently co-expressed with other receptors promoting tumor initiation and progression through PI3K/Akt signalling.

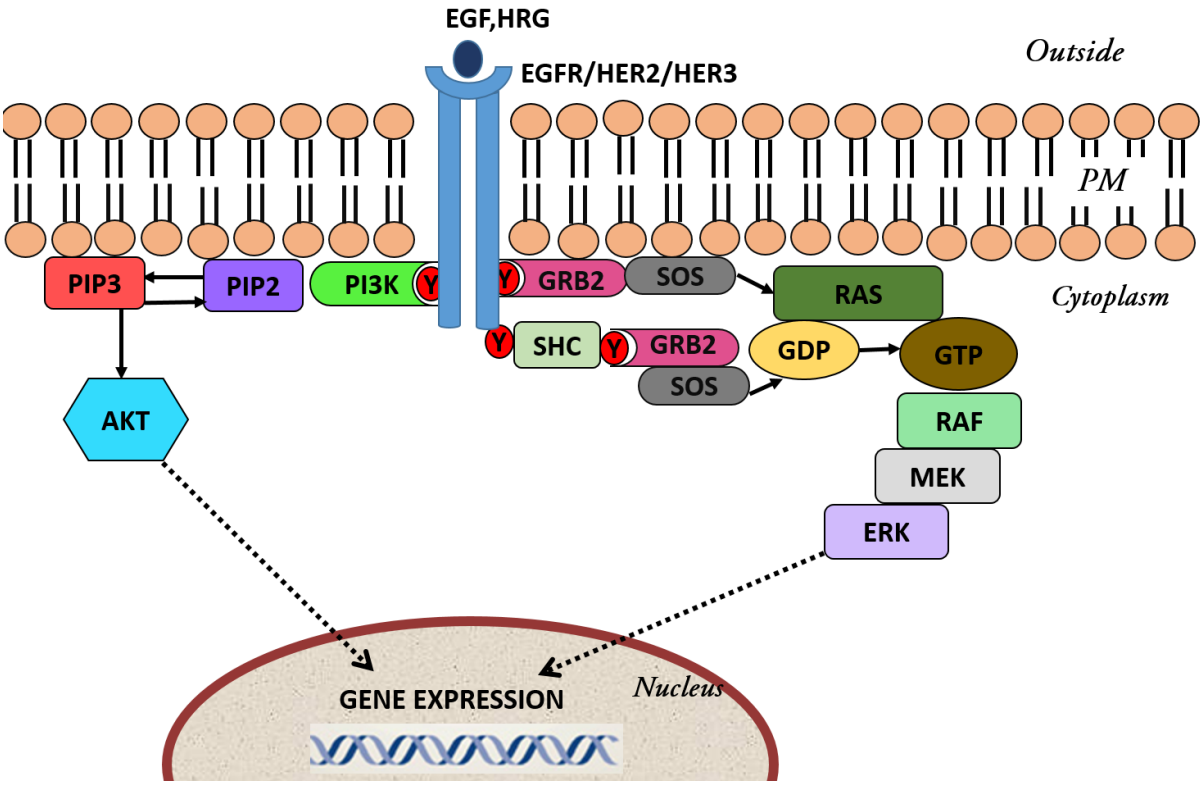


Figure 2.4 Receptor activation, trafficking and downstream signalling. PM-Plasma Membrane.

2.2.3.2 MAPK pathway

On activation of receptors by phosphorylation on tyrosine residues, these residues act as docking sites for large number of adaptor proteins such as Shc and Grb2 which subsequently recruit Ras-guanosine nucleotide exchange factor SOS. This recruitment brings it closer to small G-Protein Ras which gets activated by loading of Guanosine Tri phosphate (GTP). Ras then binds to Raf -1, translocating it from cytoplasm to PM. Active Raf dual phosphorylates to activate MEK, which in turn dual phosphorylate to activate Erk. Phosphorylated Erk can phosphorylate over 160 substrates in cytoplasm and nucleus which regulates gene expression by phosphorylating transcription factors like elk, myc and ets promoting cell

division(Orton, et al., 2008). Ras-GTP is converted back to Ras-GDP by the GTPase-activating proteins (GAP).

2.2.3.3 Phospholipase C pathway (PLCγ1)

ErbB1, ErbB2 and ErbB4 possess several PLCγ1 phosphotyrosine binding site. Activated PLCγ catalyzes the hydrolysis of phosphatidylinositol 4,5-bisphosphate to form inositol1,4,5-trisphosphate (IP3) and diacylglycerol (DAG). IP3 releases Ca2+ from the endoplasmic reticulum and DAG activates the protein serine/threonine kinase C (PKC). PKC catalyzes the phosphorylation of dozens of proteins that effects angiogenesis, cell proliferation, cell death, increased gene transcription and translation, cell migration, cell adhesion activation and several transcription factors that lead to cell proliferation and survival. One of the downstream effectors is the Raf/MEK/ERK1/2 pathway bypassing Ras.

2.2.3.4 STAT (Signal Transducer and Activator of Transcription)

STATs are cytoplasmic transcription factors that translocate to nucleus to activate target genes (Chung, et al., 1997). STAT family which consist of 7 proteins- STAT 1, STAT 2, STAT 3, STAT 4, STAT 5a, STAT 5b and STAT 6, plays a significant role in signal transduction and activation of transcription (Zhong, et al., 1994). STAT3 plays an important role in cell differentiation, proliferation, migration, apoptosis, and cell survival, depending on the signal, tissue, and cellular context

STAT1, STAT3, and STAT5, were constitutively complexed with ErbB1 and rapidly phosphorylated on tyrosine in response to EGF. The Phosphorylation of tyrosine residues on the receptor acts as a docking site for STATs (via their SH2 domain). This activates STAT by phosphorylation on its tyrosine residue, inducing it to dimerise and translocate to the nucleus. EGF ligand activates STAT1 and STAT3 which either forms homodimers of itself or STAT1-STAT3 heterodimers (Chung, et al., 1997).

2.3 Computational Modeling

The study of ErbB signaling pathway over the years have employed a combination of experimentation and computational modeling strategies. Being a constantly improving and evolving pathway, models help to suggest interesting new hypotheses

and explanations for the observed data (Orton, et al., 2008). Computational tools utilized to model biological phenomena can be categorized, broadly, as continuum or discrete. Whereas the former describe the numerical changes of the variables that represent the system, the latter can indicate how and why the dynamics involving system components operate. Continuum approaches employ classical differential equation-based models that may have numerical or approximate solutions. However, mathematical equations representing either a collection of cells or organisms or their (micro) environment, do not lend themselves as the most precise form of ontologies for biological systems. Their shortcoming is even more pronounced when simulating emergent behaviour that arises through "self-organization" which cannot be characterized "a priori".

During the past two decades, most computational models have been focusing on ErbB binding, internalization, dimerization and degradation. Kholodenko, et al. (1999) model served as the starting point for combining both experimental kinetic analysis and computational models of EGFR signaling. Kinetic models based on ODE predict time dependent profiles and steady state levels of species in a biological system where experimental data is not available (Hasdemir, et al., 2015). All species are uniformly distributed throughout the cell and are defined by differential equations whose solutions determine the concentration of the species as a function of time.

Mathematical models of varying degrees of complexity have been constructed to describe ErbB phosphorylation, trafficking and downstream signaling. Rate parameters for these models are usually set based on the literature for related systems, or are obtained by fitting to a limited set of experimental data. A critical criterion for model development is the chosen level of complexity: the model should capture the fundamental biophysical/biochemical reactions in the system, and should be detailed enough to address the mechanisms. At the same time, the availability of model parameters, and the ability to identify parameters from the available experimental data should be taken into account. Models that are too simple may be devoid of the necessary details for addressing relevant questions, while too complex models may incorporate many unknown parameters. Ideally, model scope and complexity should be addressed by integrating model construction and

experimental design at the outset, and refined through a systematic experiment-modeling iteration cycle.

Several groups, developed ordinary differential equation (ODE) models which focused primarily on dynamics of receptor activation and trafficking (Hendriks, et al., 2003; Shankaran, et al., 2008) and on understanding the transmission of signals to the downstream elements of the pathway (Birtwistle, et al., 2007; Chen, et al., 2009; Schoeberl, et al., 2002). Chen, et al. (2009)modelled one of the largest differential equation model to analyze immediate-early signaling involving ErbB1–4 receptors (EGFR, HER2/Neu2, ErbB3 and ErbB4), and the MAPK and PI3K/Akt cascades. These models consist of rate equations, describing how signaling molecules change as a function of each other and rate constants that are calibrated to the experimental data (Janes and Lauffenburger, 2013). Shankaran, et al. (2013)established a quantitative link between receptor expression level and downstream signalling by constructing an integrated mathematical model of HER activation and trafficking.

Complementary to mathematical models are the stochastic ABMs and the Gillespie algorithm (Gillespie, 1977; Hsieh, et al., 2010; Resat, et al., 2003). ABMs predict emergent phenomena that may reveal patterns which are not obtained by the aggregate stochasticity of Gillespie simulations. Agents have the advantage of modeling interactive behaviour and retain memory of past events (Figueredo, et al., 2014). Integrating experiments with model-based analysis have helped to understand these complex cellular systems(Wiley, et al., 2003).

Creamer, et al. (2012) modeled the ErbB signaling pathway by giving importance to its site specific molecular interactions. The novelty of their study lay the use of rule based methodology to understand lower scale systems. Rule-based modeling of biochemical systems involves, representation of molecules as objects and molecular interactions as rules. BioNetGen is an open-source software for specification and simulation of rule-based models (Faeder, et al., 2009). The model specification is done in BioNetGen language (BNGL). The network generation engine generates the reaction network, by iteratively applying rules to an initial set of species. The model can then be simulated in BioNetGen by ODE solvers or by Gillespie's Stochastic Simulation Algorithm (SSA). The particle-based, Network-Free Stochastic Simulator (NF-Sim), attempts to overcome the combinatorial challenge

posed by multiple molecular states and their myriad interactions (Sneddon, et al., 2011) by coarse-graining of reaction mechanisms. BioNetGen has the capability to model signal transduction systems (Blinov, et al., 2004). NFSim has been used to simulate a large model for ErbB receptor signaling, incorporating site-specific details of protein interactions (Creamer, et al., 2012).

The number of components and their interactions yields a large number of distinct combinations which lead to combinatorial complexities (Blinov, et al., 2004). In a hybrid particle/population approach, the problem of combinatorial complexity is dealt with, by treating rare, complex species as particles and plentiful species as population variables. Hogg, et al. (2014) introduced rule based models that uses hybrid particle/population approaches to model biological systems. Though simulators like NFSim use both rule based and agent based Monte Carlo simulations, they obtain predefined parameters and rate constants from the input file. A partial network expansion is implemented in BioNetGen and the hybrid model is simulated using NFsim. Walker, et al. (2008) designed an ODE model of EGFR-MAPK signaling at cellular level and scaled it up to a multi-agent rule-based model considering each cell as an agent. The Table 2.1 shows various models for ErbB signaling.

Table 2.1 Models for ErbB signalling

NAME	YEAR	DESCRIPTION
Boris N.Kholodenko	1999	Kinetic model of short term response of cell to EGF ligand
Haluk Resat	2003	Probability weighted dynamic Monte Carlo simulation of trafficking and signalling pathway
Birgit Schoeberl	2002	ODE based mathematical model for MAPK activated surface and internalised EGF receptor
Salim Khan	2003	Java-based multi-agent toolkit DECAF to simulate EGF pathway
B. S. Hendrik	2006	ODE based model for ErbB1-3 dimerization, trafficking and

		activation
M.R. Birtwistle	2007	ODE based model relating EGF and HRG stimulation on receptor stimulation and Erk and Akt activation
Zhihui Wang	2007	Multi-scale ODE based model to study expansion dynamics EGF-ERK signal transduction
M.Y. Hsieh	2008	SPS platform-Agent–based stochastic approach to study receptor density and spatial organisation
D. C. Walker	2008	Multiscale ODE model of EGFR-ligand activation of MAPK coupled with agent-based representation of individual cells.
William W. Chen	2009	ODE based model of immediate early signalling of ErbB1-4 analyzing parameter sensitivity, uncertainty and exploration factors controlling input-output parameter.
Samaga R.	2009	Large –scale logical model of signalling through 4 ErbB receptors including the Erk, JNK, p38 MAPK cascade, Akt signalling ,STAT's and PLCɣ pathway.
Matthew S Creamer	2012	Rule-based modeling ERBB receptor signalling for site-specific details of protein-protein interactions
Harish Shankaran	2013	Mathematical model of HER activation, trafficking, and phosphorylation
Toma´ sˇ Helikar	2013	Multiscale dynamical model of ErbB receptor signal transduction
Shannon E. Telesco	2013	Molecular and network model of ErbB4 activation and signalling
Justin S. Hogg	2014	Exact hybrid particle/population simulation of rule –based model from ligand binding to nuclear ERK activity

2.4 ErbB network and cancer

Loss of regulation of the ErbB receptors underlies many human diseases, most notably cancer (Wieduwilt and Moasser, 2008). EGFR overexpression plays a key role in pathogenesis of many types of cancers. Amplification and overexpression of ErbB2 is a leading cause for most of the breast and gastric cancers. HER3 is frequently co-expressed with other receptors promoting tumor initiation and progression through PI3K/Akt signalling, aiding in development of breast, castration resistant prostate cancer, platinum resistant/refractory ovarian cancer and EGFR tyrosine kinase inhibitor (TKI)-resistant non-small cell lung cancer (NSCLC)(Ma, et al., 2014). The deregulation of HER4 has shown anti-carcinogenic roles in certain tumors unlike the other HER receptors which are associated with several cancers. Overexpression of HER4 has led to increased survival of cancer (Shankaran, et al., 2013). HER4 may be only member in the ErbB family to play anticancer role in certain tumor context.

Generally, cancers are treated with a combination of surgery, radiation therapy and/or drug treatment. The mechanism of action of ErbB receptors depends heavily on their interaction with the ligands and their dimerization partners. Understanding the function and complex regulation of these receptors fuelled the development of targeted therapies for human cancer. Monoclonal antibodies (mAbs) are capable of interfering with ligand activation and in disabling heterodimeric kinase complexes by binding to the extracellular domain regions of the receptors (Wieduwilt and Moasser, 2008). The first mAb to enter clinical use was the ErbB2-targeting mAb Trastuzumab (Herceptin) by binding to the domain IV, followed by Pertuzumab which interferes with dimerization domain II region of the ErbB2 receptor. Cetuximab (Erbitux) and panitumumab (Vectibix) are other antibodies which target EGFR receptor. Inhibition of the kinase domains to block the activation downstream signaling of ErbB receptors opened doors to another class of targeted therapies, tyrosine kinase inhibitors (TKIs). Gefitinib (Iressa), erlotinib (Tarceva), and lapatinib (Tykerb) bind within the kinase domain of ErbB receptors, with gefitinib and erlotinib most active against EGFR and lapatinib equally active against EGFR and HER2 (Stamos, et al., 2002; Wood, et al., 2004). The impressive response of these drugs when used as a single agent are unfortunately balanced by the resistance

developed by the cancer cells. Combinations of targeted agents which simultaneously inhibit multiple pathways, suppressing feedback reactivation of compensatory signaling networks can therefore prevent recurrence from resistance in cancer cells (Yan, et al., 2011). Combination studies on Trastuzumab and Pertuzumab with Docetaxel has shown significant antitumor effect in an HER2-positive metastatic breast cancer compared to each antibody when used as an individual agent (Baselga, et al., 2012; Harbeck, et al., 2013; Swain, et al., 2015). Understanding the lethal mechanisms within drug combinations, the use of intermittent dosing, immuno-therapy without increasing toxicity and adaptive trial designs can greatly accelerate the development of more effective therapy combinations to improve patient care (Al-Lazikani, et al., 2012; Lopez and Banerji, 2017).

2.5 Summary

The biology of ErbB signaling and efforts at reverse engineering the pathway have been reviewed. As the reductionist approach has well reached its limit, “emergence" paradigm which complements "reduction", needs to be explored. Can the bottom-up approach inherit the legacy of reverse engineering? This is the subject of the forthcoming chapters.

3

Re-engineering ErbB signaling using Agent-based modeling

Re-engineering is closely related to reverse engineering and often used interchangeably. While both refer to the further investigation or engineering of products and systems, the methods of doing so, and the desired outcomes, are vastly different. Reverse engineering attempts to discover how something works, while re-engineering seeks to improve a current design by investigating particular aspects of it. In the computational architecture proposed here, re-engineering begins from the point where reverse engineering exits. Re-engineering with ABMs, banking on the results of reverse engineering is the main contribution of this work.

Sections 3.1 and 3.2 explain re-engineering in general and its ABM implementation. The characteristics and applications of ABM are discussed in Sections 3.3 and 3.4. The large number of software tools available for quick design and development of ABMs are listed.

3.1 Re-engineering products, processes and systems

Re-engineering refers to redesigning or restructuring by exploiting the existing information available. Business analysts have, for several years, preached about the concept of 're-engineering' or, 'business process re-engineering' leaving its mark on the pharmaceutical industry, especially in the area of research and development (Scypinski, et al., 1996). Re-engineering has also been leveraging the design and business processes of manufacturing companies (Kusiak, et al., 1994; Strandhagen and Skarlo, 1995) and robotics (Kim, et al., 2005).

Complexities in drugs targeting diseases such as cancer and autoimmune disorders, global warming, food shortages have shifted the thinking to the field of synthetic biology opening limitless potential to redesign biological systems (Way, et al., 2014). Koide, et al. (2009) focusses on converging the two disciplines, systems biology and synthetic biology for rational biological systems re-engineering. By rewiring regulatory circuits to couple new enzymes, sensors and transporters, a cell

can be re-engineered to possess control, robustness, reliability and predictability over natural organisms.

The emergence of systems re-engineering will bring a dramatic shift in the way one understands and responsibly uses biology, be it on individual cells, communities, ecologies or infectious diseases (Koide, et al., 2009). Here, the knowledge gained from reverse engineering is used to re-engineer a similar system or its variants using an Agent Based Model (**ABM**) offering opportunities to answer fundamentally new biological questions.

3.2 Re-engineering with Agent-based modeling

The history of ABMS has been a rich one for at least 40 years. Traditional mathematical approaches assume systems to be population-based, relating observables to each other via equations that may either be algebraic, or capture variability temporally (ODE) or spatiotemporally (PDE). Even though they are accurate while simulating, they fail to provide detailed solutions when applied on a heterogeneous system. Agent-based modeling is a class of algorithmic computational approaches that recognises the heterogeneity condition and being inherently hierarchical, are quite amenable to coupling with other computational paradigms. The agent-based paradigms can model multi-scale complexities and trace the emergent characteristics of the system (Kaul and Ventikos, 2015).

3.2.1 Structure of Agent Based Model

Before construction of an agent model, it is necessary to define the building blocks or primitives of the model and how they operate to simulate the system. A typical ABM has three elements (Figure 3.1):

1. Agents: A set of entities comprising the system, their attributes and goals.
2. Relationships: A set of rules that define agent behaviour for achievement of goals.
3. Environment: Agents dynamically interact with their environment and with other agents.

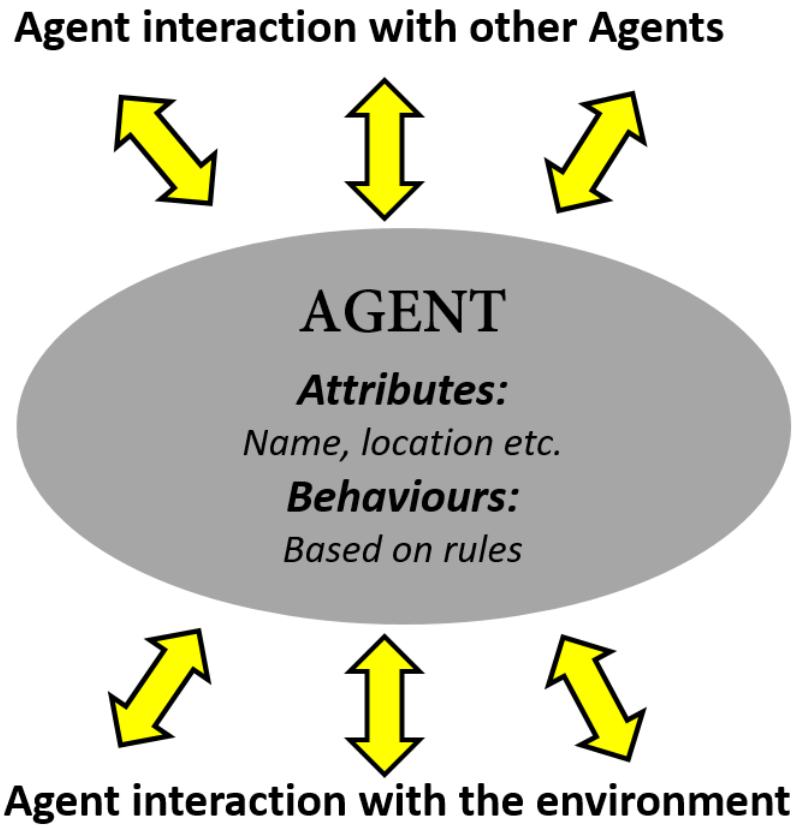


Figure 3.1 Structure of an agent

3.2.1.1 Agents

Agents are autonomous entities that interact with other agents and its environment. Autonomy refers to the agent behaviour and actions without any direct external control. Agents, though a discrete, diverse and heterogeneous entity, share an ability to adapt and modify their behaviour based on their environment (Politopoulos, 2007).

An agent is defined by its attributes which distinguishes it from other agents. Attributes can either be static or dynamic. Static refers to attributes which are not changeable, whereas dynamic attributes are changeable.

The following are the essential characteristics of an Agent (Macal and North, 2010):

- Self-contained: Agents are modular having attributes which helps to distinguish them from other fellow agents.

- Autonomous: Agents can function independently by interacting with other agents and its environment. An agent has behaviours, a set of rules that can relate information to agents decision and action.
- State: An agent has a state which consist of a set of attributes that are associated with its current situation. The behaviour of an agent is conditioned on its state. The state varies over time. In an agent-based simulation, the state at any time plays a pivotal role in providing the information needed to move the system forward.
- Social: Agents can have dynamic interaction with other agents based on a protocol which influence its behaviour.
- Adaptive: An agent can learn and adapt by modifying its behaviour based on accumulated experiences.
- Goal directed: Agents in the system are goal directed and capable of evolving by interacting repetitively, allowing unanticipated behaviours to emerge. Goals allows an agent to compare the outcome of its behaviours relative to its goals and adjust its responses for its future behaviour.
- Heterogeneous: Agent simulations often consider full range of agent diversity across the population. Agent characteristics and behaviours vary further differentiating the agents.
- Flexible: ABM can be considered more flexible by including new agents or attributes into the system.
- Stochastic: Most of the biological systems behave randomly. ABM enables to simulate the system similar to a biological system as each agent decides on its own fate.

3.2.1.2 Relationships

Agents interact with other agents termed neighbours, just as in real world systems. How agents are connected to one another is termed as the ABM topology. The typical topologies can be classified as spatial grid, network of nodes, links and mixed models. Originally, spatial ABMs were implemented in the form of cellular automata (CA)

where each agent interacts in a grid or lattice environment. Agents roam in two, three or higher dimensional spaces in the Euclidean space model. Network topology allows an agent's neighbourhood to be defined more generally with pre-specified static networks or dynamic networks that change according to the programs. In the geographic information system (GIS) topology, agents move from patch to patch over a realistic geo-spatial landscape. Locations are not important in an aspatial or "soup" model. Many ABMs also include agents interacting in multiple topologies. In rule-based models, agents interact based on the set of rules they are provided with.

3.2.1.3 Environment

An environment provides information on the spatial location of an agent relative to other agents. It can be included as a dynamic attribute which can track the movements of an agent. Compartmental models define agents as present in various compartments or moving across these compartments. The environment can also constrain agent actions (Macal and North, 2010).

3.2.2 How ABMs work

Agent-based modeling is one amongst the most powerful spatio-temporal simulation techniques; where the system modelled comprises of autonomous decision-making entities called agents. Agents interact with their fellow agents and environment based on a certain set of rules (Bonabeau, 2002). Based on these set of rules, the individual behaviour of the agents with one another and their shared environment are formalised by the model (Perez-Rodriguez, et al., 2015). Behaviours affect the agent's own actions, the actions of other agents and the environment. Rules which govern these behaviours are constructed by the investigator consistent with either theory or observed data of the system they represent (Griffin, 2006).

Figure 3.2 summarizes the concepts involved in the working of a typical ABM. ABMs are not inductive models that start with a set of data and make inferences with respect to the mechanisms that might lead to that data. Rather, they start with a set of rules that seek to reconstruct the observed pattern of data (An, et al., 2009). By modeling systems as individual agents, the full effect of heterogeneity that exists among agents in their attributes and behaviours can be observed. Interactions among these agents give rise to the behaviour of the system as a whole. Patterns, structures,

and behaviours emerge by self-organization of systems 'ground up'—agent by-agent and interaction-by-interaction (Macal and North, 2010). The two distinguishing features of agent-based simulation as compared to other simulation techniques such as discrete event simulation and system dynamics is its emphasis on heterogeneity of agents across a population and the emergence of self-organization. Therefore, modeling has the potential to replicate cellular systems at its minimum components and thus understanding the linkage from molecular level events to the emerging behaviour of the system (Perez-Rodriguez, et al., 2015)

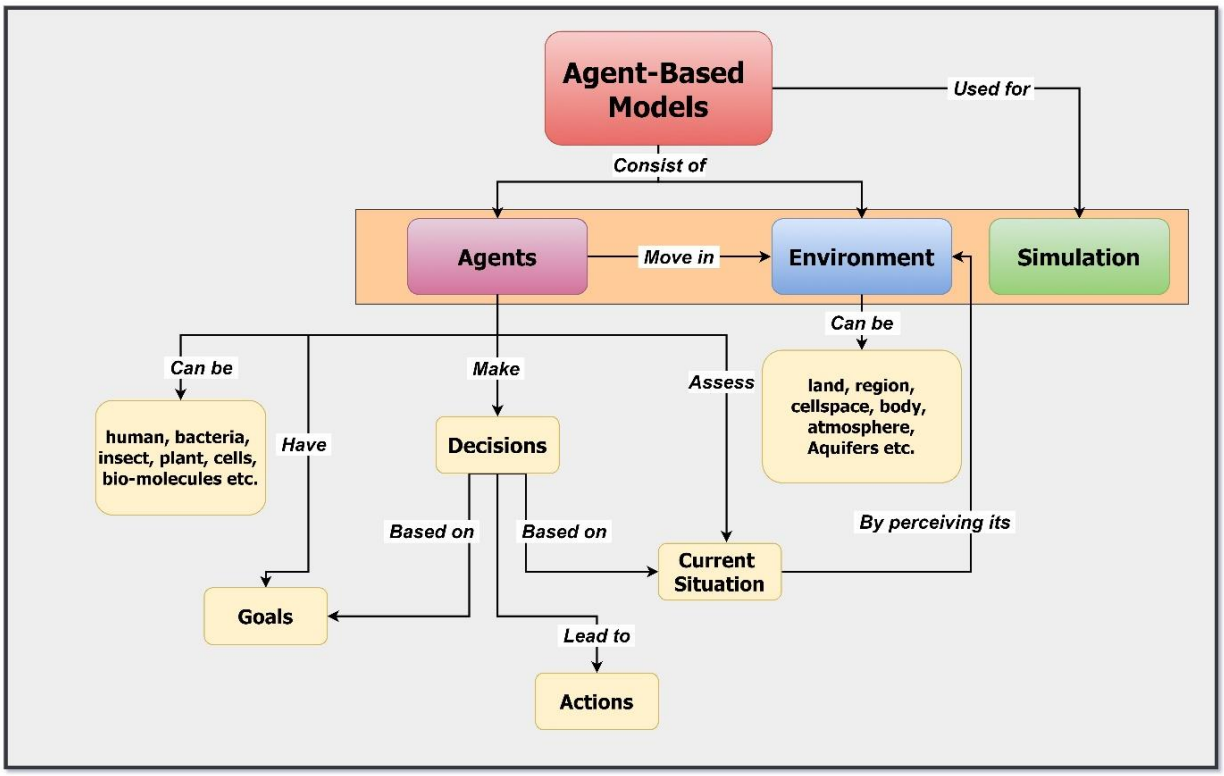


Figure 3.2 Working of a typical ABM

3.3 Characteristics of an ABM

The ABM offers an alternative to traditional equation-based modeling with ordinary or partial differential equations, Discrete event simulation (DES), Monte Carlo simulation, system dynamics, continuous simulation and combined DES/continuous simulation (Macal and North, 2010).

In traditional simulation methods (e.g., equation-based models), the investigator constructs equations whose outcomes vary, based on the variable parameters embedded in the formulae. DES (such Petri Nets or Neural Networks) lays emphasis on models that focus on process or activity as the fundamental simulation element. Individual variability cannot be handled easily in these models. Markov models assume that the probability distribution of future states depends only on the present state and, thus, these models have significant limitations when risk factors and outcomes of the system exhibit complex properties (Li, et al., 2016). 'Top-down' view is often taken in System Dynamics modeling in which a system is broken down into its constituent level components. On the contrary, ABM systems are built 'bottom-up' and system properties emerge from individual autonomous entities (agents) and their interactions.

The agent actions and interactions in ABMs are driven by rules. However, ABM differs from other rule-based systems in the following ways (An, et al., 2009):

- ABM incorporates space: Allows spatial representation of the structural relationships within a system thus supporting to model agents with "bounded knowledge". As a result, it is easier for non-mathematicians to model fairly complex topologies.
- Parallelism: Heterogeneous behaviour of an individual agent within a population of agents that interact in parallel processing environment results in aggregated system dynamics. This can lead to sophisticated level of system behaviour.
- Stochasticity: Behaviour of a single agent depends on the probability function which is incorporated into the agent's rules. This enables in producing system behavioural spaces consistent with population-level biological observation.
- Modularity: New information can be added to a system either through introduction of new agent-types or by the modification of the existing agent rules without having to re-engineer the entire simulation.

- Emergence: ABM generates system dynamics that could not have been reasonably inferred from examination of the rules of the agents alone.
- Incomplete knowledge: ABMs can be constructed even in the absence of complete knowledge, keeping the rules as simple and verifiable as possible, even at the expense of some detail.

Biological systems include multiple levels of organization (gene → protein/enzyme → cell → tissue → organ → organism) with specialized research domains having evolved with focus on the processes at each level (An, et al., 2009). ABM reaches out to reproducing the multi-scales realistically by integrating the mechanistic information generated at one level with concurrent parallel processes to produce recognizable phenomenological behaviours of the greater system as a whole. An approach proposed to represent multiple scales is to view sequential levels from an anatomic standpoint like the Basic Immune Simulator (BIS). BIS is an agent-based model created to study the interactions between the cells of the innate and adaptive immune system (Folcik, et al., 2007).

In addition, ABMs can incorporate phenomena occurring at different timescales. Population-level ordinary or partial differential equations can be used to represent soluble factors or small-scale molecules, and these can be solved on a per-agent basis within an ABM. Modeling these small-scale factors through population approaches instead of explicit agents can save considerable computational power. Furthermore, such multiscale modeling reduces the computational expense incurred if the highest resolution timescale were to be used for each phenomenon (Cosgrove, et al., 2015).

3.4 Applications

Agent-based modeling applications span a broad range of areas and disciplines spanning physical, biological, social and management sciences. The broad range of applications in ABM have been made possible by advances in the development of specialized agent-based modeling software, the availability of data at increasing levels of granularity, and advancements in computer performance.

Sugarscape model by Epstein and Axtell (1996) consists of agents that wander around in harvesting the greatest amount of sugar they can find. The environment where the agents are present is represented by a two-dimensional grid which contains sugar in some of its cells, hence the name Sugarscape.

An ABM model based on simple rules could successfully reproduce the behaviour of a real biological system. In biological sciences, agent-based modeling is used to model cell behaviour and interaction, the workings of the immune system, tissue growth, and disease processes. Walker, et al. (2004) developed The "Epitheliome", is probably one of the earliest applications, representing a rule-based model for growth and repair characteristics of epithelial tissue. Here, each cell acts as an agent and executes a set of rules according to its differentiation state, position in the cell cycle, and the immediate environment.

Applications that relate to cancer development are spatio-temporal 3D computational models of tumor growth dynamics; where each agent is a tumor cell and its environment being EGFR gene-protein interaction network and cell cycle subsystem (Zhang, et al., 2008). Wang, et al. (2013) proposed a 3D multi-scale agent-based cancer model by integrating a novel angiogenesis module into a melanoma tumor growth module. ABMs that incorporate different conditions that give rise to cancer stem cells, helps physicians to predict optimal dosage and frequency of chemotherapy that will be effective (Wang, et al., 2015).

BSim, a customisable ABM tool has been employed for building and characterizing bacterial populations (Gorochowski, et al., 2012) and their chemotaxis processes (Emonet, et al., 2005). Agent-based modeling applications are much more common in the study of infectious diseases as they have a clear path of disease transmission characterized by nonlinear, stochastic, and dynamic interactions between human beings and the environment.

(Lee, et al., 2010) developed an ABM using C++ for the Washington DC metropolitan region to design vaccination allocation strategies of the H1N1 influenza. Perez and Dragicevic (2009) proposed a model which realistically represents spatiotemporal spread of a communicable disease in an urban environment by integrating geographic information systems (GIS).

Agent based approaches have helped to predict the prevention and treatment of chronic disorders such as diabetics, cardiovascular diseases and obesity to improve population health management and medical decision-making (Li, et al., 2016). Immune Simulator is built on a general agent based framework to model the interactions between the cells of the innate and adaptive immune system (Folcik, et al., 2007).

ABM ideally provides valuable insights into the mechanisms for crowd behaviour; panic and jamming during a crisis like a fire or stampede. This suggests practical ways for minimizing the causalities of such events and the existence of an optimal escape strategy. TRANSIMS, ResortScape and SIMSTORE are ABM software packages for Traffic simulation, customer behaviour in a theme market and supermarket respectively. ABMS also offers a promising area for markets and economy by constructing real-world systems, for example shopbots; automated internet agents that compare price and quality of goods and services based on the interest of the buyer. NASDAQ simulates impact of regulations on financial markets under various conditions (Bonabeau, 2002).

(Troisi, et al., 2005) applied agent-based simulation to model molecular self-assembly which shows its applications in the field of physical sciences. Such agent-based modeling approaches have found use in investigating pattern formation in the self-assembly of nano-materials, in explaining self-organized patterns formed in granular materials.

Issues

- Simulating the behaviour of agents can be computation intensive and time consuming when it comes to modeling large systems (Bonabeau, 2002).
- The development of ABM generally requires a large amount of individual-level data for parameterization, calibration, and validation; such data are sometimes not always available to researchers (Li, et al., 2016).
- It is sometimes extremely difficult to systematically analyze how each part ABM simulation affects the output or behavior of the ABM.
- There is a high computational cost associated with running ABM as each discrete event is carried out iteratively and as the agent attribute demands large memory

space. This bottleneck can be solved by implementing high performance supercomputers and other distributed computing platforms such as using Graphical Processing Units (GPUs) that can significantly reduce simulation run times by parallelising the algorithms.

- Applications of ABM in a social, economic and political background involve issues as they most often involve human agents, with potentially irrational behaviour, subjective choices, and complex psychology (Bonabeau, 2002).

3.5 Software Tools for ABM

A number of different platforms have been developed, in order to assist in the development of ABM. They mainly vary in how much support they provide. Among the most popular ABM toolkits are listed (Berryman and Angus, 2010; North, et al., 2013):

Most of the prior agent-based models were developed using the Swarm (http://www.swarm.org/index.php/Swarmmainpage) modeling software designed by Langton and others to model ALife (Macal and North, 2010). Swarm is a free and open source toolkit with both Objective-C and Java bindings.

Mason (http://cs.gmu.edu/~eclab/projects/mason/) is a Java based, open source, ABM framework that serves as basis for a wide range of multi-agent simulation tasks ranging from swarm robotics to machine learning to social complexity environments (Luke, et al., 2005; Perez-Rodriguez, et al., 2015). Swarm and Mason, offer a set of software libraries that helps to program an ABM model.

RePast3(RecursivePorusAgentSimulationToolkit)(http://repast.sourceforge.net/)is a family of three free and open source agent-based modeling libraries (North, et al., 2013). The three libraries are Java-based Repast J, C#-based Repast .NET, and NQP (Not Quite Python)-based Repast Py. They offers tools for quick construction of agent-based models.

NetLogo (http://ccl.northwestern.edu/netlogo/) is a free and open source agent-based simulation environment that uses a modified version of the Logo programming language. They provide a set of libraries and has a graphical tool for quickly constructing interfaces for ABM.

StarLogo (http://education.mit.edu/starlogo/) is a free, open source library and environment that uses a Java interpreter and interface.

SPARK (Simple Platform for Agent-based Representation of Knowledge) developed by University of Pittsburgh is specifically designed as a biomedical ABM toolkit.

Most ABMs tools available are blackbox packages, as they provide fixed set of rules and are often too restricted to capture the wide range of phenomena that one might want to model (Berryman and Angus, 2010).

ABM is increasingly popular in biology due to its ability to represent multiple scales of system decomposition, intertwine complicated behaviours, and its spatial-temporal implementation. ABM has the potential to replicate cellular systems at its minimum components and to understand the linkage of the molecular events that lead to an emergent behaviour of the system. The spatial nature of ABM emphasizes on the behaviours driven by local interactions that match closely with the mechanisms of stimulus and response observed (Perez-Rodriguez, et al., 2015). Modeling ErbB networks involve agents at micro-scale, where agents are bio-molecules inside the cell. The spatial-temporal agent interactions refer to biochemical reactions that occur within the cell.

3.6 Summary

The characteristics and working of an ABM have been explained. The applications and software tools available to build ABMs have been reviewed. The code for the models developed here, are independent programs which gives the freedom to design algorithms and modify them in future.

4

The Modeling Architecture

Life Sciences has turned into a data intensive science with advancement in sophisticated experimental technologies coupled with increasing capabilities of computing hardware and software. Reverse engineering has worked very well at this data-rich first level. At the second level, this data has to be integrated by re-engineering into a working model. The modeling architecture developed here is more of a paradigm than a methodology, for analytics of the biological big data generated by sophisticated instrumentation and computational power.

Section 4.1 introduces the working principle of the Pipeline Architecture. Section 4.2 illustrates the algorithmic details agent-based re-engineering model which forms the latter part of the pipeline. As a proof-of-concept, an implementation has been carried out on HER1-HER2 receptor activation and trafficking dynamics (Section 4.3). The results of the *in silico* experiments are shown in Section 4.4, followed by the summary of the chapter.

4.1 The Pipeline Architecture

The two main concepts that define the pipeline architecture are reverse engineering and re-engineering. While reverse engineering, an engineer works backwards from a product, disassembling it, in order to understand the processes involved in making it. It is generally used to get the know-how and copy a competitor's product. With the knowledge thus gained, one can, not only re-engineer the product, but can also make alterations and adaptations or even design new products. In systems biology parlance, this translates to reverse engineering by a model-based analysis of the system which generates molecular level data on the biochemical reactions. This data

can be pipelined to create the system *in silico* with an ABM which can be used to study new scenarios and test hypotheses under different cellular conditions.

Two modeling formalisms have been used in sequence to create an implementation of the pipeline architecture. It begins with reverse engineering a biological system using mathematical modeling. The parameters estimated by the mathematical model, along with the factual knowledge available, are used to re-engineer the system by the bottom-up approach of agent-based modeling. Information and process flow through the pipeline architecture is shown in Figure 4.1.

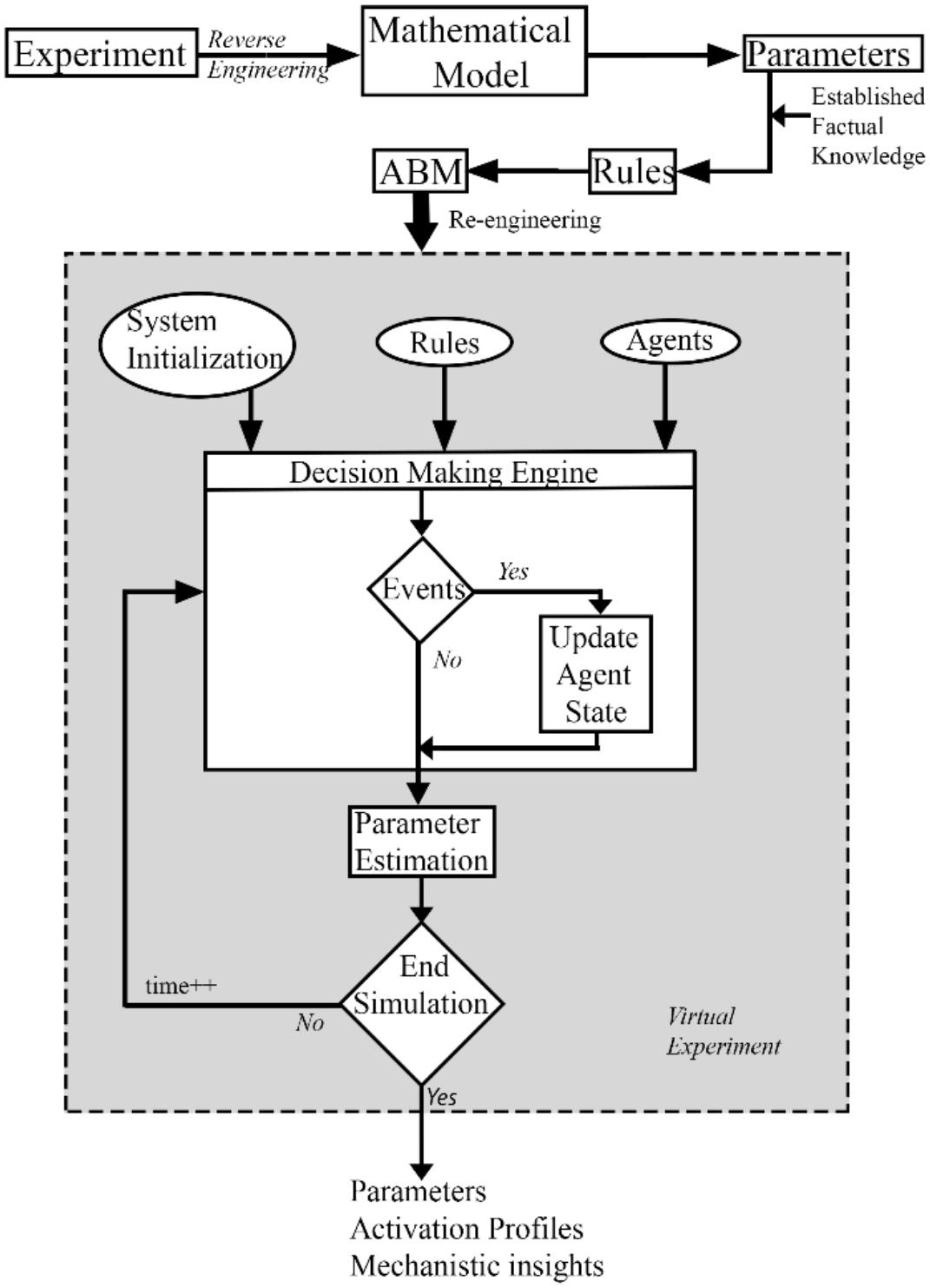


Figure 4.1 Information and process flow through the modeling pipeline

Integration of experimental measurements with computational predictions is a proven way to arrive at the mechanistic details of the system. The reverse engineering step estimates the rate constant parameters of the biochemical reactions, by fitting a kinetic model to the experimental data. These parameters,

together with the established knowledge about the pathway, have been encapsulated into a set of if-then rules for the ABM.

The ABM is a spatio-temporal simulation of the experiment where the system is represented as a set of agents (biomolecules), whose actions and interactions result in the occurrence of various cellular events. The data from reverse engineering, namely, the rate parameters are the primary source of input piped to the agent-based re-engineering model. The agent modeling formalism draws upon this information to drive the local behavior of the biomolecular entities of the system. The decision-making rules capture the course of events that occur during the experiment.

The agent model for re-engineering was developed on a user-friendly framework that provides a template for defining molecular species (agents) and user-defined functions to implement the decision-making logic. The C++ code runs on both Windows (VC.NET) and Linux with R interface for visualization.

4.2 The Agent-based re-engineering model

An ABM contains three main components: 1) agents; 2) the environment or the spatial domain within which the agents exist; 3) a rule-set governing relationships and means of interaction between agents and their environment. The system behavior emerges from the collective interactions of these components over time. A summary of the terminology relating to ABM are illustrated in Appendix I.

4.2.1 Creating the agent world

4.2.1.1 Agents

The system is represented as autonomous, interacting molecular agents, residing in cellular compartments. These agents interact with their fellow agents and environment based on the rules provided by the investigator. Each agent is defined by a set of attributes which include *unique id* generated for the agent, *name* of the agent, *location, check variable* etc. It is then possible to create many agents of the same type, each with different values for its attributes, thus producing a population of individual agents. The agents, characterized by their attributes are stored in a

vector. It is assumed that, in one time step, a maximum of only one event can occur to an agent that can lead to a state change.

4.2.1.2 Compartments

The agents are positioned in compartments, which are determined by the agent attribute–*location*. The agents are allowed to freely interact within a compartment and to also move from one compartment to the other based on their behavior. As the agent leaves the confines of one compartment, their location is set to the new compartment.

4.2.1.3 Rules: Decision makers

The rules are obtained from literature, experimental and computational knowledge. Rules fire events so as to maintain the relative number of molecular species.

The rules are of two types:

i) ***Macro Rules*** which represent the factual knowledge from various experimental and modeling studies. These rules are general in nature.

ii) ***Micro Rules*** focus on the fine details. For e.g., where experimental data is available, rules can be formulated from the reaction rates.

The total time over which the biological system is being simulated is divided into discrete time-steps. At each time-step, an agent must determine how to respond to factors in its environment, using its associated rule-set and current state.

4.2.2 The Algorithm

Input: N←Total agents, Rules, T←time, t← time step=1 unit
Output: Mechanistic insights, Estimated rate values, Quantitative estimations

Main Algorithm

```
Initialization of Agents in compartments; t=0
for each t until T do
        for each Agent until N do
                Find location, type, and state of Agent
                Call Decision Engine ()
        end for
```

Re-calculate reaction rates based on Agent species concentration
Calculate quantitative estimates
Increment t
end for

Decision Engine ()

find the rule subset applicable to the agent
apply all rules in the subset
if all rate rules favour an event **then**
event occurs, update agent state
// *events such as Binding, Dimerization, Internalization, Dissociation and Degradation*
else if there is a conflict condition **then**
randomly choose one event
event occurs, update agent state
else
no event or state change occurs
end if

An algorithm is developed for a time course that agrees with the experiment. For each time step, an agent is selected randomly. The selected agents are free to interact within a compartment. The decision engine drives the simulation by selecting the rules applicable to an agent and deciding which rule to apply so that the rate rules are satisfied. If the rules favor a particular event, an agent is made to update its state. On conflict conditions, an event is chosen randomly. Reaction rates are calculated from actual quantities of species undergoing a state change during the simulation. Average rates are calculated at the end of each time step and local rates are updated within each time step.

The agents, characterized by their attributes are stored in a vector data structure. An agent of type 'a' has n agents (a_1, a_2,....a_n). A Virtual Agent Vector stores the new state of the agents and the original vector gets updated at the end of each timestep (Figure 4.2). It is assumed that, in one time step, a maximum of only one event can occur to an agent that can lead to a state change. An action on one agent may trigger an event on another, leading to a state change for both, e.g. if agent A dimerizes with agent B, then dimer AB is formed. Therefore, asynchronous updation is required, which is achieved by a flag that is set for both agents A and B, so that, B is not examined further during that time step.

During a time step, all agents interact with each other in parallel, but in a computational implementation the agents can be drawn from the agent vector in sequential order only. The order in which agents are processed can introduce bias. A randomized schedule is implemented so that one agent does not get priority over another.

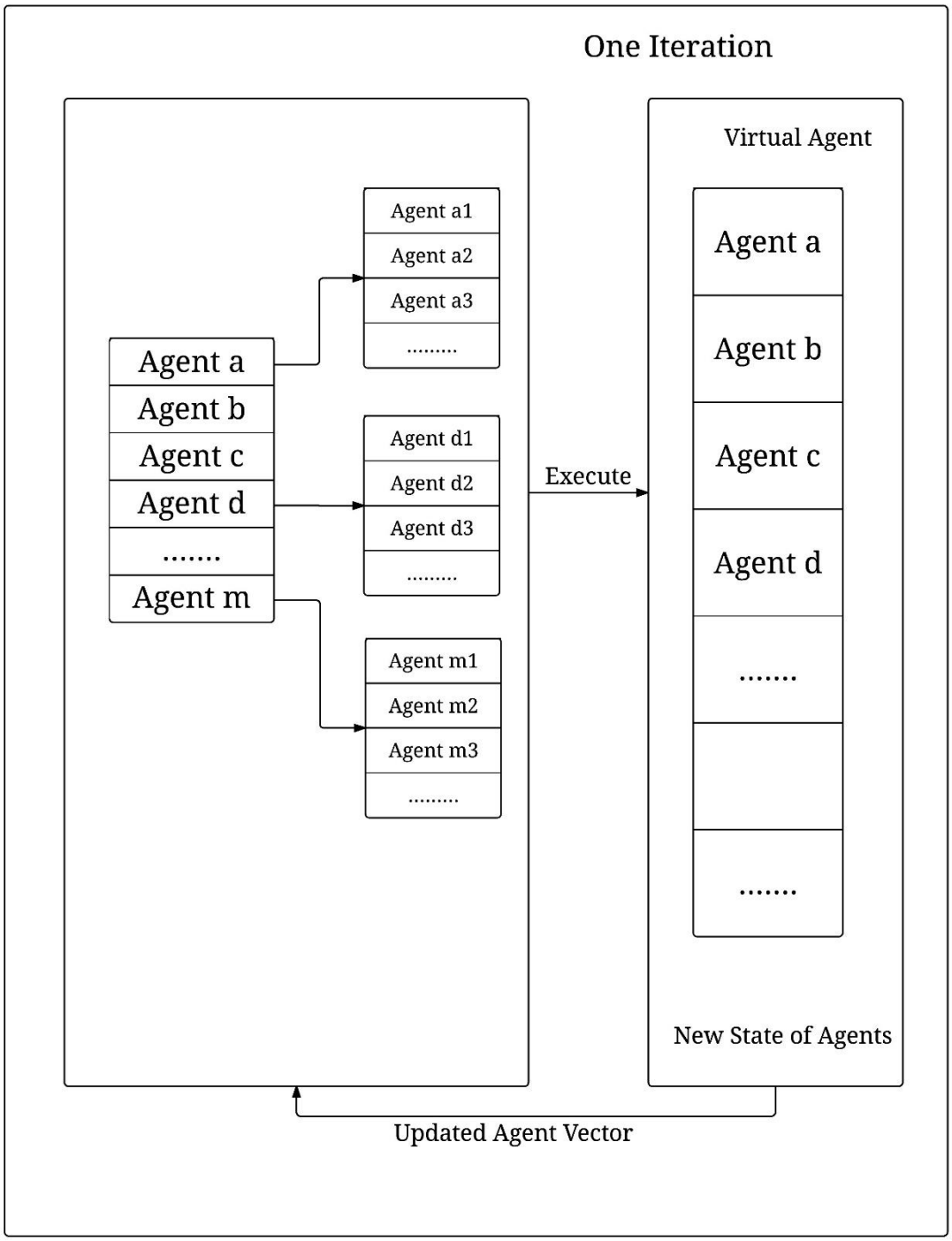


Figure 4.2 ABM Implementation Data Structure

4.3 Implementation: HER1-HER2 receptor activation and trafficking

To demonstrate the pipeline architecture, we focus on signal transduction, in particular, the ErbB signaling pathway. This pathway is one of the most characterized networks with its success largely accredited to the fact that there was prolific activity in both experiments and computational models. The major reason for this being the fact that over-expression of EGFR and HER2 receptors of the ErbB family, have been attributed to the development of many types of human cancers making it an attractive molecular target for therapeutic interventions (Ahmad, et al., 2011; Yarden and Sliwkowski, 2001).

4.3.1 Reverse Engineering with cell line experiments

A mathematical model developed by Shankaran et al. (2008) for activation and trafficking of ErbB receptors, is considered as the reverse engineered system. The model, which will henceforth be referred to as *12Model*, predicts receptor (EGFR and HER2) activation and trafficking by fitting to experimental data. The independent parameters for dimer formation affinities, trafficking rates and relative phosphorylation levels were estimated by the *12Model*. The experimental system considered was that of the HME (Human Mammary Epithelial Cells) parental cell line transduced with HER2 gene to clone other cell lines-expressing low (17L), medium (24H) and high levels (A11H) of HER2 (Table 4.1).

Table 4.1 Receptor abundance in cell lines (Courtesy Shankaran et al. (2008))

List of cell lines		**Abundance of EGFR and HER2 (molecules/cell)**	
Notation	**Description**	**EGFR**	**HER2**
Par	Parental cell line	2.0×10^5	3.0×10^4
17L	Low HER2 expresser	1.8×10^5	1.1×10^5
24H	Medium HER2 expresser	1.6×10^5	6.0×10^5
A11H	High HER2 expresser	3.0×10^5	1.5×10^6

These cells were activated by adding 100 ng/ml EGF to the culture medium and incubating for 5 to 120 min. ELISA assays were done to quantify the receptor mass and phosphorylation levels.

The kinetic network model addresses two processes: (i) ligand binding, receptor dimer formation and phosphorylation, and (ii) receptor trafficking events: internalization and degradation. The system has ten species: receptor monomers (R1: EGFR and R2: HER2) which dimerize and phosphorylate to form active homo and heterodimers (R11, R12, and R22) present, both at the cell surface (s) and internal compartments (i). For example, R_{1s} and R_{1i} are EGFR monomer present on surface and the internal compartment respectively. Monomers interact to form active dimers with a forward rate k_{fs} and dissociate with a reverse rate k_{rs}. Internalized dimers dissociate at a rate k_{ri}. Monomers interact to form active dimers with a forward rate k_{fs} and dissociate with a reverse rate k_{rs}. Internalized dimers dissociate at a rate k_{ri}. Monomers and dimers internalize at a rate k_t and k_e respectively. V_{R1} and V_{R2} are the zero order synthesis rates of EGFR and HER2. The following rate equations describe how the molecular abundances of the ten modeled species change with time in the kinetic model:

$$dR_{1s}/dt = -k_{fs}R_{1s}^2 + 2k_{r11s}R_{11s}^* - k_{fs}R_{1s}R_{2s} + k_{r12s}R_{12s}^* - k_{t1}R_{1s} + V_{R1}$$

$$dR_{2s}/dt = -k_{fs}R_{2s}^2 + 2k_{r22s}R_{22s}^* - k_{fs}R_{1s}R_{2s} + k_{r12s}R_{12s}^* - k_{t2}R_{2s} + V_{R2}$$

$$dR_{11s}^*/dt = \tfrac{1}{2}k_{fs}R_{1s}^2 - k_{r11s}R_{11s}^* - k_{e11}R_{11s}^*$$

$$dR_{22s}^*/dt = \tfrac{1}{2}k_{fs}R_{2s}^2 - k_{r22s}R_{22s}^* - k_{e22}R_{22s}^*$$

$$dR_{12s}^*/dt = k_{fs}R_{1s}R_{2s} - k_{r12s}R_{12s}^* - k_{e12}R_{12s}^*$$

$$dR_{1i}/dt = 2k_{r11i}R_{11i}^* + k_{r12i}R_{12i}^* + k_{t1}R_{1s} - k_{d1}R_{1i}$$

$$dR_{2i}/dt = 2k_{r22i}R_{22i}^* + k_{r12i}R_{12i}^* + k_{t2}R_{2s} - k_{d2}R_{2i}$$

$$dR_{11i}^*/dt = -k_{r11i}R_{11i}^* + k_{e11}R_{11s}^* - k_{d11}R_{11i}^*$$

$$dR_{22i}^*/dt = -k_{r22i}R_{22i}^* + k_{e22}R_{22s}^* - k_{d22}R_{22i}^*$$

$$dR_{12i}^*/dt = -k_{r12i}R_{12i}^* + k_{e12}R_{12s}^* - k_{d12}R_{12i}^*$$

The symbols with star (*) are used to denote phosphorylated (active) molecules. The mathematical model estimates values of 20 unknown parameters for

active receptor dimer formation affinities, trafficking rates and relative phosphorylation levels using non-linear least square regression to simultaneously fit the experimental data collected from the four cell lines.

4.3.2 Agent-based Re-Engineering

The output from the experimental and mathematical model of *12Model* is the input for the Agent-based model.

4.3.2.1 Initialization

The model consists of two types of agents: ligands (EGF) and receptors (ErbB1 and ErbB2). The monomer receptors, on ligand binding interact to form various dimers as shown in Figure 4.3.

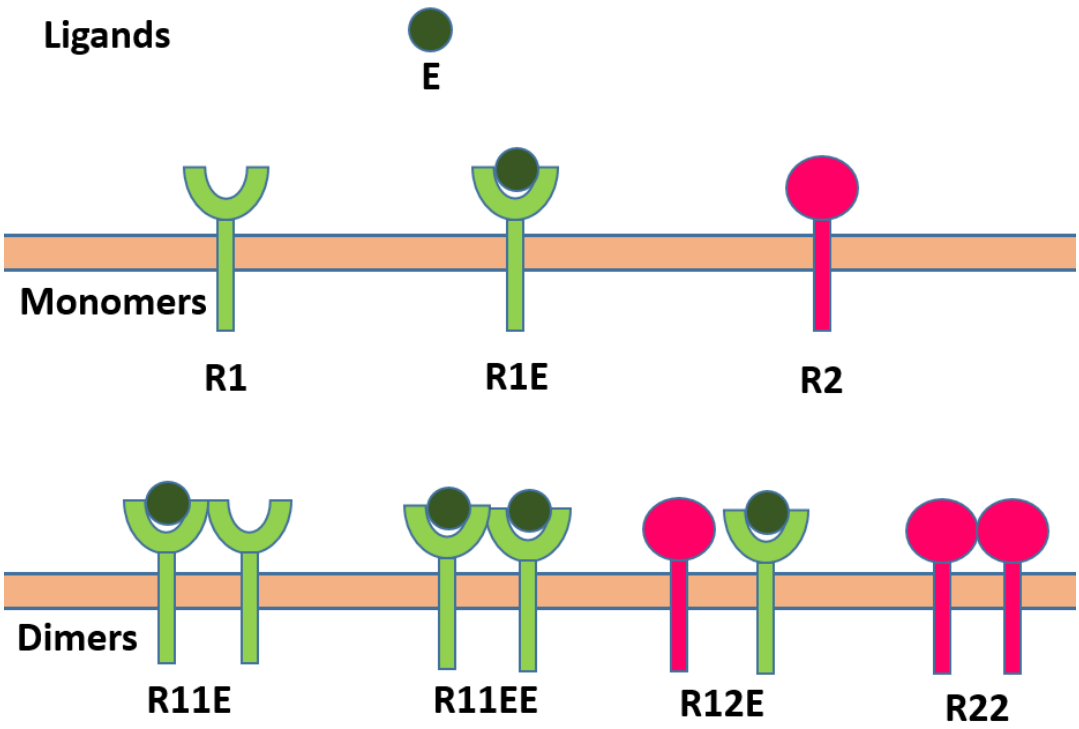


Figure 4.3 Ligand, Monomer and dimer agents. E- EGF, R1-ErbB1 monomer, R1E-EGF bound ErbB1 monomer, R2-ErbB2 monomer, R11E-Single EGF bound ErbB1 homodimer, R11EE-Double EGF bound ErbB1 homodimer, R12E-Single EGF bound ErbB1–ErbB2 heterodimer, R22-ErbB2 homodimer.

Every simulation is a virtual experiment conducted on a cell line. The number and type of agents are initialized to the same value as given in Table 4.1. An agent is defined by attributes like (Figure 4.4):

Ligand Agent	Receptor Agent
unique ligand id	***unique receptor id***
location (Outside, PM)	***location*** (PM,EE,LE)
name (EGF)	***name*** (ErbB1/ErbB2/ErbB3)
bound state (bound/free)	***bound state*** (bound/free)
receptor id (id of the receptor bound to ligand)	***dimer state***(monomer/dimer)
	phosphorylation **status** (Yes/No)
	ligand id (id of the bound ligand)

Figure 4.4 Attributes of Receptor and ligand agent

The 3D spatial domain is divided into five compartments: (1) an extracellular domain (Outside), (2) plasma membrane (PM), (3) early endosome (EE), (4) late endosome (LE) and (5) lysosome. In Figure 4.5, the Outside compartment refers to the exterior part of the cell where the ligand is present. The receptors present on the PM can be in ligand-bound, monomeric or dimeric forms. On internalization, receptors move to the EE compartment from where they are either recycled to PM or sent to LE for degradation in the lysosome. In the course of signal activation, monomers change to dimers and vice versa. This can result in mobility between compartments. The events that bring about these changes are ligand binding, dimerization, internalization, recycling, dissociation and degradation. The input parameters for the simulation are given in Appendix II.

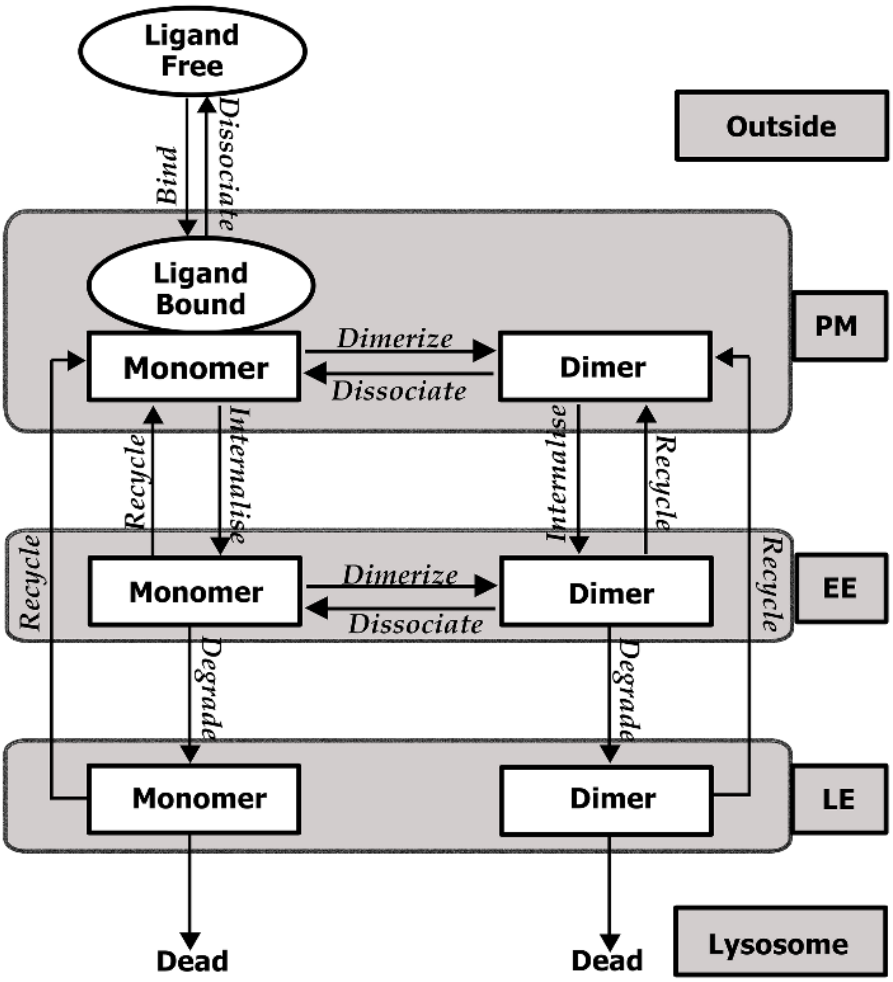


Figure 4.5 Schematic for receptor activation and trafficking

4.3.2.2 Decision Engine

The dynamics of receptor activation and trafficking are determined by the behavior of agents and the interactions between them (Refer flow chart in Appendix III). The behavioral rules have been formulated from i) the established facts about the ErbB pathway and ii) parameters estimated from reverse engineering. Macro rules are derived from the factual knowledge from the experimental and modeling studies. The micro rules are derived from the reaction rates given by the mathematical model fitted to the experimental data. When agents undergo dimerization, internalization, dissociation or degradation, the rates at which they do so, should comply with the micro rules. In our case, micro rules are experiment-specific. These rules, together, simulate the conditions in an actual experiment. Some typical rules are listed in Table 4.2.

Table 4.2 Macro and micro rules for EGFR, HER2 model.

Macro Rules	Micro Rules
(1) Phosphorylation increases with higher receptor expression level. *(2) The peak EGFR phosphorylation levels are similar for Parental, 17L, and 24H, whereas it is ~4 times greater in A11H.* *(3) Dissociation on PM is the highest for HER2-HER2, and the least for HER1-HER1* *(4) EGFR monomers are internalized and degraded slightly faster than HER2.* *(5) EGFR homodimers are internalized much rapidly compared to EGFR-HER2 heterodimers and HER2 homo-dimers.* *(6) EGFR has comparable phosphorylation levels in homodimers and heterodimers.* *(7) HER2 phosphorylation is significantly higher in heterodimers than in its homodimers.* *(8) Dissociation rates in the internalized complexes (EE) is the highest for HER2-HER2,followed by HER1-HER2 and the least for HER1-HER1*	*(1) The ratio of ligand dissociation rate in the internal compartment (EE) to the surface (PM) is in the range 7 to 31.* *(2) The EGFR degradation rate is in the range 0.008-0.33/min.* *(3) Ratio of pf11/pf12 =1.05, Ratio of pf21/pf22 =1.05* *(4) HER2 monomer: degradation is faster than its internalization.* *(5) EGFR homodimer:* *a. Internalization rate is four times the dissociation rate on the PM.* *b. Dissociation on EE is seven times the dissociation on the PM.* *(6) HER2 homodimer:* *a. Internalization rate is 12 times faster than the HER2 monomer.* *b. Dissociation rate on PM is nearly 80 times its internalization rate.* *c. The ratio of dissociation in EE to dissociation in the PM is 11.* *d. Degrades at the same rate as HER2 monomer .* *(7) EGFR-HER2 heterodimer:* a. *Internalizes 8 times faster than HER2 monomer and 3 times faster than EGFR monomer.* b. *Dissociation rate on PM is 10 times its internalization rate.* c. *Dissociation on the PM is 4 times faster than on EE.* d. *The degradation rate is equal to HER2 monomer degradation rate.*

The rules are framed based on the dependency map (Figure 4.6), e.g. internalization of EGFR homodimer (ke11) depends on the internalization of other species (ke12, ke22 and kt1) whereas its dissociation in EE (kr11i) depends only on its dissociation rate at the surface (kr11s). As the number of rules increases, the map becomes complex but, if the dependencies are intuitively defined, the decision engine will need to work only on a small set of rules for each species.

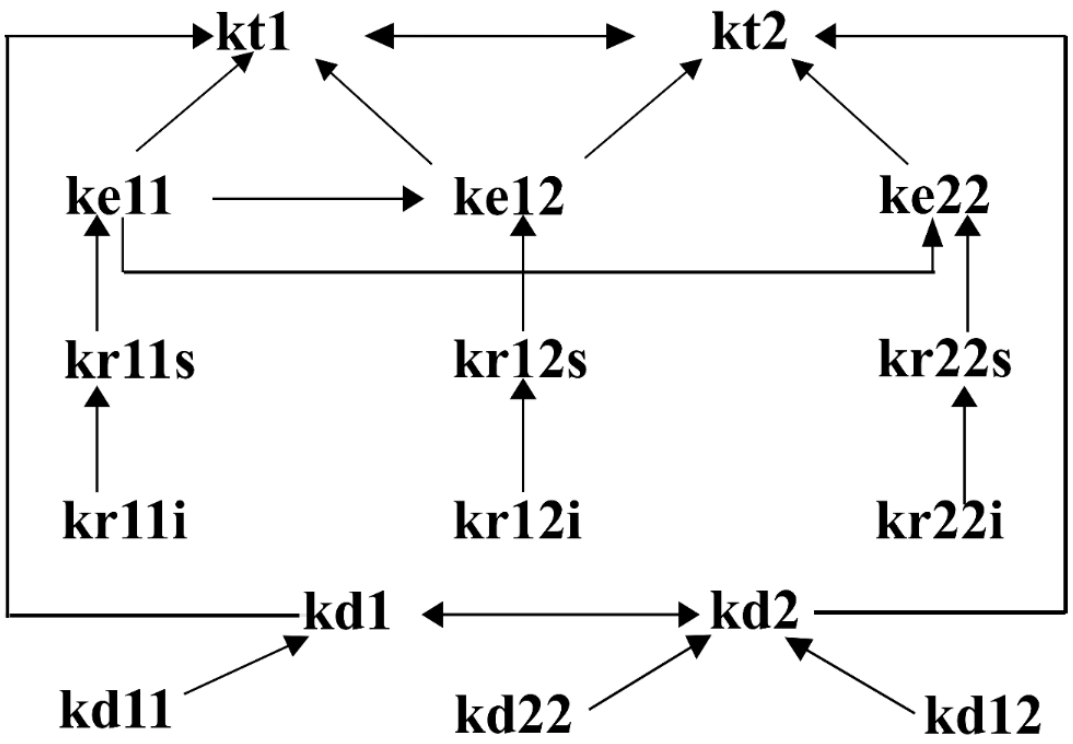


Figure 4.6 Dependency map of rate constants: kr – dissociation rate, kd – degradation rate, kt – internalization rate for monomers, ke-internalization rate for dimers. 's' – surface (PM), 'i' –internal compartment (EE), 1 – EGFR, 2 – HER2, e.g. kr12i is the heterodimer dissociation rate in EE

Examples for implementation of micro and macro rules for the model are given below.

rules_v[] is the rule vector for rate constants; for e.g., rules_v[3] and rules_v[4] are for rule vector EFGR homodimer internalization rate. The following enumerates how the rule vectors are used in our model.

Initial values rules_v[] = 0, it can take the value 1 or -1.

Set value to 1 if rate rule is satisfied or -1 if it shows a reverse trend.

Macro Rule 5 implementation:

Rule (5): EGFR homodimers are internalized much rapidly than EGFR-HER2 heterodimers and HER2 homodimers.

if (ke11_avt > ke12_avt) rules_v[3]=1;

if (ke11_avt <= ke12_avt) rules_v[3]=-1;

if (ke11_avt > ke22_avt) rules_v[4]=1;

if (ke11_avt <= ke22_avt) rules_v[4]=-1;

ke11_avt, ke22_avt, ke12_avt, and kr11s_avt variables refer to average values of EGFR and HER2 internalization and dissociation rate calculated after each time step.

Micro Rule 5a implementation:

Rule: (5) a. EGFR homodimer Internalization rate is four times the dissociation rate on the PM.

if (ke11_avt > (4.0*kr11s_avt)) rules_v[9]=1;

else rules_v[9]=-1;

The pseudocode for the events of activation and trafficking are given in Appendix IV.

4.3.2.3 Parameter Estimation

Reaction rates are calculated from actual quantities of species undergoing a state change during the simulation. The rates are calculated at each time step using the following calculations:

Dimerization rate (kf) on PM and EE

kf1s=c_1_11_p/c_1_p

kf2s=c_2_22_p/c_2_p

kf12s=c_1_12_p/min_c1_c2_p

kf21s=c_2_12_p/min_c1_c2_p

kf1i=c_1_11_e/c_1_e

kf2i=c_2_22_e/c_2_e

kf12i=c_1_12_e/min_c1_c2_e

kf21i=c_2_12_e/min_c1_c2_e

’s’-surface (PM) and ’i’-internal compartment (EE)

c_1_p, c_2_p – Monomers of ErbB1 and ErbB2 on PM (p) and EE (e).
c_1_e, c_2_e

min_c1_c2_p – min(EGFR and HER2 monomer) on PM and EE
min_c1_c2_e

c_1_11_p, c_2_22_p – Monomers of ErbB1 and ErbB2 on PM (p) and EE (e)
c_1_11_e, c_2_22_e

c_1_12_p, c_2_12_p – Monomers of ErbB1 and ErbB2 on PM (p) and EE (e)
c_1_12_e, c_2_12_e

Dissociation rate (kr) on PM and EE

kr11s=c_11_1_p/c_11_p

kr22s=c_22_2_p/c_22_p

kr12s=c_12_1_2_p/c_12_p

kr11i=c_11_1_e/c_11_e

kr22i=c_22_2_e/c_22_e

kr12i=c_12_1_2_e/c_12_e

c_11_1_p, c_22_2_p – Homodimers dissociated to monomers on PM (p) and EE (e)
c_11_1_e, c_22_2_e

c_12_1_2_p – EGFR-HER2 heterodimers dissociated to EGFR and HER2
c_12_1_2_e monomers on PM (p) and EE (e)

c_12_p,c_12_e – EGFR-HER2 heterodimers on PM (p) and EE (e)

c_11_p,c_22_p – Homodimers on PM (p) and EE (e)
c_11_e,c_22_e

Internalization rate (kt,ke)

kt1=c_1_int/c_1_p

kt2=c_2_int/c_2_p

ke11=c_11_int/c_11_p

ke12=c_12_int/c_12_p

ke22=c_22_int/c_22_p

c_1_int,c_2_int –Monomers internalised

c_11_int,c_22_int –Homodimers internalised

c_12_int –EGFR-HER2 heterodimers internalised

Degradation rate (kt,ke)

kd1=c_1_deg/c_1_e

kd2=c_2_deg/c_2_e

kd11=c_11_deg/c_11_e

kd12=c_12_deg/c_12_e

kd22=c_22_deg/c_22_e

c_1_deg, c_2_deg - Monomers degraded

c_11_deg, c_22_deg - Homodimers degraded

c_12_deg - EGFR-HER2 heterodimers degraded

4.4 Simulation Results

To demonstrate the re-engineering process, the biological experiment is simulated using ABM. We investigate, if the knowledge built into the ABM rules can reproduce the experimental results. Assumptions made for the cell environment are:

(1) EGF binding rate and dissociation rates were 0.83/min and 0.12/min respectively.

(2) Average dimerization rate (kfs) was assumed to be 0.2/min for all cell lines.

(3) At the start of simulation, 52% of EGFR and 55% of HER2 are at the cell surface.

(4) Internalization and degradation rates for monomers (kt1, kt2, kd1, and kd2) are initialized to the values obtained from the mathematical model assuming that these events occur even before ligand addition. All other parameters are estimated by the ABM.

(5) Agents are recycled during dissociation in EE, internalization, and degradation. The recycling fractions were taken from literature (Shankaran, et al., 2013).

(6) Only dimerization and degradation occur after 40 minutes.

Given the abundance of EGFR and HER2 in the 4 HME cell lines (Table 4.1), twenty simulation runs were made for each cell line, and the parameter values were

averaged. The simulations were run for 100 min with each time step equal to one minute.

4.4.1 Receptor Activation Patterns

Receptor mass values and phosphorylation levels were quantified by multiplying receptor/dimer abundances with constant factors derived from the measurements in the ELISA assays. The total receptor count determined by the ABM model was translated to quantities (no units), using the following equations obtained from (Shankaran, et al., 2008).

Total receptor mass of EGFR and HER2 monomers on PM (Plasma Membrane) and EE (Early Endosome).

mR1t=R1t***m1**
mR2t=R2t***m2**

Total mass of phosphorylated EGFR and HER2 dimers on PM and EE.

mRP1t= (R11s+R11i+R12s+R12i)***m1**
mRP2t= (R22s+R22i+R21s+R21i)***m2**

Mass of phosphorylated EGFR and HER2 dimers on EE.

mRP1i= (R11i+R12i)***m1**
mRP2i= (R22i+R12i)***m2**

't'-Total present on PM and EE

'i'-Total present on EE (internal compartment).

R1t, R2t - Total monomer abundance

R11s, R22s, **R11i**, **R22i** - Total number of homodimers on the PM and EE

R12s, **R21s**, **R12i**, **R21i** - Total number of EGFR- HER2 heterodimers on the PM and EE

m1 and **m2** are the normalization factors used to convert the receptor abundances expressed in number of molecules to mass values mR measured in their ELISA experiments; m1=0.0000032, m2=0.00000435.

pRP1t= (**pf11***(R11s+R11i)) + (**pf12***(R12s+R12i))

pRP2t= (**pf22***(R22s+R22i)) + (**pf21***(R21s+R21i))

pRP1i= (**pf11***R11i) + (**pf12***R12i)

pRP2i= (**pf22***R22i) + (**pf21***R21i)

The total receptor phosphorylation of EGFR and HER2 on PM and EE is referred as **pRP1t** and **pRP2t**. The phosphorylated receptor levels in EE is given as **pRP1i** and **pRP2i**. The 4 pf (phosphorylation factors) values are the dimer-specific phosphorylation multipliers used to convert abundance of activated receptors to extent of phosphorylation measured in their ELISA experiment; pf11=0.00043, pf12=0.00041, pf22=0.000051, and pf21=0.00023.

The percentage phosphorylation contribution of EGFR homodimer and HER2 homodimer is calculated by **homo1_percent** and **homo2_percent**.

homo1_percent= (100***pf11***R11t)/pRP1t

homo2_percent= (100***pf22***R22t)/pRP2t

The peak phosphorylation levels are attained at about 5 min after ligand addition. The peak EGFR phosphorylation values are similar for Par, 17L and 24H, while it is ~3 times higher for A11H (Figure 4.7). The HER2 phosphorylation levels also increase with receptor expression levels in the cell lines (Figure 4.8).

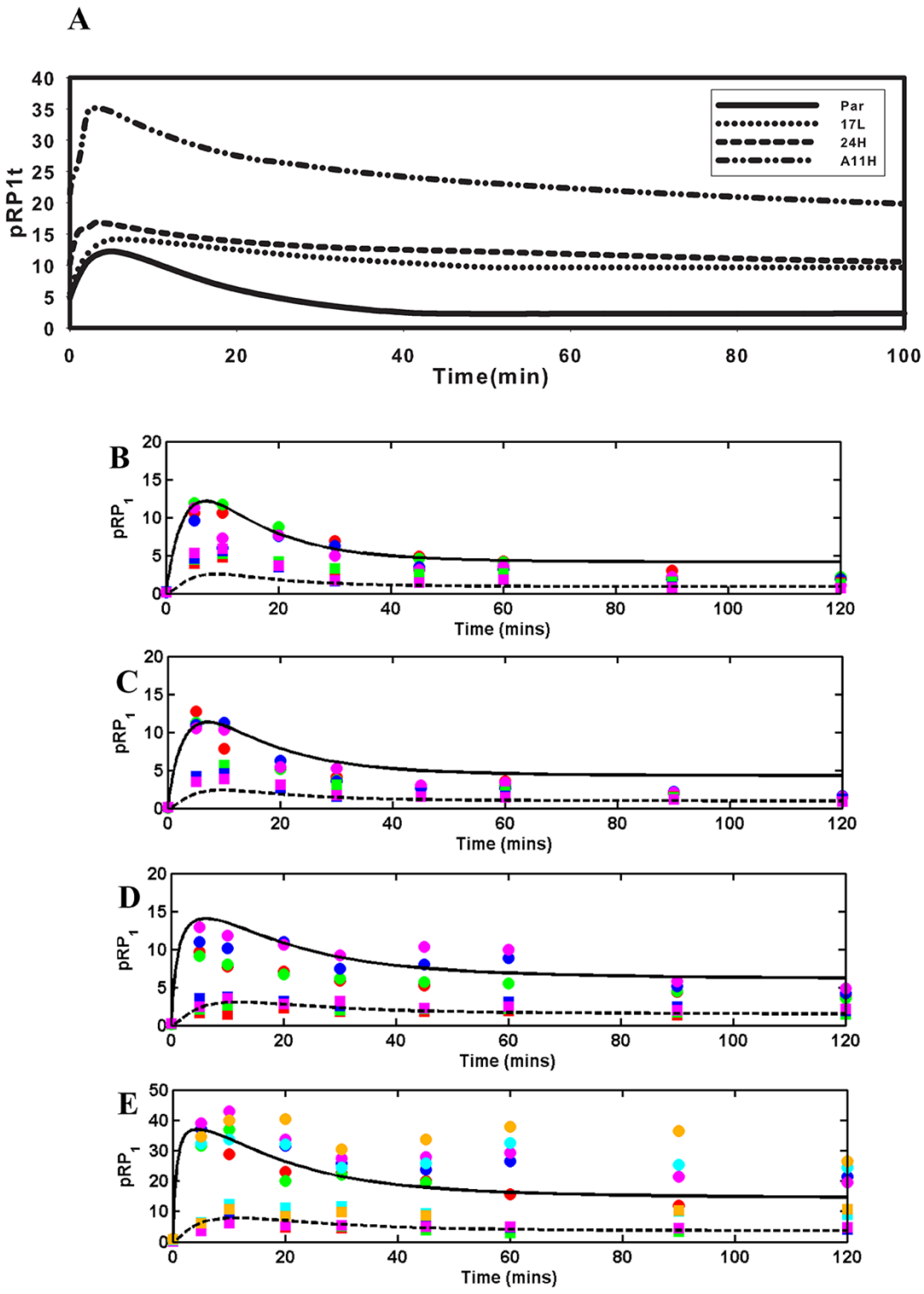


Figure 4.7 Total receptor phosphorylation levels of EGFR in Par, 17L, 24H and A11H cell lines: A- total receptor phosphorylation in the cell obtained from ABM model. B, C, D, E- obtained from mathematical model for total receptor phosphorylation in the cell (solid line) and in the internal compartments (dotted line) fitted to the experimental data points (circles for total cell and squares for internal compartments).Multiple biological replicates are denoted by distinct colors.

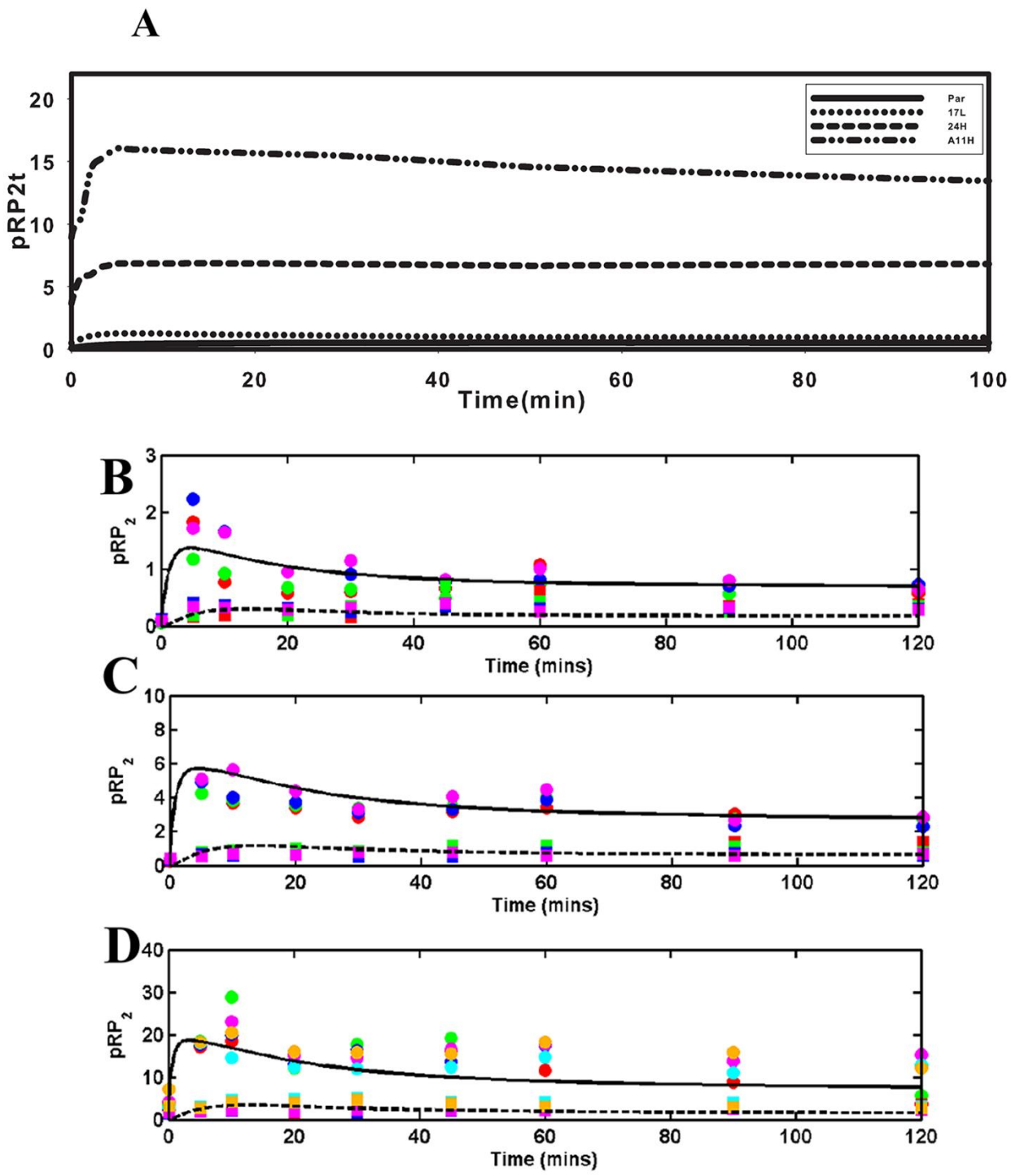


Figure 4.8 Total receptor phosphorylation level of HER2 in Par, 17L, 24H and A11H cell line: A- total receptor phosphorylation in the cell obtained from ABM model. B, C, D - for 17L, 24H and A11H obtained from mathematical model for total receptor phosphorylation in the cell (solid line) and in the internal compartments (dotted line) fitted to the experimental data points (circles for total cell and squares for internal compartments).

The error plots were plotted for EGFR and HER2 phosphorylation for the four cell lines (Figure 4.9 and 4.10). Error bars indicate the standard deviation over 20 simulations.

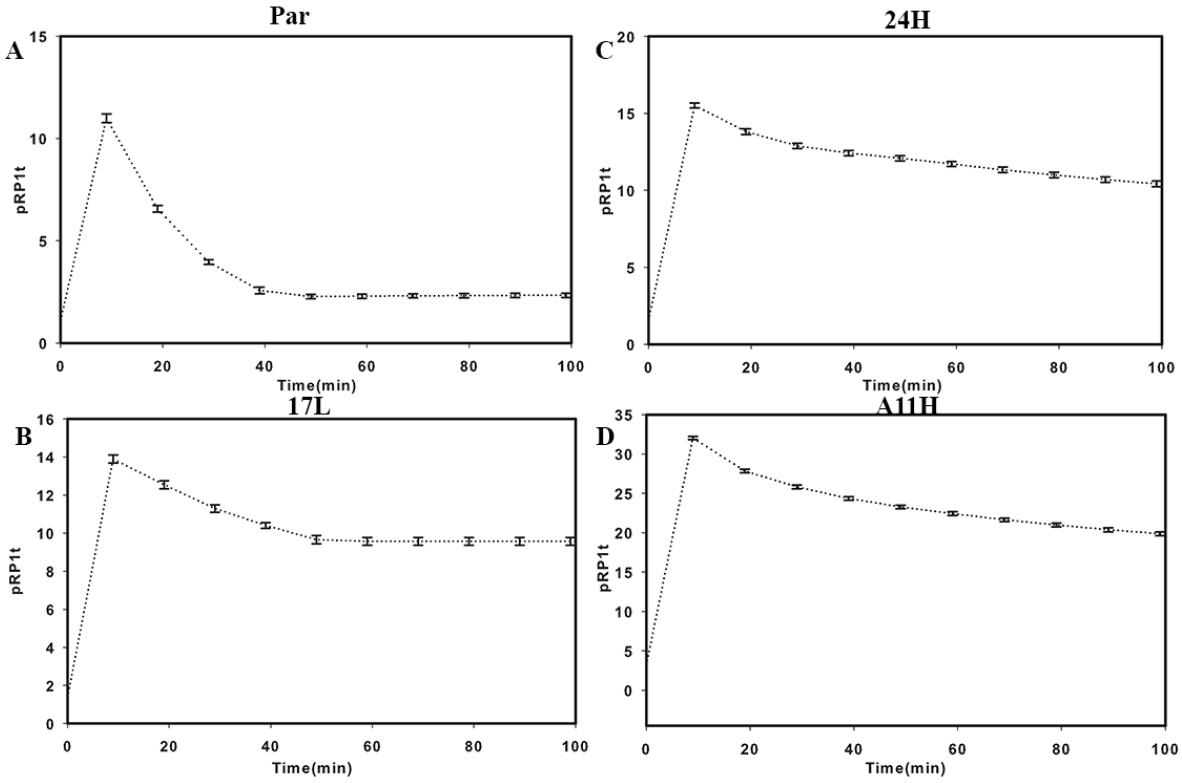


Figure 4.9 Total receptor phosphorylation levels of EGFR in Par, 17L, 24H, A11H with error bars.

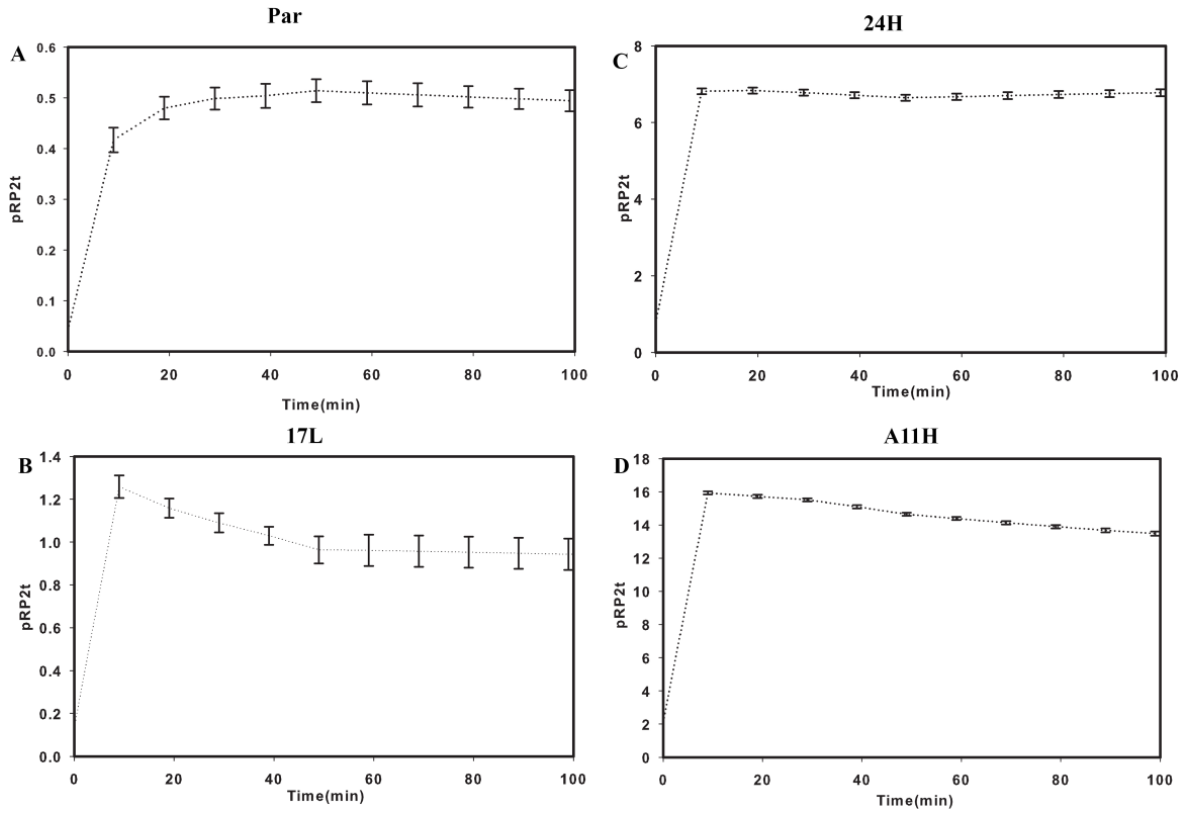


Figure 4.10 Total receptor phosphorylation level of HER2 in Par, 17L, 24H, A11H with error bars.

The HER dimerization pattern is difficult to measure experimentally, but this is not a limitation in the agent scenario, where absolute quantities of each dimer are available throughout the simulation. Figure 4.11A shows that EGFR homodimers account for more than 90% of the total EGFR phosphorylation in the parental cell line. This decreases with increasing HER2 expression in other cell lines due to the formation of heterodimers. In A11H, the hetero-dimer contributions are ~ 70%. HER2 dimers are allowed to form a minute before the beginning of the simulation. The rate at which they are formed is proportional to the fraction of HER2 monomers present in the cell line. With the addition of ligand, HER2 homodimers decrease due to the formation of heterodimers. The contribution of HER2 homodimers increases from 0 in Par to 38% in A11H (Figure 4.12A).

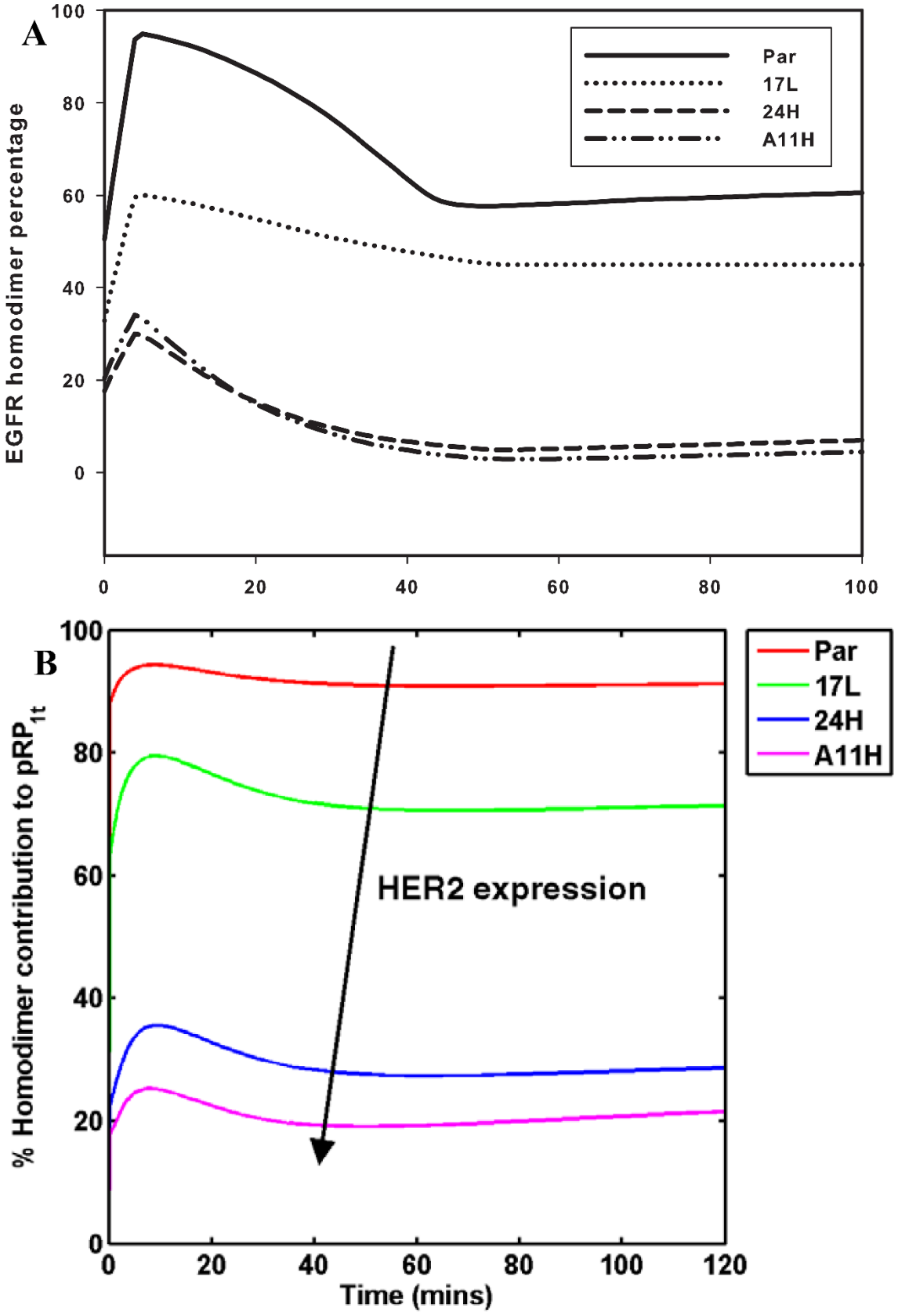


Figure 4.11 Dimer contributions to receptor phosphorylation. A, B fractional contribution of EGFR homodimers towards total EGFR phosphorylation signal using ABM model and mathematical model respectively.

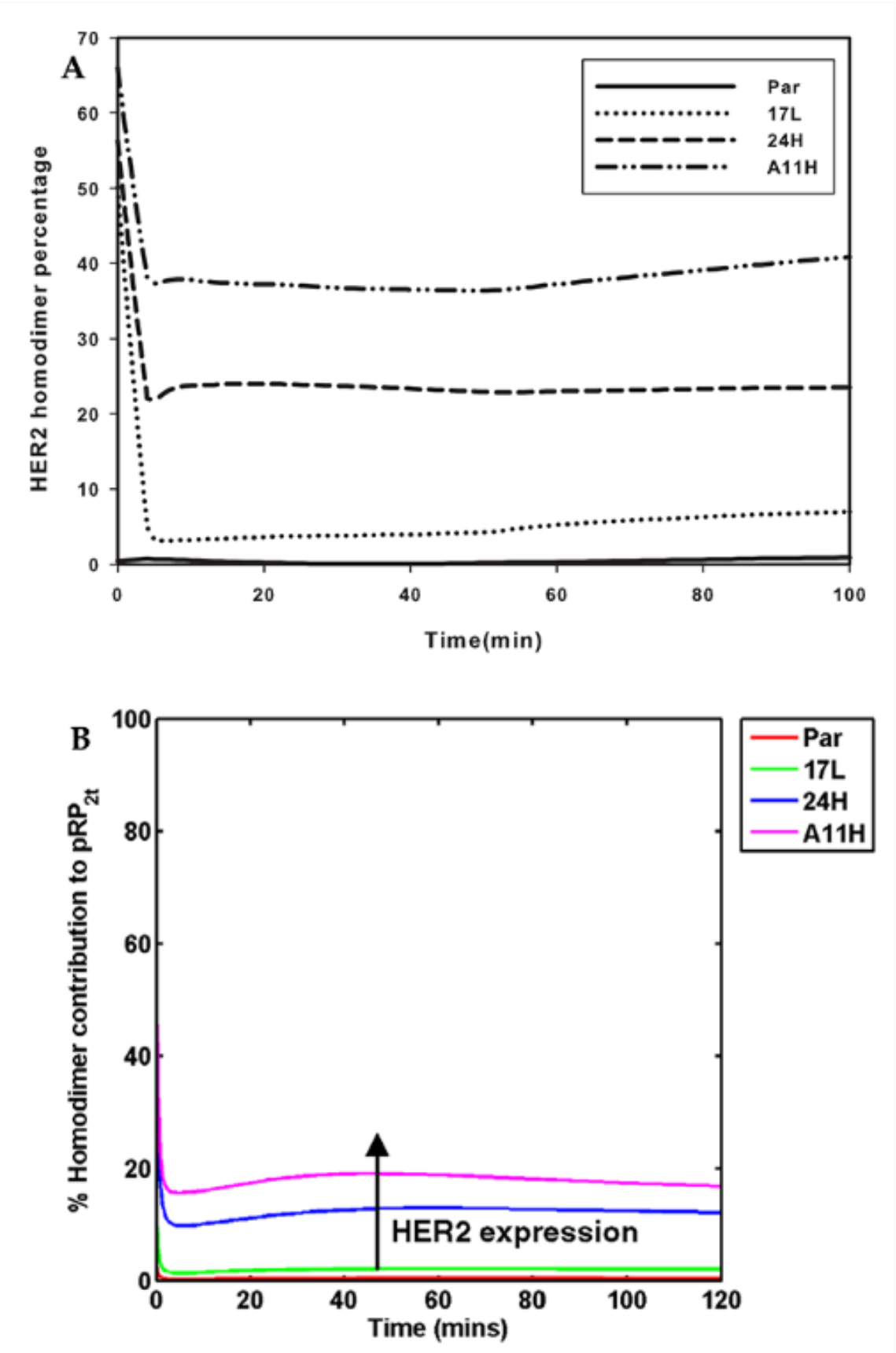


Figure 4.12 Dimer contributions to receptor phosphorylation. A, B- fractional contribution of HER2 homodimers towards total HER2 phosphorylation signal using ABM model and mathematical model respectively.

Comparison of the slopes of the receptor mass curves show HER2 gets degraded at a slower rate than EGFR (Figure 4.13 and Figure 4.14).

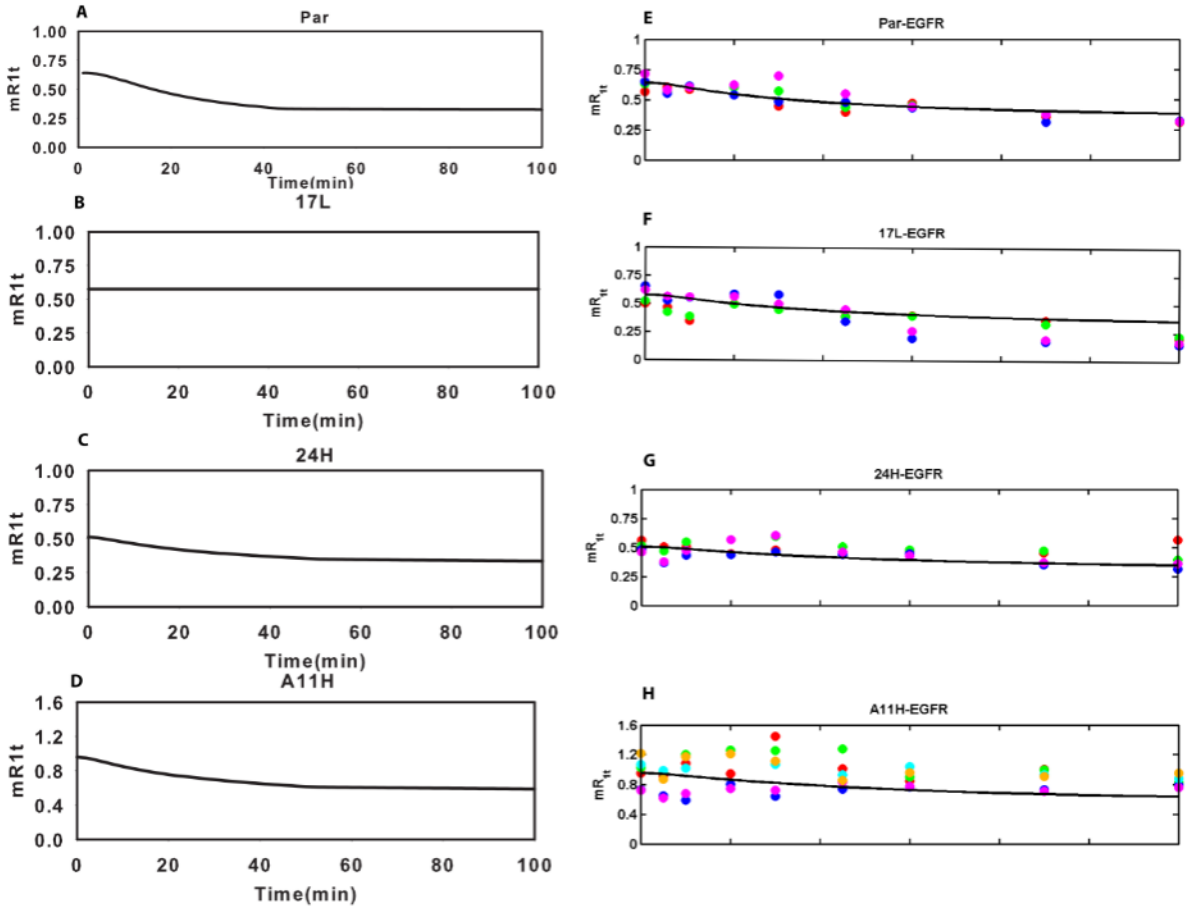


Figure 4.13 Total receptor mass of EGFR in Par, 17L, 24H and A11H cell lines: A, B, C, D - total receptor mass of EGFR in the cell obtained from the ABM model. E, F, G, H - total receptor mass of EGFR in the cell obtained from the mathematical model where the model fits (lines) to the experimental data (points).Multiple biological replicates are denoted by distinct colors.

It is evident that the phosphorylation and dimerization patterns obtained from the ABM in the simulated system of cell lines co-expressing EGFR and HER2 are found to be in agreement with the results of the mathematical model.

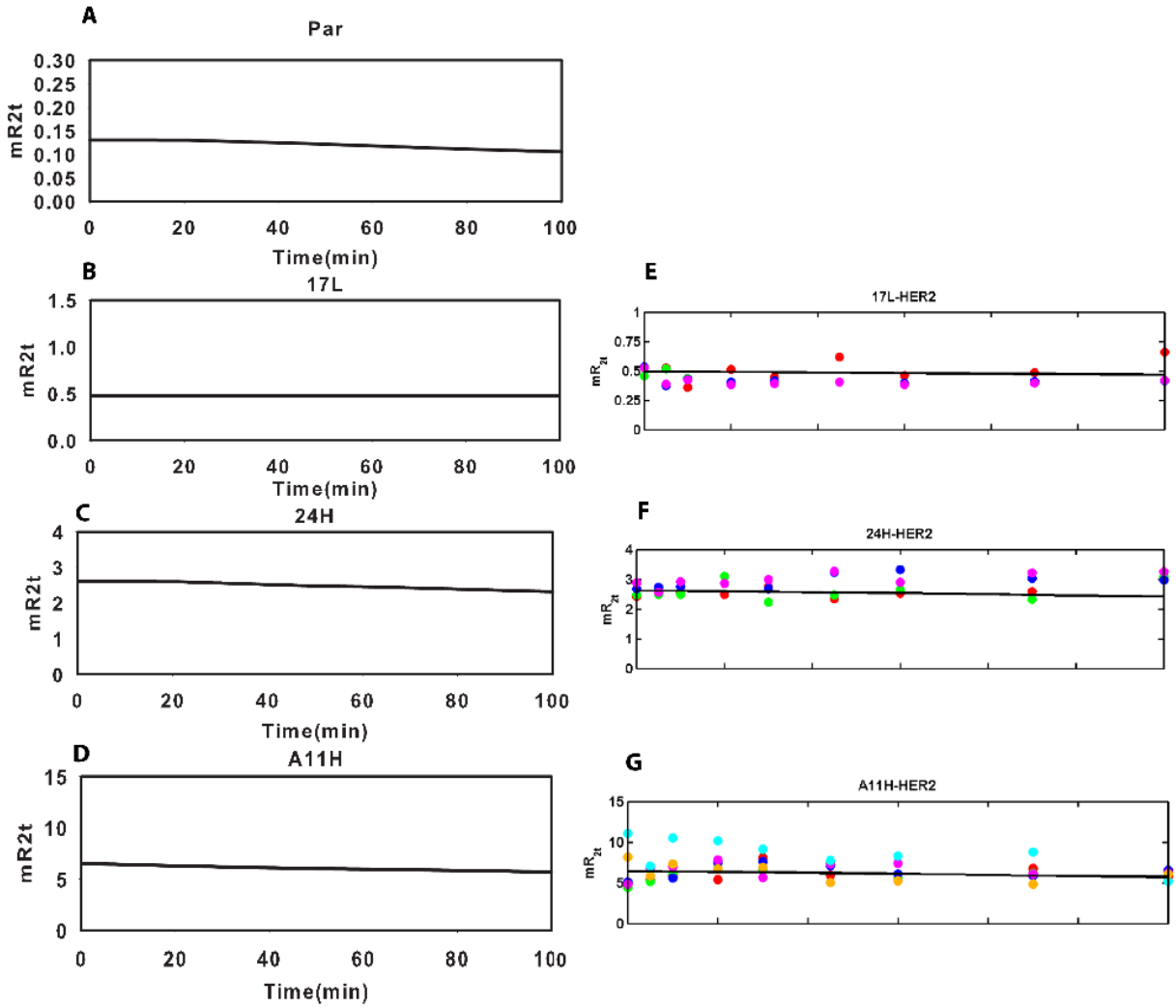


Figure 4.14 Total receptor mass of HER2 in Par, 17L, 24H and A11H cell lines: A, B, C, D - total receptor mass of HER2 in the cell obtained from the ABM model. E, F, G - total receptor mass of HER2 obtained from the mathematical model where the model fits (lines) to the experimental data (points). Multiple biological replicates are denoted by distinct colors

4.4.2 Prediction of Parameter Estimates

The individual-based representation of the system offers a simple way to predict the rates of activation and trafficking events. At the end of each time interval, the quantities of the ten species, namely, two monomers and three dimers located at the surface and in the internal compartment, are re-calculated and values of the rate parameters for activation and trafficking are obtained from actual concentrations of species undergoing state change.

The decision engine calculates the local rates within the time step and updates average rates at the end of a time step. The local rates are set to zero at the start of

each time step. The rates are calculated for dimerization, internalization, dissociation and degradation (Section 4.3.2.3). The rate constants evolve as the simulation progresses (Figure 4.15-4.18).

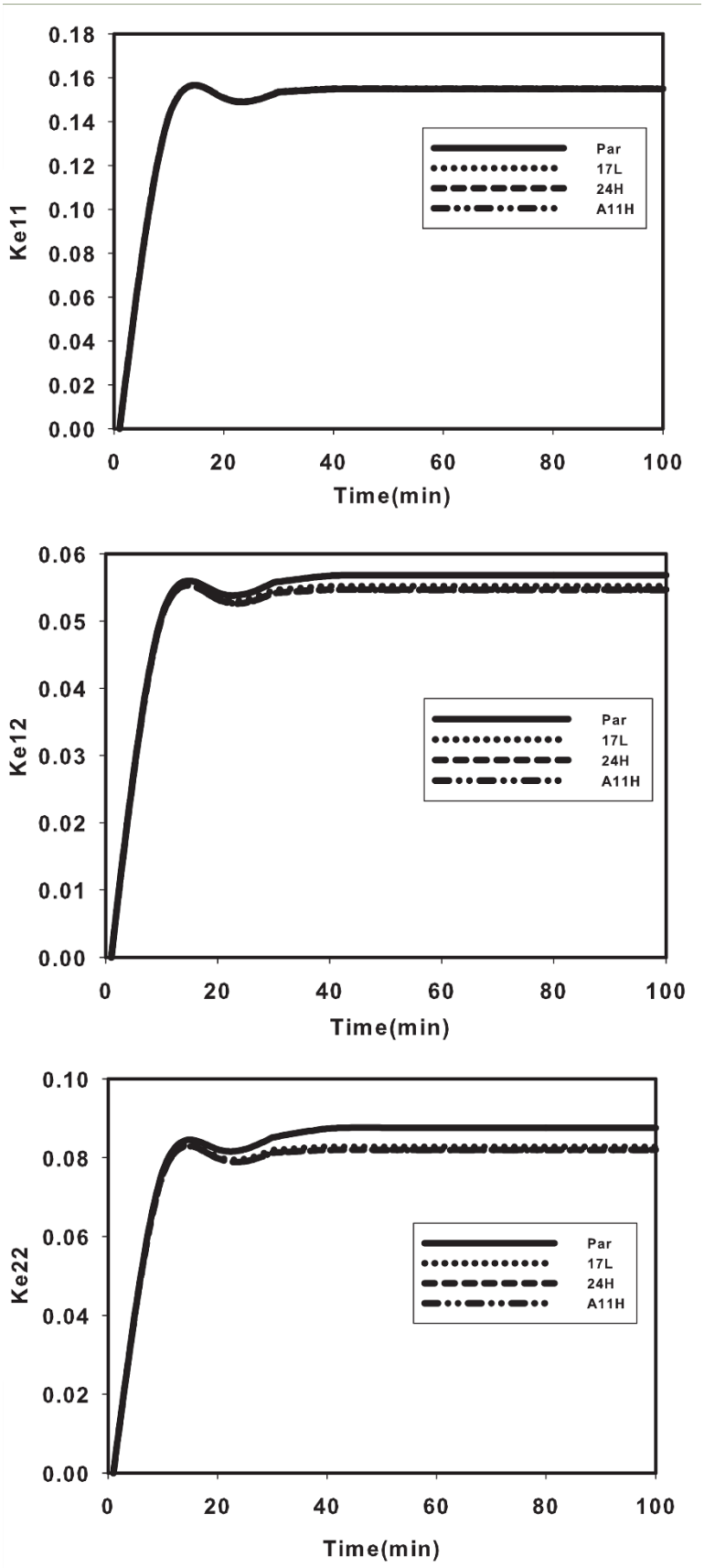


Figure 4.15 Plot showing the values of internalization rate constants (Ke) calculated using ABM at intervals of 10 (min) as the simulation progresses. The rate constant values, increase and become steady after 30-40 minutes.

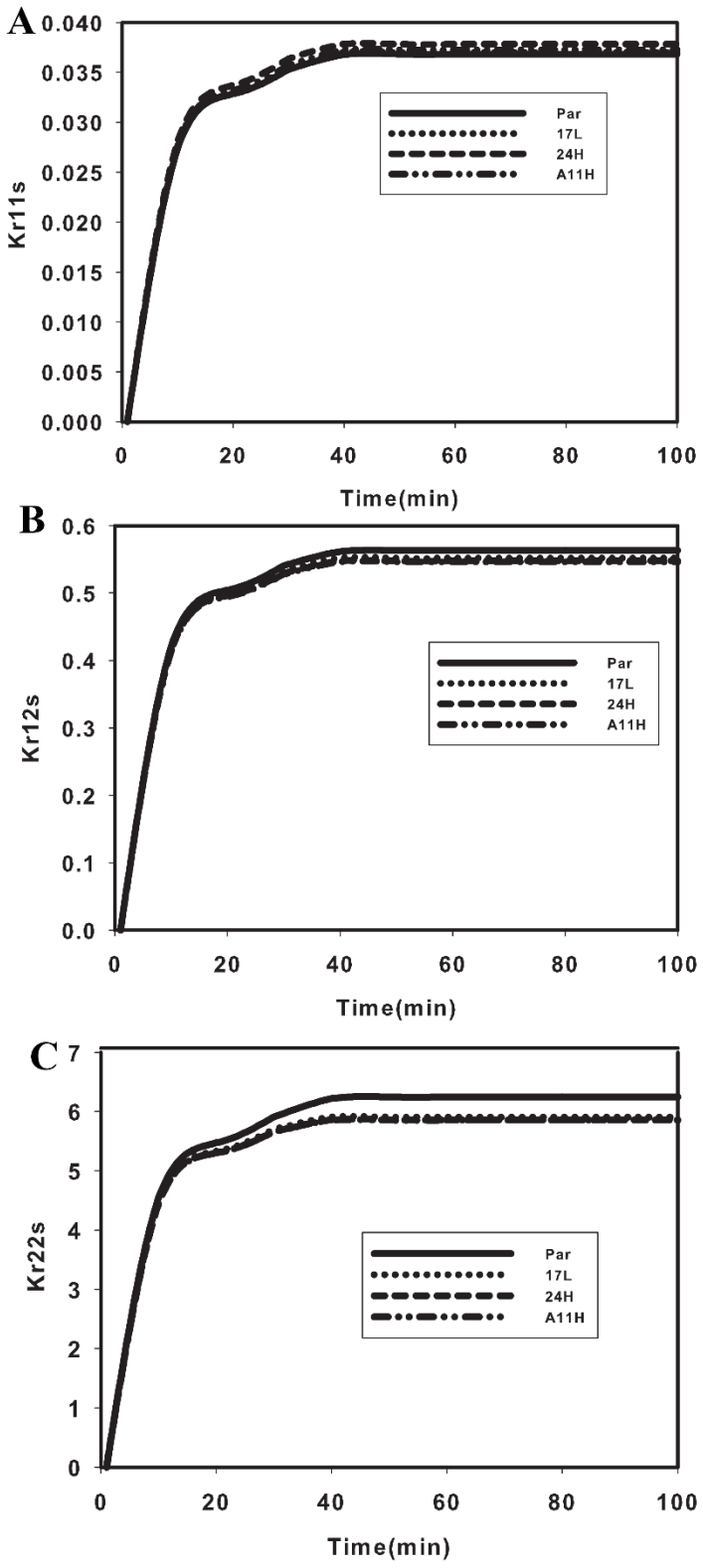


Figure 4.16 Plot showing the values of dissociation rate constants on PM (Krs) calculated using ABM at intervals of 10 (min) as the simulation progresses. The rate constant values, increase and become steady after 30-40 minutes.

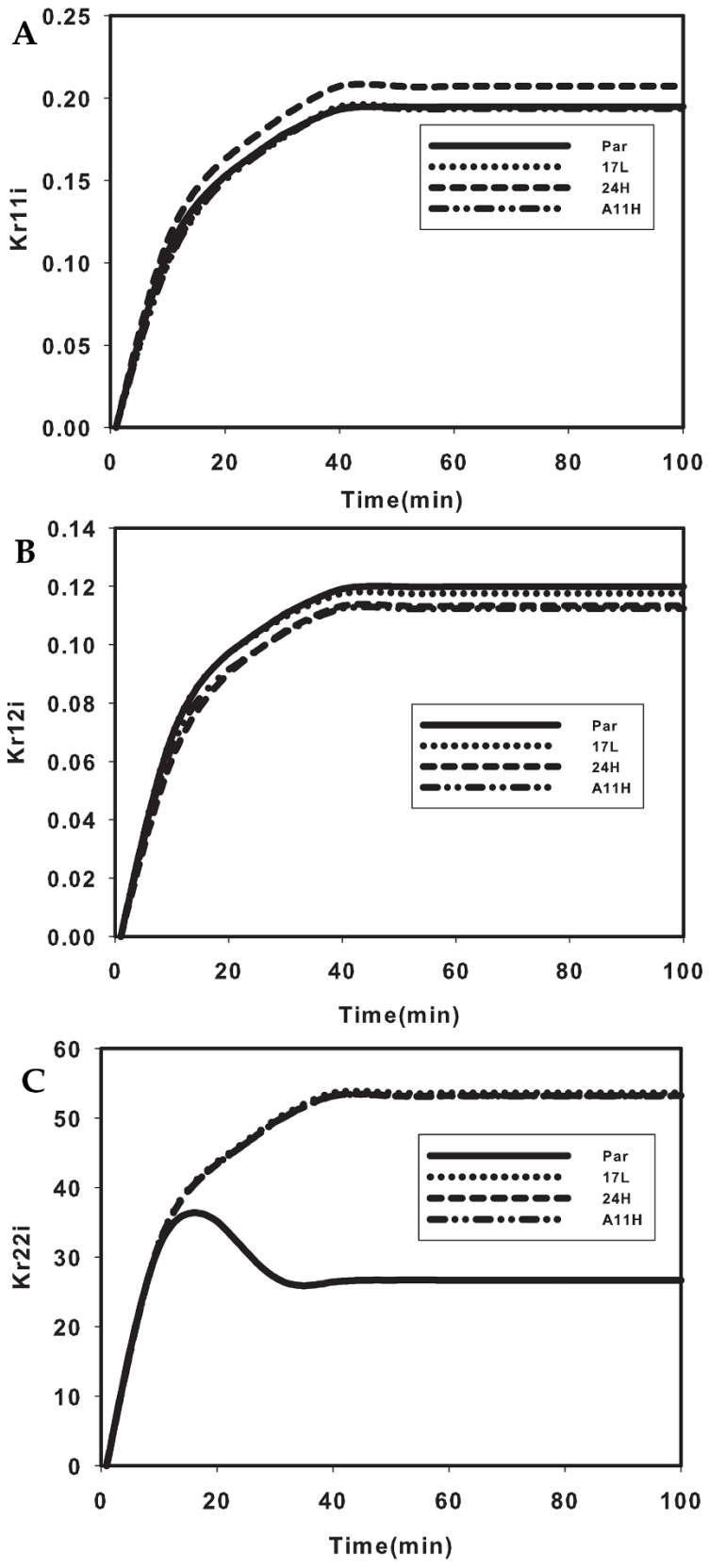


Figure 4.17 Plot showing the values of dissociation rate constants on EE (Kri) calculated using ABM at intervals of 10 (min) as the simulation progresses.

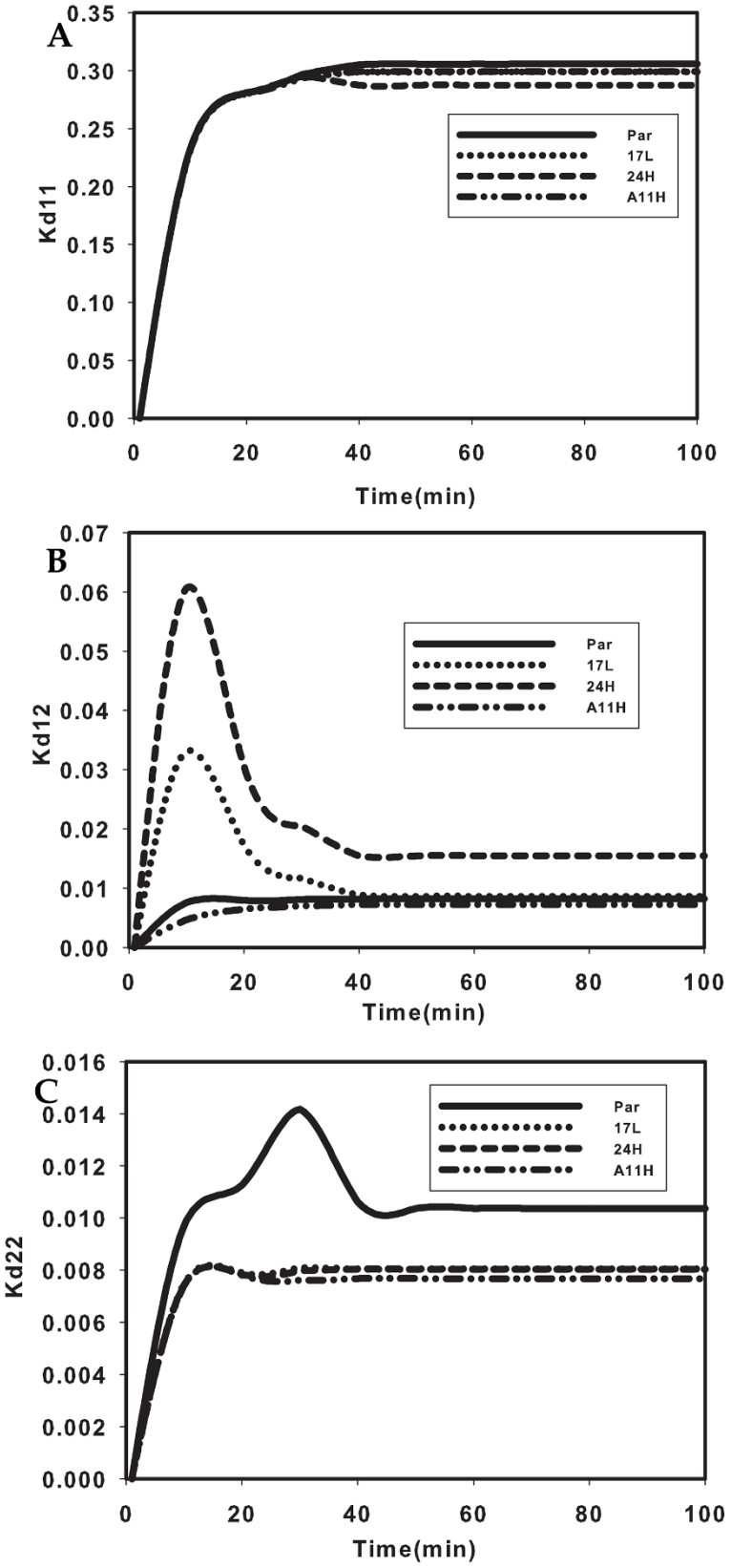


Figure 4.18 Plot showing the values of degradation rate (Kd) calculated using ABM at intervals of 10 (min) as the simulation progresses.

Table 4.3 shows values of rate constants for each cell line, averaged over the 20 simulations. Parameters generated by the ABM are compared with best fit parameters and confidence intervals from the mathematical model.

Table 4.3 Average values for rate constants obtained from ABM model for Par, 17L, 24H and A11H cell lines.

Parameter(/min)	Average value from simulation				Parameter(/min)	Average value from simulation			
	Parental	17L	24H	A11H		Parental	17L	24H	A11H
kfs	0.027	0.021	0.014	0.017	**kd1**	0.019	0.019	0.019	0.018
kr11s	0.032	0.032	0.033	0.032	**kd2**	0.008	0.008	0.008	0.008
kr12s	0.491	0.495	0.479	0.478	**ke11**	0.139	0.139	0.139	0.139
kr22s	5.425	5.312	5.126	5.123	**ke12**	0.05	0.05	0.049	0.049
kr11i	0.164	0.162	0.173	0.161	**ke22**	0.078	0.076	0.073	0.073
kr12i	0.102	0.098	0.096	0.095	**kd11**	0.268	0.263	0.255	0.263
kr22i	24.72	47.07	45.17	45.152	**kd12**	0.007	0.013	0.022	0.006
kt1	0.018	0.018	0.018	0.018	**kd22**	0.012	0.007	0.007	0.007
kt2	0.007	0.007	0.007	0.007					

The cell line dependent rates are in the range predicted by the mathematical model. Estimated average values for parameters using ABM, compared with best fit parameters and confidence intervals from the mathematical model.

The simulation space has been populated with as many agents as the number of receptors, ligands and proteins expressed in the cell lines. However, as the number of agents increases, the computational limit and processing time become a bottleneck for realistic simulations. The model will then have to be run on a reduced system (75%, 50% abundance) and a warning message given, when the quantities of species fall below the minimum threshold value needed for satisfying the rates of reactions.

4.5 Summary

As proof-of-concept, two complementary modeling formalisms have been pipelined to study the dynamics of ErbB receptor activation and trafficking. The first formalism is a mathematical model which solves rate equations to track molecular abundances over time. The aggregate characteristics of equation-based modeling, drives the individual-based agent model. The rules were designed so as to account for all the significant events leading to the state changes of the molecular species. Reverse engineering with computational models generates metadata that is currently used

only to make first-level inferences. The macro and micro rules are information structures that plough back the knowledge gained from research, to conduct *in silico* experiments for validation, predictions and testing hypotheses. The model predicts phosphorylated dimer levels for any given receptor expression profile and the emerging characteristics were validated with the original reverse-engineered results of the experiment. ABM also offers a novel method for calculating rate constants from the changing species concentration.

5

Application: Prediction of ErbB signal activation using the data analytics pipeline

In the previous chapter, we proposed an architecture for a modeling pipeline where the agent-based approach was used for re-engineering the system by learning from data generated by reverse engineering. This work is an application of the data-driven modeling pipeline. The model has been extended to study the activation kinetics in HME cells co-expressing ErbB1-3 receptors and their trafficking in the presence of ligands EGF and HRG. Activation patterns of Erk and Akt, triggered by phosphorylated dimers, is investigated by adding Erk and Akt as agents in the system. The effect of an ErbB2 blocking drug (Pertuzumab) on HER2 expressing cell lines was also carried out.

5.1 Model Description

5.1.1 Reverse engineering of ErbB signaling

The integrated mathematical model of HER activation, and trafficking parameterized with a comprehensive set of HER phosphorylation and abundance data collected in a panel of HME cells expressing varying levels of HER1-3 by Shankaran, et al. (2013) is taken as the reverse engineered system for this study, henceforth referred to as *123Model*. In their experiment, the HME parental cell line (low level of HER2 and HER3; designated as HER2-HER3-) was transduced with HER2 gene and HER3 gene to clone 24H (HER2+HER3-) and B5 (HER2-HER3+) cell lines, respectively. Further, 24H cell line was inserted with the HER3 gene to obtain the D20 (HER2+HER3+) cell line that expressed all the three receptors. These cells were activated by the addition of known concentration of EGF and HRG to the culture medium and incubated for 5 to 120 min. ELISA assays were done to quantify the receptor mass and phosphorylation levels.

The *123Model* takes into account, different types of ligands, monomers, dimers and signaling molecules that exist in three distinct compartments – the cell surface or plasma membrane (PM), early endosomes (EE) and late endosomes (LE). The model estimates values of 47 unknown parameters for active receptor dimer formation affinities, trafficking rates and relative phosphorylation levels using nonlinear least square regression to simultaneously fit the experimental data collected from the four cell lines. Overall root-mean-squared error (RMSE) between the experimental data and model predictions were used to assess the goodness of the fit.

The output data from the reverse engineered model forms the input for the agent-based re-engineering model.

5.1.2 Agent-based Re-engineering of ErbB signaling

5.1.2.1 Initialization

The key molecular entities in the system are represented as agents. Agents are of four types: ligands (EGF/HRG), receptors (ErbB1/ErbB2/ErbB3) (Figure 5.1) and signaling proteins (Erk and Akt). The 3D spatial domain is divided into six compartments: (1) an extracellular domain (Outside), (2) plasma membrane (PM), (3) early endosome (EE), (4) late endosome (LE) (5) lysosome and (6) cytoplasm.

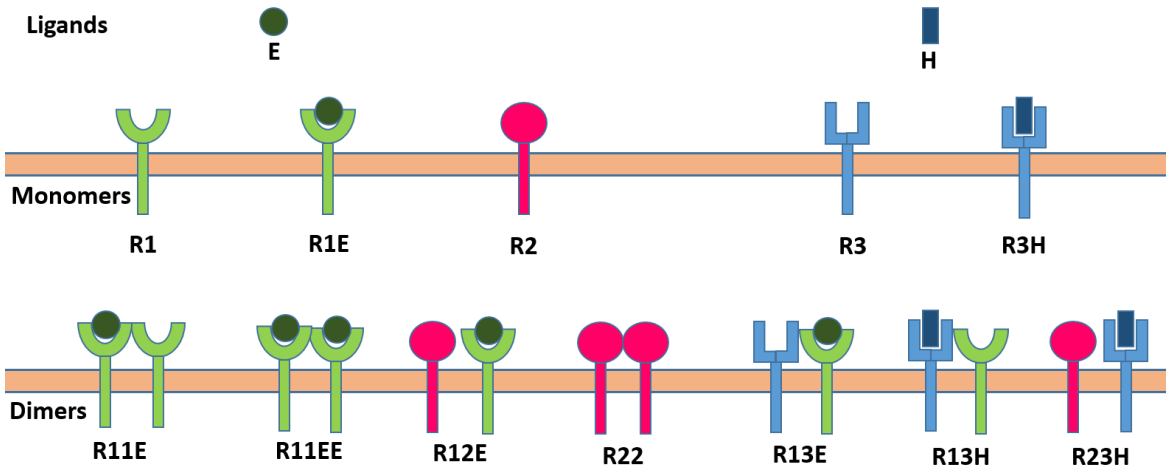


Figure 5.1 Ligand and receptor entities: E- EGF, H – HRG, R1-ErbB1 monomer, R1E-EGF bound ErbB1 monomer, R2-ErbB2 monomer, R3-ErbB3 monomer, R3H-HRG bound ErbB3 monomer, R11E-Single EGF bound ErbB1 homodimer, R11EE-Double EGF bound ErbB1 homodimer, R12E-Single EGF bound ErbB1–ErbB2 heterodimer, R22- ErbB2 homodimer, R13E-Single EGF bound ErbB1–ErbB3 heterodimer, R13H-Single HRG bound ErbB1–ErbB3 heterodimer, R23H-Single HRG bound ErbB2–ErbB3 heterodimer.

Initially, ligands, EGF and HRG, are present in the outside compartment, which on binding to receptors move to the plasma membrane (PM) (Figure 5.2). Bound/unbound monomers and dimers are internalised to the early endosome (EE). From EE, they are destined to accumulate in the late endosome (LE) for degradation or are recycled to return to the PM. Once degraded they move to the lysosome. Erk and Akt are sentinel proteins present in the compartment designated as cytoplasm. Synthesis of agents occurs at regular intervals from the external compartment which supplies new ligand and receptor agents to the Outside and PM respectively. The number and type of agents are initialized to their abundance in the cell lines (Table 5.1). Ligand agent and receptor agents are defined same as in Chapter 4. The sentinel protein agents-Erk and Akt are defined by a set of attributes (Figure 5.2).

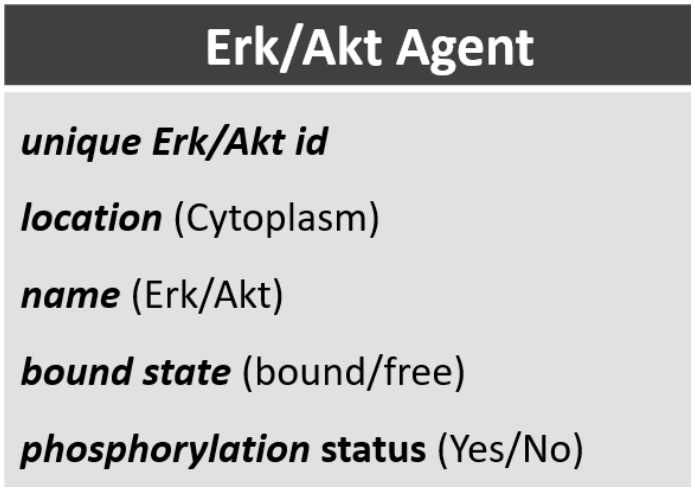


Figure 5.2 Attributes of Erk/Akt agent.

Table 5.1 HER expression levels in cell lines (Courtesy (Shankaran Harish et al. 2013).

List of cells lines	Abundance of EGFR, HER2 and HER3 (molecules/cell)		
Notation	**EGFR**	**HER2**	**HER3**
Par (HER2-3-)	200000	30000	973
24H (HER2+3-)	136938	600000	2000
B5 (HER2-3+)	189126	30000	28000
D20 (HER2+3+)	85896	643777	28770

The cellular events that lead to signal activation are ligand binding, dimerization, phosphorylation, internalization, recycling, dimer dissociation and degradation (Figure 5.3).

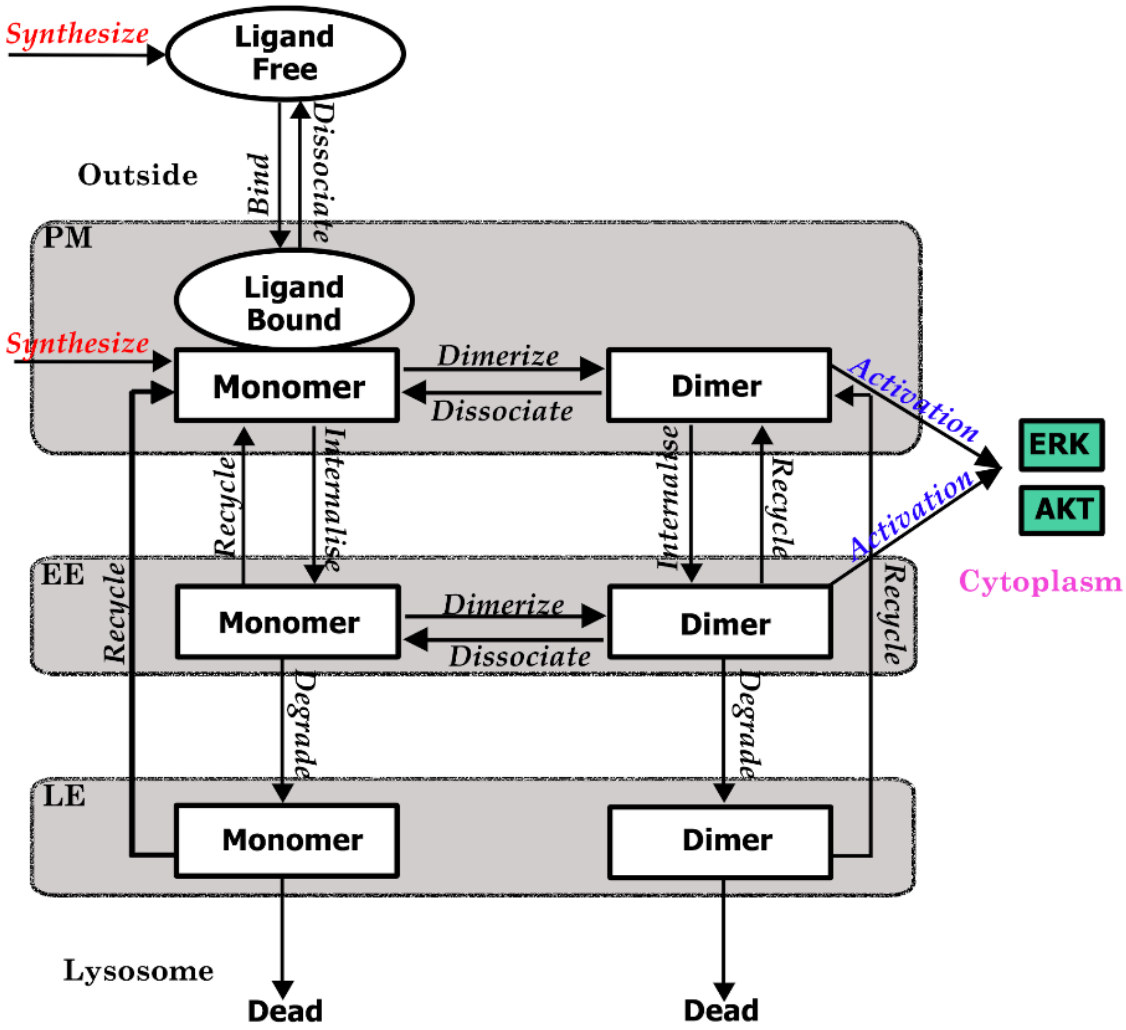


Figure 5.3 Schematic of signalling events in the cellular compartments.

5.1.2.2 Action rules

The agent interactions leading to signaling events are determined by macro and micro rules (Table 5.2). The rules are defined based on the dependency map as shown in Figure 5.4. The additional rate constants in the *123Model* due to the presence of ErbB3 and its ligand HRG, are marked in red. The dependencies are intuitively defined and are converted into rules using the values of reaction rates.

Table 5.2 Macro and Micro rules for EGFR, HER2 and HER3 model

Macro Rules	Micro Rules
1. *EGFR monomer internalization rate is higher than internalization rate of HER2 and HER3 monomer.* 2. *EGFR homodimers undergo dissociation more than internalization.* 3. *EGFR-HER2 and HER2 homodimer dissociates faster on PM than EE.* 4. *Erk correlated with HER1 signaling from HER1 homodimer and HER1-HER2 heterodimer* 5. *AKT activation correlates more with HER3 phosphorylation and signaling from HER3-HER1 and HER2-HER3 dimers.* 6. *EGFR- HER2 heterodimer internalises faster than HER2 monomer* 7. *Internalization rate of EGFR homodimer is greater than its degradation rate* 8. *HER2 heterodimers dissociation rate is greater than its degradation rate* 9. *HER2 homodimer degrades at a slower rate compared to its dissociation rate* 10. *EGFR-HER2 dissociates faster compared to its internalization.* 11. *HER2 homodimer dissociates faster compared to its internalization.*	1. *EGFR-HER2 internalization rate is 10 times more than HER2 monomer internalization rate* 2. *EGFR-HER2 dissociation rate on PM is more than 100 times its internalization rate* 3. *EGFR-HER3 heterodimer internalization rate is 2 times more than HER3 monomer internalization rate* 4. *EGFR homodimers on EE dissociates 4 times slower compared to them on PM* 5. *The probability of dimerization for single EGF bound EGFR homodimer, double EGF bound EGFR homodimer, EGFR-HER2 heterodimer and EGFR-HER3 heterodimer are 0.053, 0.22, 0.46 and 0.0074 respectively for 24H cell line.* 6. *The probability of dimerization for EGFR-HER2 heterodimer, HER2 homodimer and HER3-HER2 heterodimer are 0.19, 0.02 and 0.15 respectively for D20 cell line.* 7. *The probability of dimerization for EGF bound EGFR-HER3 heterodimer, HRG bound EGFR-HER3 heterodimer and HER3-HER2 heterodimer are 0.0018, 0.0047 and 0.07 respectively for B5 cell line.*

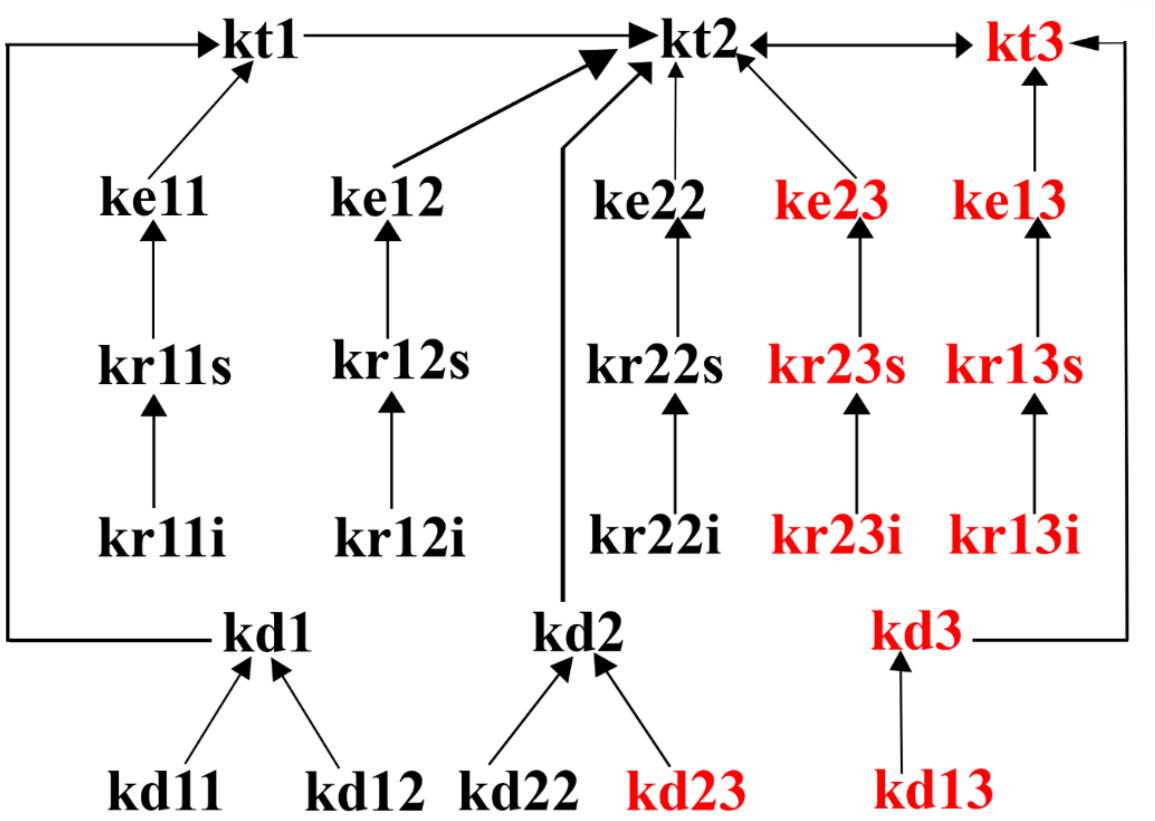


Figure 5.4 Dependency map of rate constants: kr – dissociation rate, kd – degradation rate, kt – internalization rate for monomers, ke-internalization rate for dimers. 's' – surface (PM), 'i' –internal compartment (EE), 1 – EGFR, 2 –HER2, 3-HER3 .e.g. kr23s is the heterodimer HER2-HER3 dissociation rate on PM

The constructed receptor activation model has been used to quantitatively predict the relative contributions of the HER receptor types and their dimers to the activations of Erk and Akt kinases. Erk and Akt undergoes dimer-specific activation (macro rules 4 and 5 derived from (Gong, et al., 2015)) during the time course of the simulation.

5.2 Results

In silico experiments replicate the signaling behaviour in cell lines. The activation of ErbB1-3 receptors, on binding with ligands EGF and HRG, their trafficking and stimulation of Akt and Erk pathways have been simulated. Given the abundance of EGFR, HER2 and HER3 in the 4 HME cell lines (Table 5.1), twenty simulation runs were made for each cell line, and the parameter values were averaged. The simulations were run for 120 min with each time step equal to 1 min. The major assumptions made for the experiments are:

(1) EGF binding rate and dissociation rates were 0.63/min and 0.12/min, respectively.

(2) HRG binding rate and dissociation rates were 0.73/min and 0.34/min, respectively.

(3) Dimerization probabilities are calculated for each cell line from dimer contributions to HER phosphorylation predicted by (Shankaran, et al., 2013).

(4) Ligand quantities were calculated by increasing the numbers till they reach saturating levels. The numbers were fixed at - EGF =200000, HRG = 30000.

(5) At the start of simulation, 2% of EGFR, 1% of HER2 and 0.05% of HER3 are present in the early endosome (EE).

(6) Since HER3 lacks kinase activity, the phosphorylation effect of HER3 homodimers is not considered.

(7) Internalization and degradation rates for monomers (kt1, kt2, kt3, kd1, kd2 and kd3) are initialized to the values obtained from the mathematical model assuming that these events occur even before ligand addition. All other parameters are estimated by the ABM.

(8) Agents are recycled during dissociation in EE, internalization and degradation. The recycling fractions were taken from literature (Shankaran, et al., 2013).

(9) Newly synthesized receptors and ligands are added to the cell surface and outside respectively.

(10) The number of Erk and Akt agents are fixed at 20000 in all cell lines.

(11) Drug is introduced, 20 min after the beginning of the simulation. Its action is immediate and continues till the end.

5.2.1 Prediction of receptor activation and signaling

The receptor phosphorylation levels were quantified by multiplying receptor/dimer abundances with constant phosphorylation efficiency factors derived from the measurements in the ELISA assays. The timecourse measurements of levels of phosphorylated HER1-3 for the four cell lines, Par, 24H, B5 and D20 in response to addition of ligands EGF and HRG are shown in Figure 5.5. The phosphorylation values were calculated by considering the contribution from activated dimers in the

external (PM) as well as internal compartments (EE). The peak values for phosphorylation are attained at 5-10 min after ligand addition.

The EGFR phosphorylation is found to be high in all four cell lines with highest level in 24H (HER2+3-) cells followed by D20, B5 and Par (Figure 5.5A). HER2 activation levels are much higher in the HER2+ cell lines (24H and D20) compared to the HER2- cell lines (Figure 5.5B). HER3 phosphorylation peaks slowly and is maximum for the HER2+3+ cell line (D20) (Figure 5.5C).

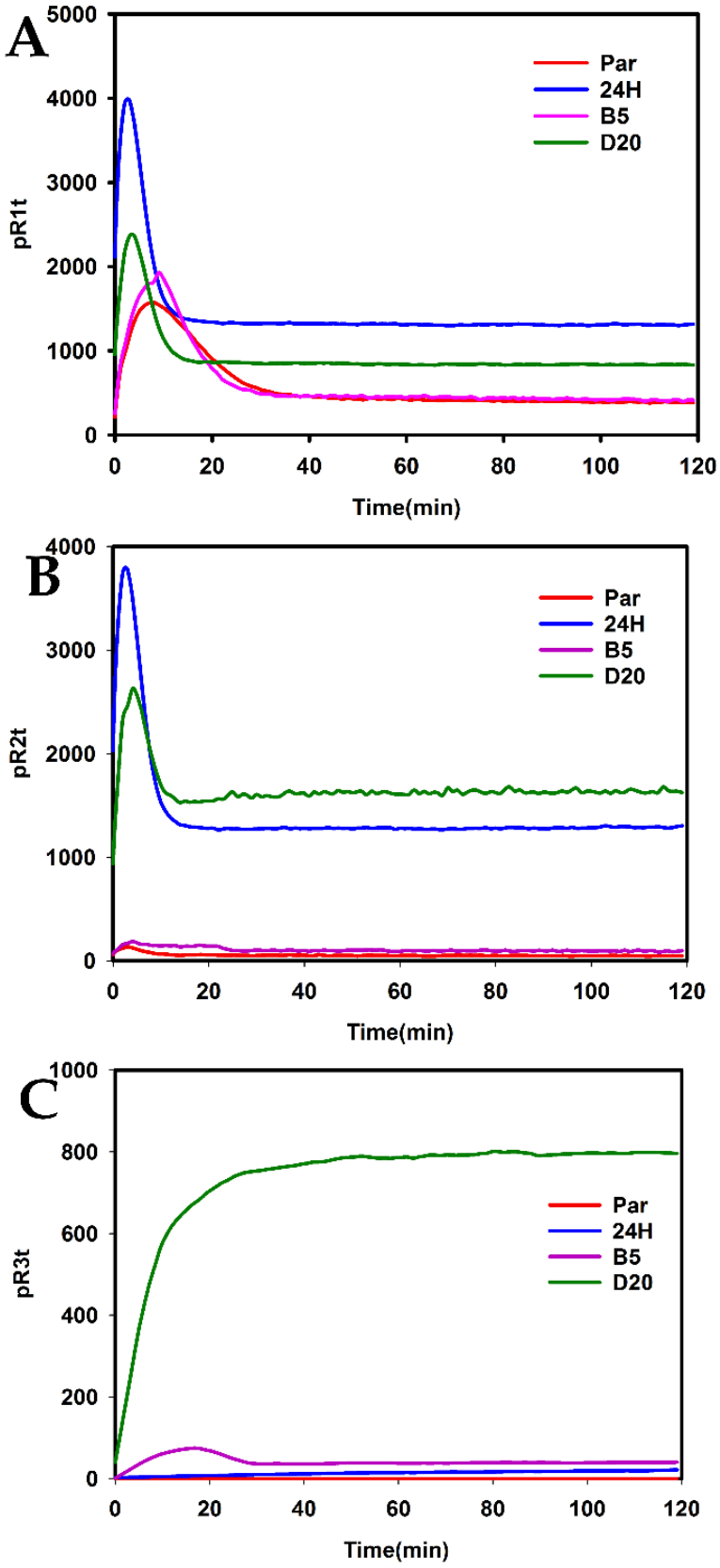


Figure 5.5 Phosphorylation levels for (A) EGFR (B) HER2 and (C) HER3 receptors in Par (HER2-/3-), 24H(HER2+/3-), B5(HER2-/3+) and D20 (HER2+/3+) cell lines in response to activation by ligands EGF and HRG.

The HER expression levels determine the dimerization pattern which in turn govern the activation levels of signaling molecules Erk and Akt (Figure 5.6 and Figure 5.7). The model predictions for dimer contributions were considered at 100 min following ligand addition, after which the phosphorylation levels were found to be relatively steady. An Erk agent is activated if it finds dimers EGFR-EGFR or EGFR-HER2. The Erk agents are activated most in 24H and D20 cells and this is accredited to the formation of heterodimers between EGFR and HER2. In the HER2- cells of B5 and Par, the activation is due to the homodimers of EGFR (Figure 5.6). Figure 5.5.1 shows the error plots for total receptor phosphorylation.

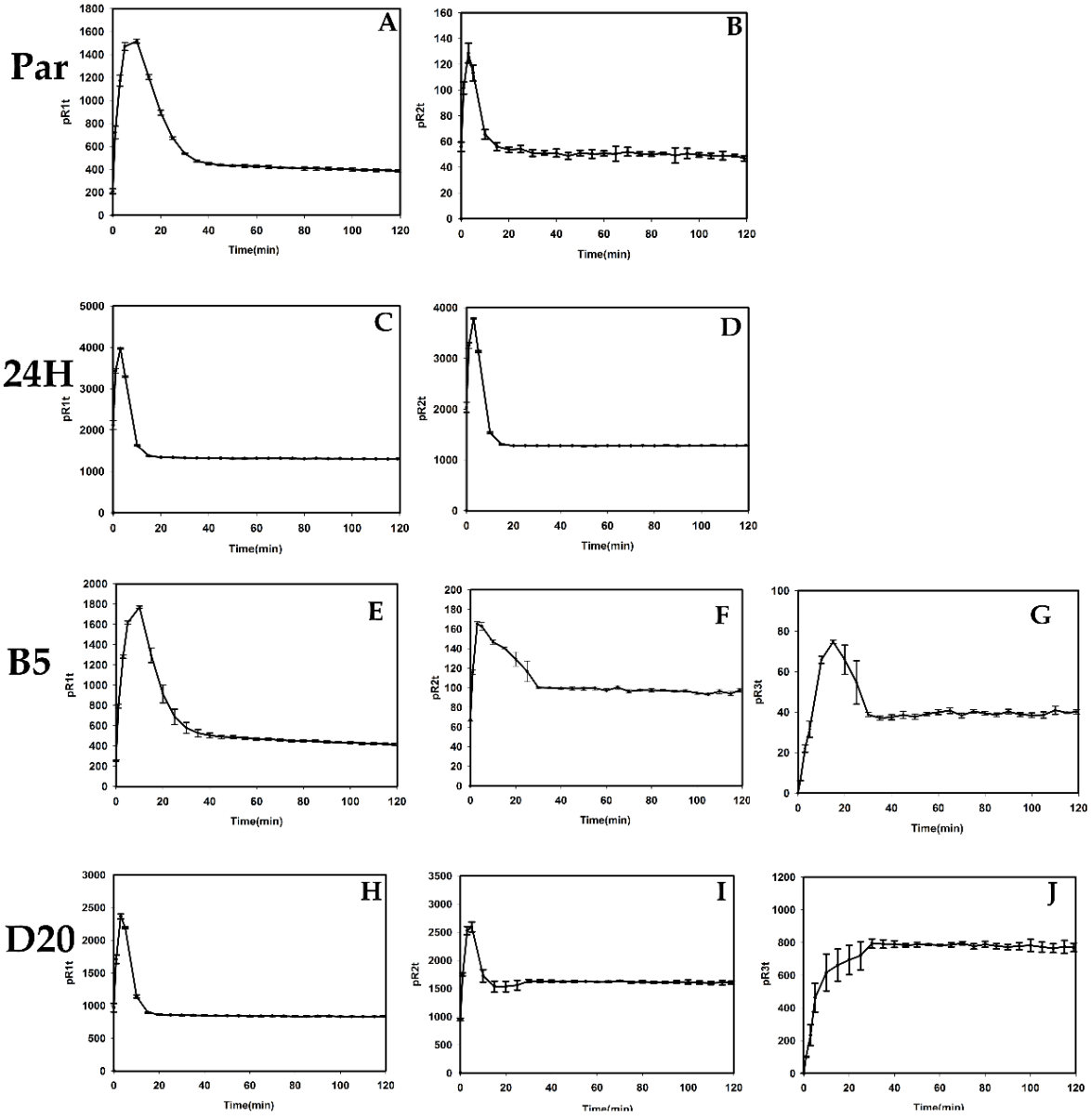


Figure 5.5.1. Error plots for total receptor phosphorylation in Par, 24H, B5 and D20 cell lines. Error bars indicate the standard error over 20 simulations.

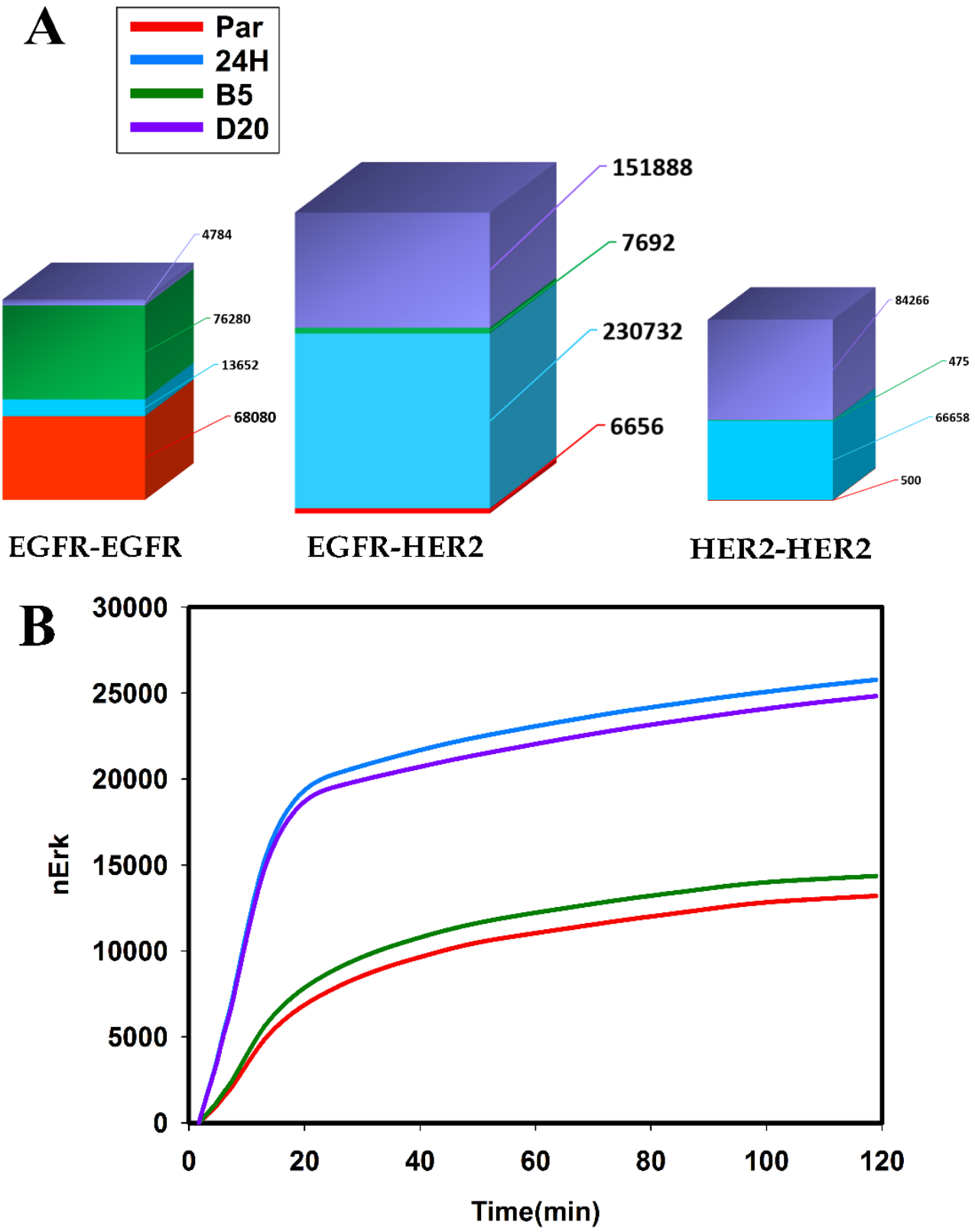


Figure 5.6 Model predictions on dimer contribution to Erk activation dynamics. (A) Dimer contribution from EGFR-EGFR, HER2-HER2and EGFR-HER2 as stacked bars for Par(HER2-/3-), 24H(HER2+/3-), B5(HER2-/3+) and D20 (HER2+/3+) cell lines at time=60 min. (B) Erk activation in the cell lines.

An Akt agent is activated, if it finds dimers of EGFR-HER3 or HER2-HER3. The Akt agents are activated most in D20 cells and this is due to the formation of large numbers of HER2-HER3 heterodimers. In the B5 (HER2-3+) cells, the EGFR-HER3 contribution is maximum relative to other cell lines. Akt activation is less for 24H and Par due to the absence of HER3 (Figure 5.7). Error bars for Erk and Akt activation are plotted in Figure 5.6.1.

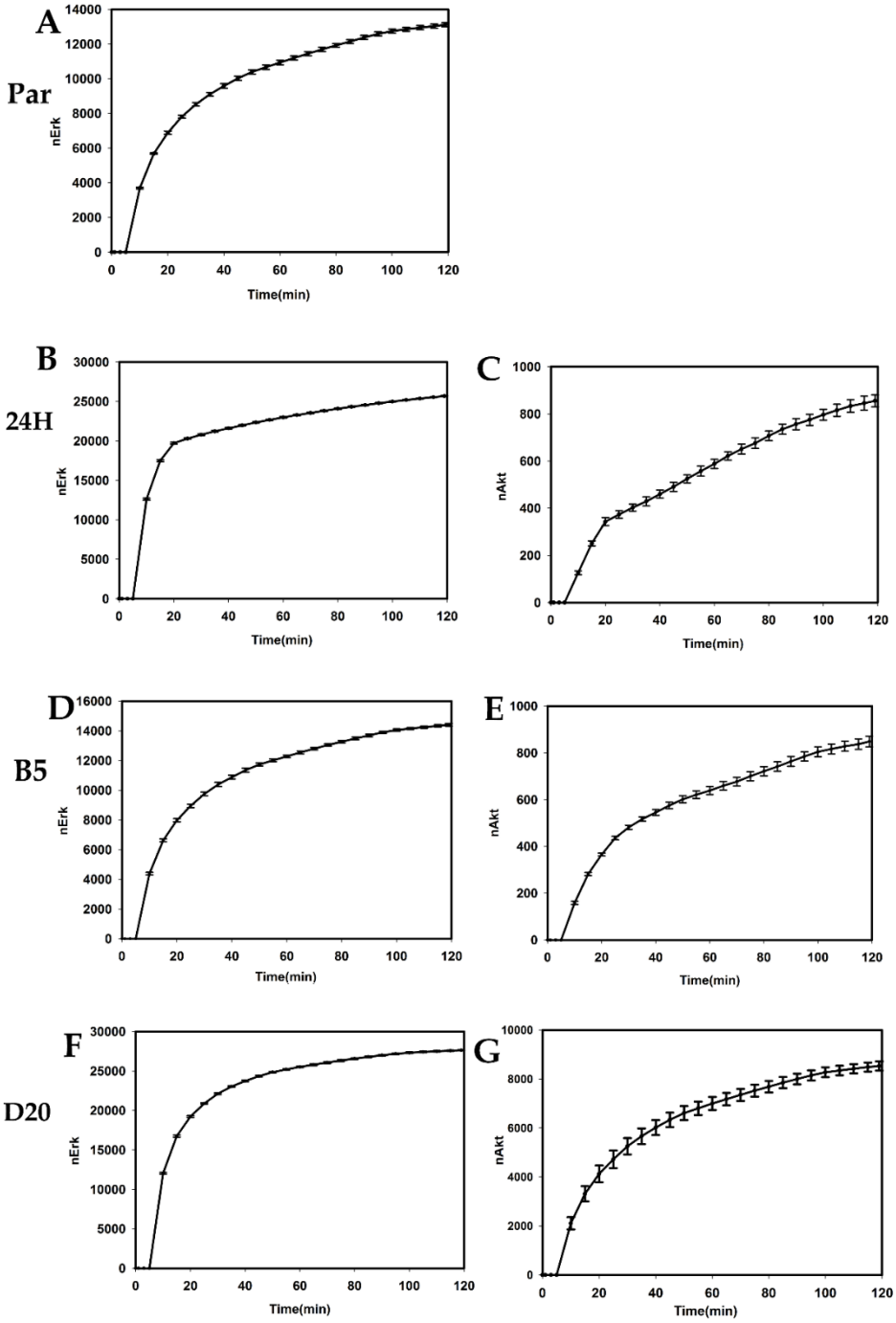


Figure 5.6.1 Error plots for Erk and Akt activation in Par, 24H, B5 and D20 cell lines. Error bars indicate the standard error over 20 simulations.

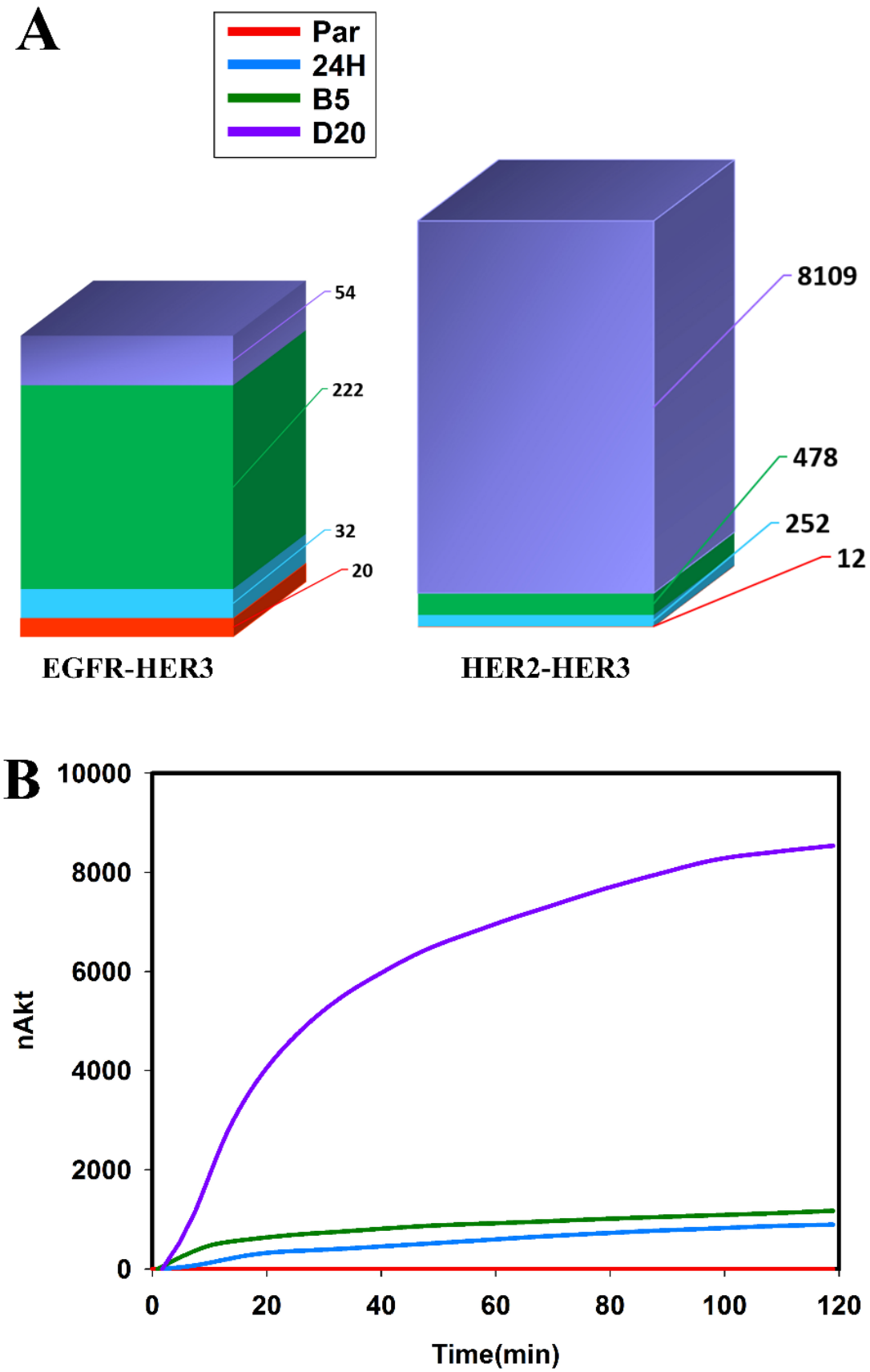


Figure 5.7 Model predictions on dimer contribution to Akt activation dynamics. (A) Dimer contribution from EGFR-HER3 andHER2-HER3 as stacked bars for Par(HER2-/3-), 24H(HER2+/3-), B5(HER2-/3+) and D20 (HER2+/3+) cell lines at time=60 min. (B) Akt activation in the cell lines.

5.2.2 Parameter Estimation

The rate constants are calculated for activation and trafficking events from the changing species concentration. At the end of each time interval (1 min), the quantities of species were re-calculated and values of the rate parameters were obtained from actual concentrations of species undergoing dimerization, dissociation, internalization, recycling and degradation.

Table 5.3 Average value for rate parameters in Par, 24H, B5 and D20 cell lines

Parameters	Average value from simulations				Parameters	Average value from simulations			
	Par	24H	B5	D20		Par	24H	B5	D20
kr11s	2.04	1.06	2.71	0.87	**ke11**	0.27	0.27	0.48	0.34
kr22s	-	10.90	-	11.11	**ke12**	0.27	0.32	0.27	0.32
kr23s	-	8.20	20.52	39.97	**ke22**	-	0.07	-	0.08
kr12s	5.31	5.84	5.37	5.83	**ke13**	-	0.31	0.07	0.27
kr13s	-	0.42	10.28	0.57	**ke23**	-	0.21	1.04	0.76
kr11i	8.66	2.43	8.26	1.79	**kd1**	0.08	0.08	0.07	0.07
kr22i	-	0.51	-	0.51	**kd2**	0.03	0.007	0.01	0.008
kr12i	0.30	0.33	0.30	0.33	**kd3**	0.07	0.08	0.07	0.08
kr23i	-	0.92	2.49	11.72	**kd11**	0.03	0.02	0.03	0.02
kr13i	-	0.46	6.30	0.92	**kd12**	0.03	0.03	0.03	0.03
kt1	0.07	0.07	0.07	0.07	**kd22**	-	0.007	-	0.008
kt2	0.01	0.01	0.01	0.02	**kd13**	-	0.05	0.04	0.05
kt3	0.01	0.05	0.01	0.01	**kd23**	-	0.08	0.07	0.07

Table 5.3 shows values of rate constants for each cell line, averaged over the 20 simulations. Rate constants for some of the species that do not form at all, are left blank. In Par (HER2-3-) cells, homodimers of HER2 and heterodimers of HER3, do not form. Similarly, in B5, homodimers of HER2 are insignificant.

5.2.3 Receptor blocking experiments

Studies on receptor blocking were done by determining the receptor phosphorylation in various cell lines in the presence of the drug 2C4 (Pertuzumab) (Figure 5.8-10). Pertuzumab is a HER2 dimerization inhibitor, a monoclonal antibody that binds to domain 2 of HER2 blocking its dimerization. Model simulations assume that the addition of drug renders 80% of the cellular HER2 unavailable for receptor dimerization.

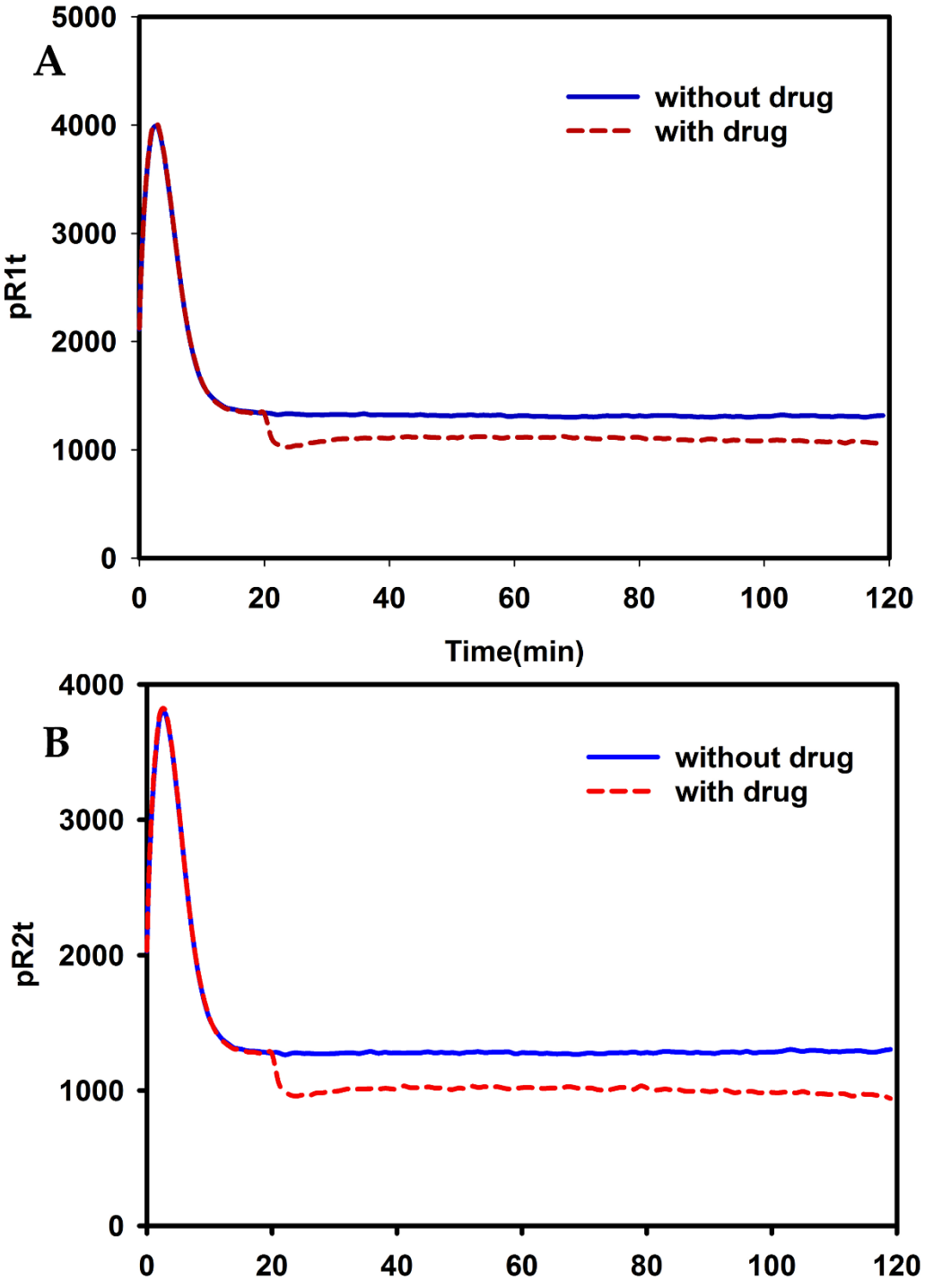


Figure 5.8 HER receptor phosphorylation pattern of 24H on inhibition of HER2 by Pertuzumab.
(A) HER1 and (B) HER2 phosphorylation in the absence and presence of Pertuzumab.

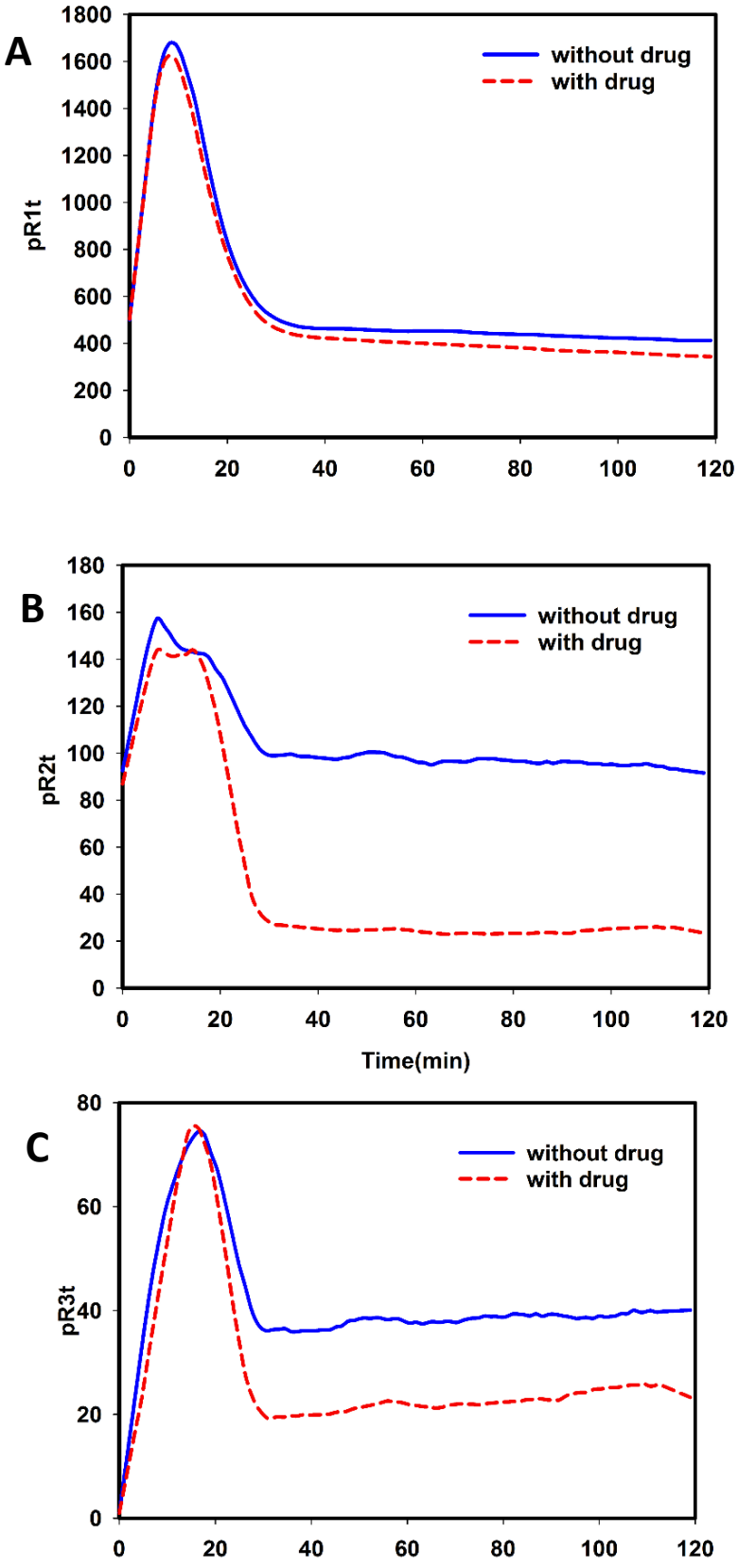


Figure 5.9 HER receptor phosphorylation pattern in B5 on inhibition of HER2 by Pertuzumab.
(A) HER1, (B) HER2 and (C) HER3 phosphorylation in the absence and presence of Pertuzumab.

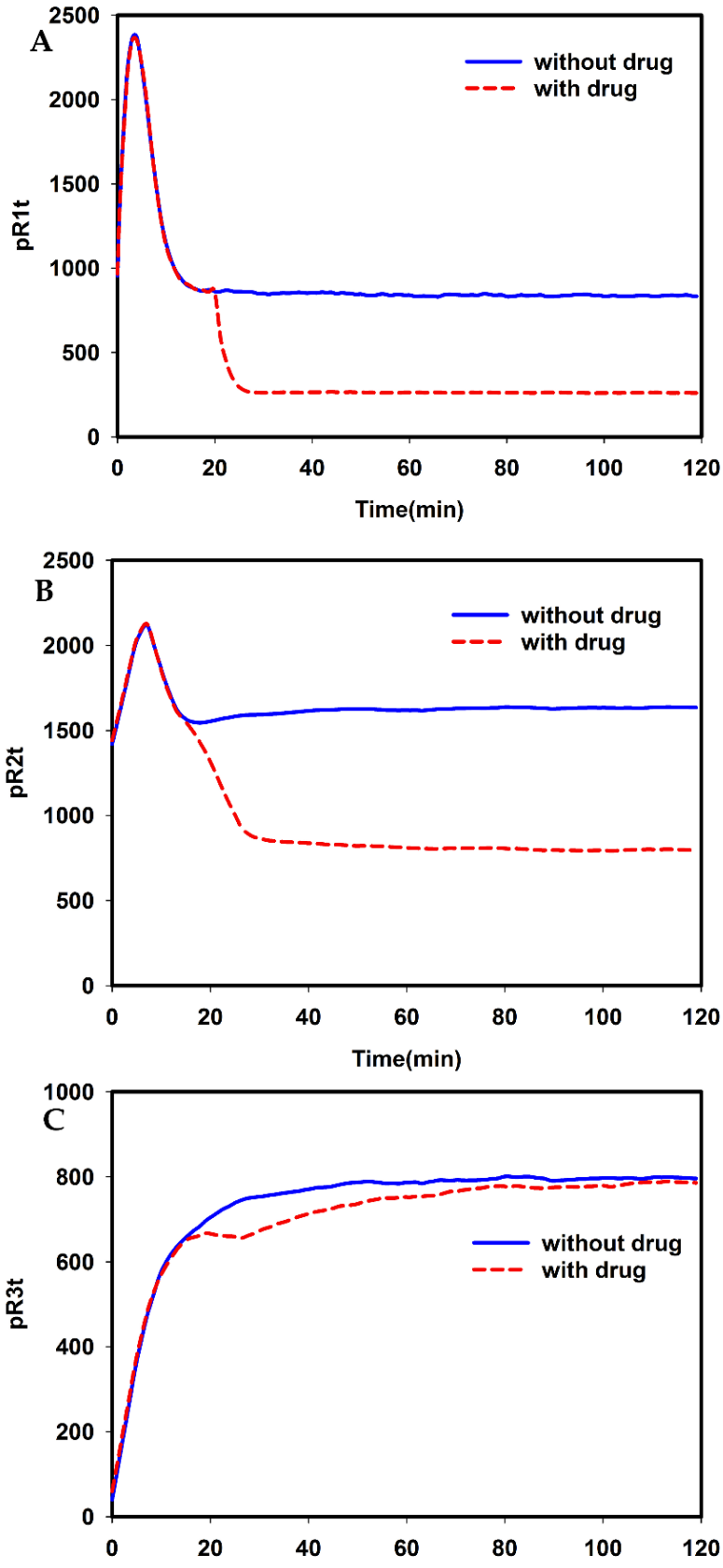


Figure 5.10 HER receptor phosphorylation pattern in D20 on inhibition of HER2 by Pertuzumab. (A) HER1, (B) HER2 and (C) HER3 phosphorylation in the absence and presence of Pertuzumab.

In 24H, EGFR and HER2, phosphorylation falls by about 15% due to less formation of EGFR-HER2 and HER2-HER2 dimers and HER3 expression being very less, there is no impact of the drug on HER3 phosphorylation (Figure 5.8 A-B). In B5, there is a steep fall in HER2 and HER3 phosphorylation levels on drug addition (Figure 5.9 A-C). This is due to HER2-HER3 hetero dimers not being formed. In the D20 cells, HER1 and HER2 activation show a sharp decline because blocking of HER2 affects formation of HER1-HER2 dimers whereas HER3 activation is held by more of HER1-HER3 dimers forming (Figure 5.10 A-C).

When the drug blocks the formation of HER1-HER2, Erk activation shows sharp decline in 24H (Figure 5.11A). In D20, Erk activation dips only by a small amount because of more EGFR homodimer formation due to the absence of HER2 (Figure 5.11B). Akt activation, being linked to the HER3 dimers, shows a gradual rise in the HER3+ cells of B5 and D20 peaking at lower values than the level achieved by the original system before treatment with drug (Figure 5.11 C-D).

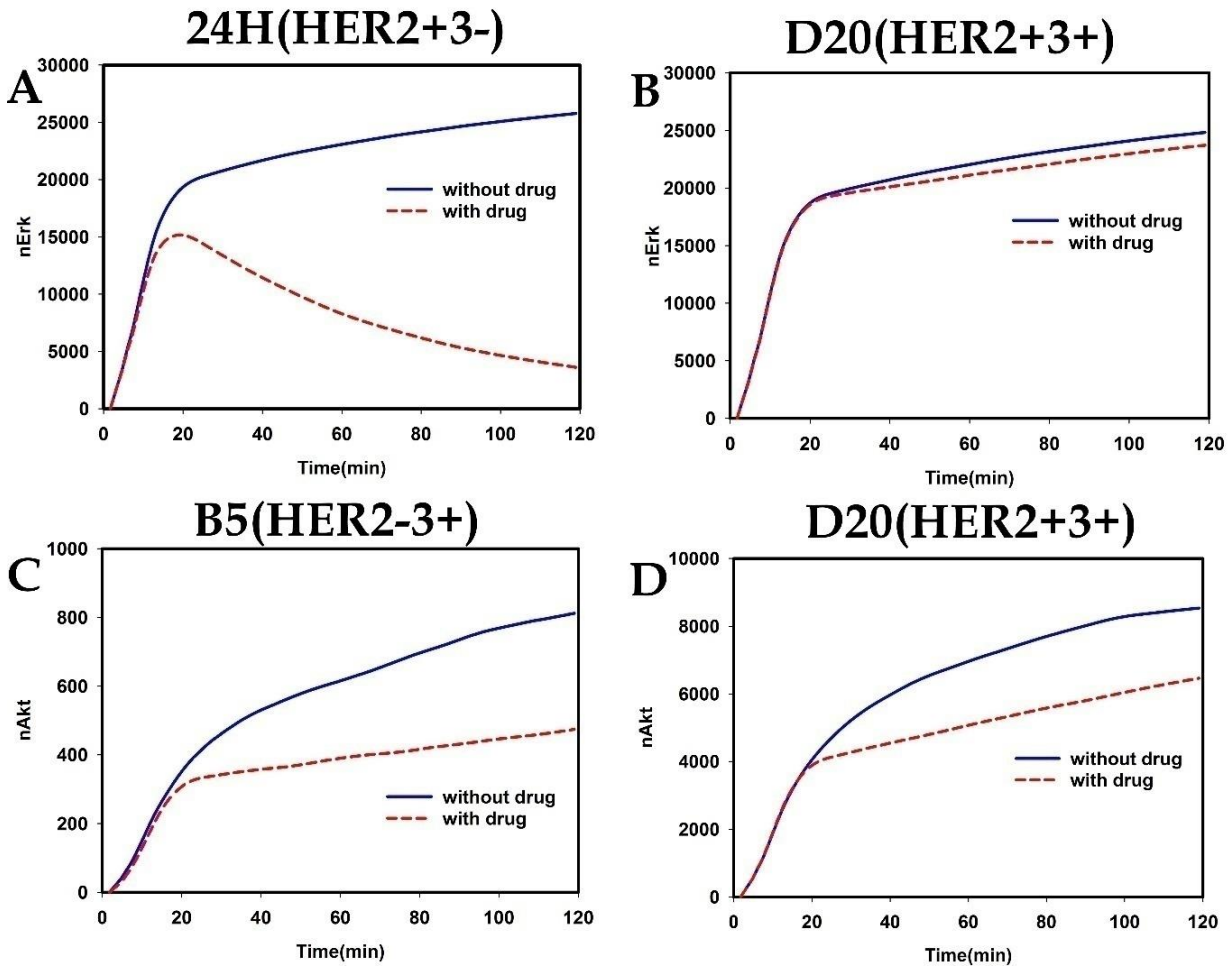


Figure 5.11 Effect of Pertuzumab on Erk and Akt activation. Erk activation in the presence and absence of drug in (A) 24H (HER2+3-) and (B) D20 cell lines. Akt activation in the presence and absence of drug in B5 (C) and D20 (D) cell lines.

A key aspect of receptor signaling is the prediction of how the changes in receptor phosphorylation levels alter the activation patterns of the downstream elements of the involved signaling pathways. The role of HER receptors seems to be different in different cell lines and under different conditions. Dimers formed and pathways that are activated depend upon the specifics of the cell system and treatment conditions. In order to make *in silico* predictions of ErbB signal activation and HER-initiated cellular responses from receptor expression profiles, it becomes necessary to strengthen the data analytics pipeline from more data generated from cell lines and computational models.

5.3 Summary

The results of the *123Model* are taken as the reverse engineered system that predicts receptor (EGFR, HER2 and HER3) activation by ligands (EGF and HRG) and their trafficking within the cell. The data from the model-based study is passed to the agent-based re-engineering model completing the analytics pipeline. The model predicts the relative contributions of the HER receptor in four distinct cell lines and how the changes in the receptor phosphorylation levels alter activation patterns in the downstream cascades of Erk and Akt pathways. Though the absolute values of quantitative estimates cannot be used directly, they provide the key link to correlate receptor expression and signaling outcomes.

By establishing the link between receptor expression, dimer concentrations and their translation to signaling outcomes, the re-engineering model can stand in, for *in silico* clinical trials to study the action mechanisms of HER-targeting drugs and provide insights into the problem of acquired resistance.

6

Application: Effect of combination drugs on HER2 targeted therapies for cancer

6.1 Introduction

Clinical trials in oncology are known to have the highest failure rate compared to other therapeutic areas due to the adaptability of cancer cells at multi-scales to survive any changing environment. Developing treatments involving synergistic drug combinations against cancer has enhanced therapeutic efficiencies. However, analyzing the large number of compound combinations experimentally, is infeasible. The ability of *in silico* methods in combing the innumerably large search space to derive multidrug combinations have enabled model-based analysis to gain traction in the pharmaceutical industries. This chapter is an application of the modeling architecture to develop a model for *in silico* trials for combination drug therapy in HER2 over-expressed breast cancer cell lines.

In recent years, the mechanisms of action and the effect of these drugs have been reverse engineered by *in vitro* and clinical studies. With the existing state of knowledge relating to disease state and effect of drug-treatment, an agent-based model for re-engineering the action of drugs in combination, can consolidate the efforts of the past while giving the opportunity to simulate new scenarios. We use the re-engineered ABM based on the *123Model* developed in Chapter 5 as the platform for studying the effect of drugs on activation of receptors, their downstream signaling, and effect on cancer progression. The drugs taken for this study are Pertuzumab and Trastuzumab as they have shown improved anti-tumor activity in comparison to single-agent therapy in various breast cancer related studies.

Section 6.2 reviews the current state-of-the-art in the field of drug combination studies both *in vitro* and *in silico*. The mechanism of action of Pertuzumab and Trastuzumab on HER2 are described in Section 6.3. Section 6.4 describes the implementation of the model and Section 6.5 illustrates the simulation

results. Section 6.6 summarizes the significance of the model and its future applications.

6.2 Background Literature

Cancer initiation, invasion and metastasis span multiple length and time scales. The combinatorial interplay of molecular pathways acts in concert to deregulate the cellular function complicating the reprogramming of signal flow and interpretation of experimental studies (Chakrabarti, et al., 2012). ABMs in Systems Pharmacology is a fledgling area promising novel insights into complex systems and engineering challenges. ABM captures the systematic organization of data across distinct spatio-temporal scales, enabling the understanding of the effect of disease prognosis and response to treatment. In designing ABMs, hypotheses can be developed as to how individual components or pathways contribute to tissue- or organism-level effects which can be further examined with experimental or clinical studies (Cosgrove, et al., 2015). ABMs assist in identifying key features of a mechanistic target by scaling and extrapolating observations from *in vitro* and *in vivo* studies on animals into possible outcomes *in vivo* in humans (Viceconti, 2016).

Breast cancer has been the most frequently diagnosed life-threatening cancer among women worldwide. Based on the genes expressed, there are five intrinsic subtypes of breast cancer: luminal A, luminal B, HER2 over-expression, basal and normal-like tumors (Dai, et al., 2015). HER2 subtype of breast cancer is associated with over-expression of Human HER2 (ErbB2), a member of transmembrane tyrosine kinase receptor in EGFR (epidermal growth factor receptor) family.

The reduced response of standard therapies for effective inhibition of HER2 signaling in cancer cells have led to the design of targeted therapies. The anti-HER2 treatment strategies have improved over the past few decades and increased the survival median of patients from 1.5 years to 5 years (Araki, et al., 2017). Drugs like humanized therapeutic monoclonal antibodies Trastuzumab & Pertuzumab, taxane based chemotherapy, antibody drug conjugate T-DM1 and dual EGFR/HER2 tyrosine kinase inhibitor lapatinib have shown improved prognosis (Shah and Osipo, 2016). Despite the availability of new drugs, their success is largely hindered by the

resistance that cancerous cells develop towards the drug or due to the lack of improvement when used as a single agent.

Devising combinations of targeted treatments can circumvent some mechanisms of resistance to yield clinical benefits (Lopez and Banerji, 2017). Drugs when used in combination can be additive, superadditive (synergistic) or subadditive (antagonistic). When the combination effect is consistent with the individual drug potencies it is said to be additive. In contrast, antagonism reduces the effect in combination. Drug synergy refers to administering different drugs to enhance the effects, with significant reduction in the dosage of the individual drugs thus reducing its toxicity. The combined effect is greater than their predicted individual potencies (Tallarida, 2011) and are highly efficacious and therapeutically more specific. The Chou-Talalay method quantifies the synergy of drug combinations based on the median-effect equation derived from the mass action law principle. The combination index (CI) offers a quantitative definition for the additive effect (CI=1), synergism (CI<1), and antagonism (CI>1) in drug combinations (Chou, 2010). Isobolograms are graphical constructions based on the assumption of constant relative potency. The coordinate systems comprise of individual drug dosages and commonly a "line of additivity" that distinguishes additive from synergetic and antagonistic interactions (Tallarida, 2011).

As a result of their medicinal potential, combinatorial drug pairs are researched both experimentally and computationally. Current efforts to identify combinatorial anti-cancer therapies rely on experimental methods (Oak, et al., 2012) and clinical trials to assess their safety and efficacy (Mayer, 2015). However researchers face a tremendous challenge in the large search space required for higher-order combinations, multiple drug dosages, temporal optimization of drug administration, costly experimental screening approaches and diversity in cancer cell types and patients.

Computational models have also been developed, with the aim to rationalize and economize the experimental bottlenecks, to concomitantly predict drug responses and efficient anticancer drug combinations (Flobak, et al., 2015). Computational systems biology enable to explore and understand the behaviour of cancer cells under different therapeutic conditions that would otherwise be

impossible with experimental studies. Even though ABM is sparingly used in pharmaceutical context, adopting an ABM approach permits the integration of existing knowledge into a software platform capable of hypothesis testing and designing of new experiments. They are also advantageous in incorporating heterogeneity and to model phenomena occurring across different temporal and spatial scales (Cosgrove, et al., 2015).

In this study, the *123Model* (Chapter 5) is used to explore the effect of drugs Pertuzumab and Trastuzumab and their combinations on breast cancer cell lines so as to effectively analyze drug synergies for effective treatments.

6.3 Biology of action mechanism of drugs

Trastuzumab and Pertuzumab are humanized anti-HER2 monoclonal antibodies designed to potently inhibit HER2-mediated signaling pathways.

Trastuzumab targets the extracellular domain IV of HER2 by the mechanism of antibody-dependent cellular cytotoxicity (ADCC), internalization and degradation, and interfering with dimerization of HER2 (Figure 6.1) (Vu and Claret, 2012). Pertuzumab blocks the dimerization domain II of HER2. Pertuzumab efficiently inhibits ligand-induced dimerization, whereas Trastuzumab inhibits ligand-independent dimerization. Trastuzumab binds tighter than Pertuzumab to HER2 (Lua, et al., 2015). Clinical studies on the combination of Trastuzumab and Pertuzumab has shown significant antitumor effect in an HER2-positive metastatic breast cancer compared to each antibody as a single agent (Baselga, et al., 2012; Harbeck, et al., 2013; Swain, et al., 2015). Although, both show synergistic results in tumor inhibition, the mechanisms underlying the synergy of Trastuzumab and Pertuzumab remains enigmatic. Molecular modeling (Fuentes, et al., 2011) and experimental studies (Lua, et al., 2015) have suggested the colocalization of the two antibodies on to the extracellular domain of HER2 for its synergism. Another proposed mechanism is the inhibition of ligand-independent HER3-HER2 heterodimerization by Trastuzumab on truncation of the extracellular region of HER2, which produces active truncated p95 HER2 and the ligand dependent inhibition of HER2 by Pertuzumab (Goltsov, et al., 2014).

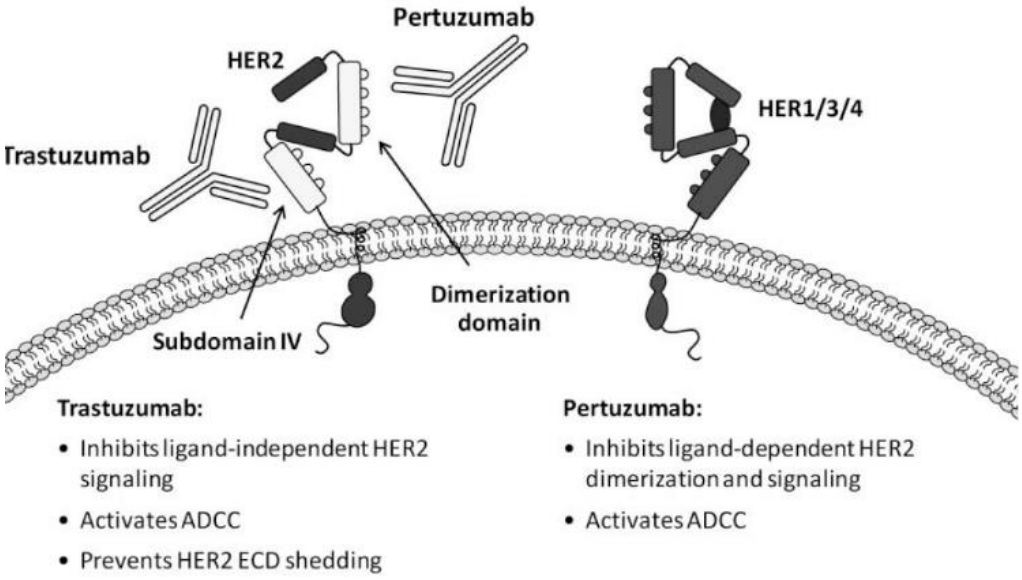


Figure 6.1 Mechanisms of action of Pertuzumab and Trastuzumab (Courtesy Harbeck, et al. (2013))

6.4 The Model

Studies were done by adding drugs Trastuzumab and Pertuzumab individually, and in combination to the D20 and artificial cell line (denoted as A1). This model has been extended to simulate the response kinetics of the signaling network, on addition of HER2 inhibitor drugs and their effect on receptor and Akt activation, tumorigenesis and evolution.

6.4.1 Agents and cell compartments

The elegance of the ABM based on the *123Model* is that it can be easily extended to include drug studies by considering the drug molecules as yet another type of agent with its own set of attributes and actions. The model in addition to HER receptor activation, trafficking, downstream signaling, simulates inhibition of HER2 by Pertuzumab (P) and Trastuzumab (T) separately, and in combination (P+T). This takes the number of agent types to four - ligands (EGF/HRG), receptors (ErbB1/ErbB2/ErbB3), signaling proteins (Erk and Akt) and drugs (Trastuzumab and Pertuzumab).The 3D spatial domain is divided into seven compartments: (1) an extracellular domain (Outside), (2) plasma membrane (PM), (3) early endosome (EE), (4) late endosome (LE) (5) lysosome and (6) cytoplasm.

Similar to the events mentioned in the previous chapter, ligands EGF and HRG bind to receptors on plasma membrane (PM) from the outside compartment to initiate signaling. The bound/unbound monomers and dimers are internalised to the early endosome (EE) from where they either accumulate in the late endosome (LE) for degradation or recycle back to the PM. On being flagged for degradation, they move to the lysosome. Sentinel proteins, Erk and Akt, present in the cytoplasm, are activated by signaling through receptor activation. Drugs, Trastuzumab and Pertuzumab, are present in the "Outside" compartment and bind to ErbB2 on different domains and inactivate it through various mechanisms. Drug and receptor agent attributes are defined in Figure 6.2.

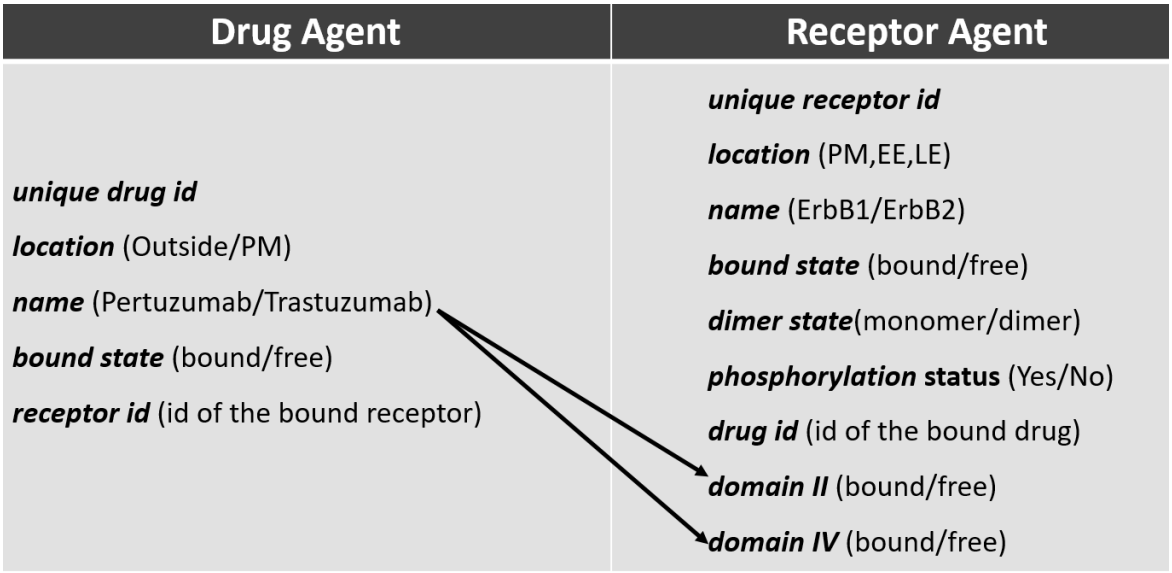


Figure 6.2 Linkages of Drug and Receptor Agent attributes

The drug studies were done on two cell lines where HER2 is overexpressed-D20 and an artificial cell line A1 (Table 6.1). These cell lines were studied on three scenarios - P, T and P+T for various doses.

Table 6.1 HER expression levels in artificial (A1) and D20 cell lines

List of cell lines	Abundance of EGFR, HER2 and HER3 (molecules/cell)		
Notation	**EGFR**	**HER2**	**HER3**
A1	80000	300000	40000
D20	85896	643777	28770

6.4.2 Rules

The drug action is based on the outcomes of drug studies conducted in animal models, clinical trials and computational studies. These are encapsulated as a set of macro rules as shown in Table 6.2. It may be noted that there are no micro rules because, only the action mechanism is being tested to observe the ErbB2-targeted inhibition activity of the drugs. There are no experiment specific rules, like the data of a drug trial.

Table 6.2 Macro Rules for Drug action

Macro Rules
(1) Pertuzumab binds to domain II and Trastuzumab binds to domain IV of HER2
(2) Pertuzumab blocks ligand dependent receptor heterodimerisation of HER2 and HER3
(3) Trastuzumab blocks ligand independent dimerization of HER2
(4) Trastuzumab inhibits the shedding of the extracellular domain (ECD) of HER2
(5) Trastuzumab and Pertuzumab can cause different inhibitory effects on PI3K and MAPK pathways despite targeting the same receptor HER2
(6) Pertuzumab blocks the association of HER2 with HER3 when cells are stimulated with HER3 ligand, HRG
(7) Trastuzumab and Pertuzumab attract immune cells to tumor sites that overexpress HER2 by antibody-dependent cellular cytotoxicity (ADCC)
(8) Trastuzumab and Pertuzumab inhibits HER2 activation and suppresses Akt phosphorylation

As in Figure 6.1, a drug agent can bind to domain II or/and IV of the receptor agent. Domain II attribute of a receptor agent is set to “bound” on binding of drug agent Pertuzumab. Whereas, domain IV is set to “bound” on binding of Trastuzumab. In case of combination, both domains are set to “bound” on binding of the drug agent.

Mechanism of action of drugs P and T		
i.	**P binds to Domain II**	➢ Inhibits formation of ligand-dependent HER2 heterodimers ➢ Activates ADCC
ii.	**T binds to Domain IV**	➢ Inhibits mutation of HER2 to p95 HER2 ➢ Blocks ligand-independent HER2 signaling ➢ Flags cells for ADCC
iii.	**P+T binds to Domain II and IV**	➢ Inhibits formation of ligand dependent and independent HER2 dimers ➢ Activates ADCC ➢ Inhibits mutation of HER2 to p95 HER2

In drug combinations, (i), (ii) and (iii) events can occur depending on whether P, T or P+T is bound to the receptor.

ADCC and Tumor size
➢ **When drug bound HER2 (denoted as $E2^D$) crosses a threshold value, ADCC occurs. A fixed number of $E2^D$ are removed from the system.** ➢ **Tumor size (Tsize) is a global variable that is incremented based on the activation value of Akt. When ADCC occurs, Tsize shrinks proportional to the amount of ADCC.**

6.5 Simulation Results

In silico drug experiments were carried out 20 times for each cell line and parameter values averaged. The major assumptions in the simulation are the following:

(1) Drug is introduced at the beginning of the simulation.

(2) Agents including drugs are recycled during dissociation in EE, internalization and degradation.

(3) Drugs bind to the receptor at the rate of 0.1/min.

(4) Drug dosage is translated to corresponding numbers of 25000/50000/75000/ 100000 drug agents. Actual relationship between dosage and the number of agents in the system has to be worked out in real trials.

(5) The T+P combination is in the ratio 40-60%.

Each timestep is a period in which the agents can interact with each other and the drugs to carry out a unit action. The timecourse continues to trace the system. Drug agents are replenished when their numbers fall below a critical number. In actual trials this can vary from 2-3 weeks.

6.5.1 Simulation on Artificial cell line

As Pertuzumab and Trastuzumab bind to different domains of HER2 and suppress different aspects of HER2 signaling, their combination is expected to inhibit HER2 more effectively. A1 was treated with Pertuzumab and Trastuzumab, both alone, and in combination to study their effect on Akt activation (Figure 6.3). Akt showed a markedly decreased activation in combination compared to either drug alone. The curve (in red) is the control before introduction of any drug into the system. The phosphorylation values for Akt (responsible for proliferative activity of cells) were determined as a function of HER2-HER3 and HER3-HER1 dimer activation.

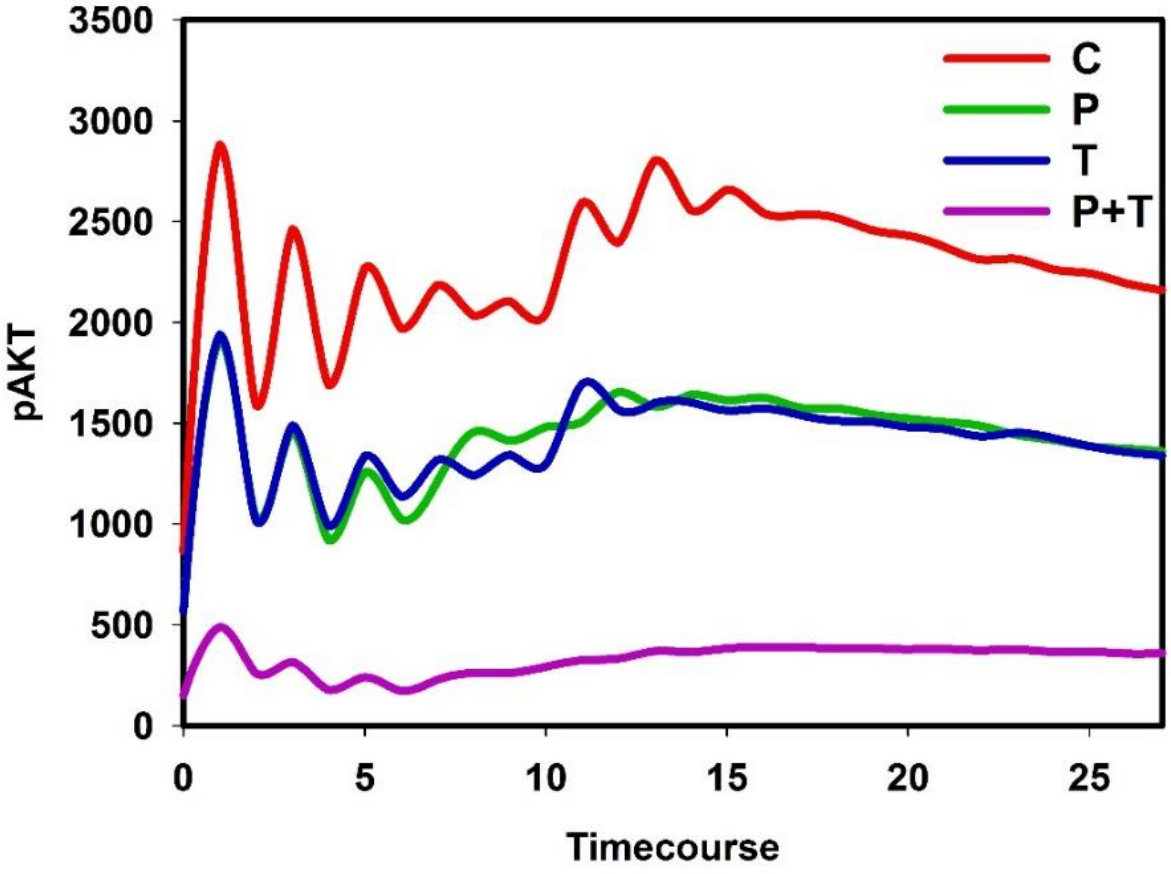


Figure 6.3 Akt activation plotted in A1 cell line on addition of Pertuzumab (P), Trastuzumab (T) and both (P+T) along with the Control(C).

Figure 6.4 demonstrates the enhancement of ADCC when both the drugs are added in combination. Trastuzumab exhibits slightly higher ADCC compared to Pertuzumab. This holds consistent with the hypothesis that ADCC can be a mechanism for the decrease in signal activation.

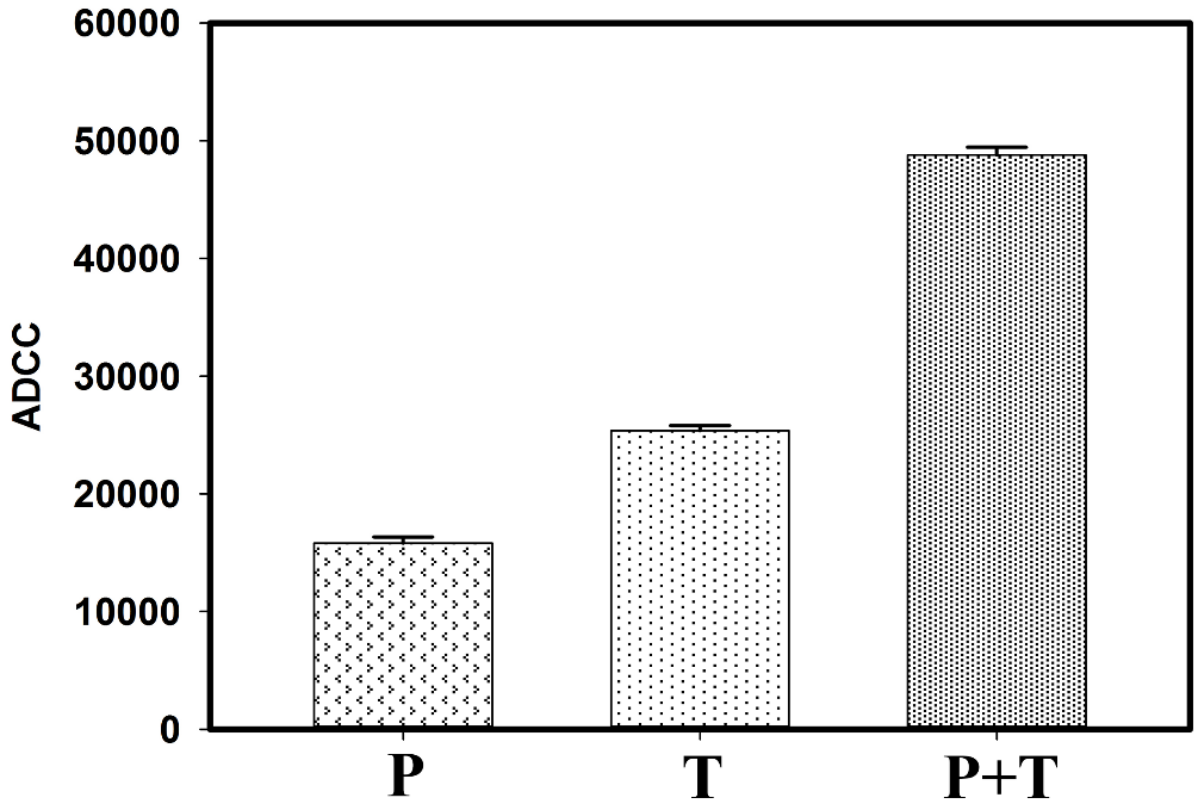


Figure 6.4 ADCC of A1 cells on treatment with Pertuzumab (P), Trastuzumab (T) and both (P+T). ADCC measured for P-50000, T-50000, P+T- 75000; Bars represent standard error for 20 simulations.

Drugs exhibited significant anti-tumor activity in comparison with the control. Out of the three simulations, T+P combination therapy showed a significantly stronger anti-tumor activity compared to their monotherapy (Figure 6.5). T+P showed ~25% decreased anti-tumor activity compared to T and P alone. However, Pertuzumab and Trastuzumab did not show substantial difference in their anti-tumor activity.

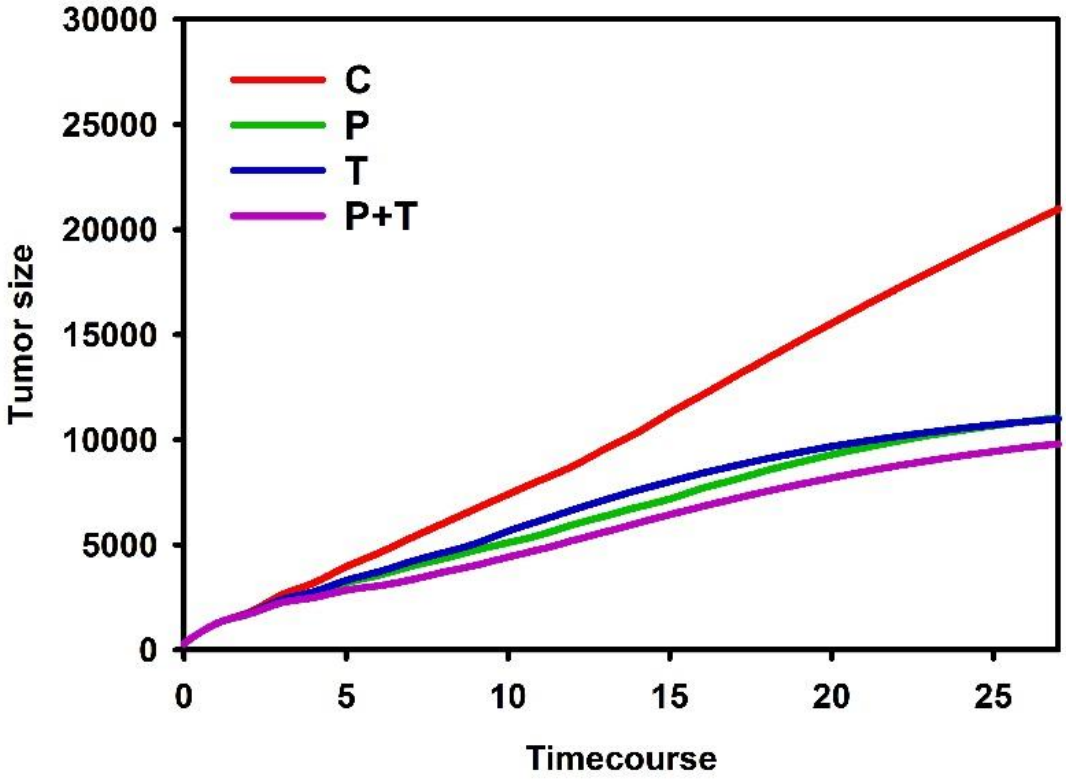


Figure 6.5 Tumor size in A1 cell line treated with P-50000, T-5000 and T+P-75000.

To examine the changes in Akt activation on treatment with different dosages, A1 cell line was treated with Trastuzumab, Pertuzumab, and both drugs simultaneously for doses (represented as agents in the model) ranging from 25000-100000 agents. P+T proved to be saturated after a dose of 75000 agents (Figure 6.6 C). Whereas, Akt activation decreases on increasing the dose of Pertuzumab (Figure 6.6 A) and Trastuzumab (Figure 6.6 B). The anti-proliferative activity of pAkt is more sensitive to P than to T.

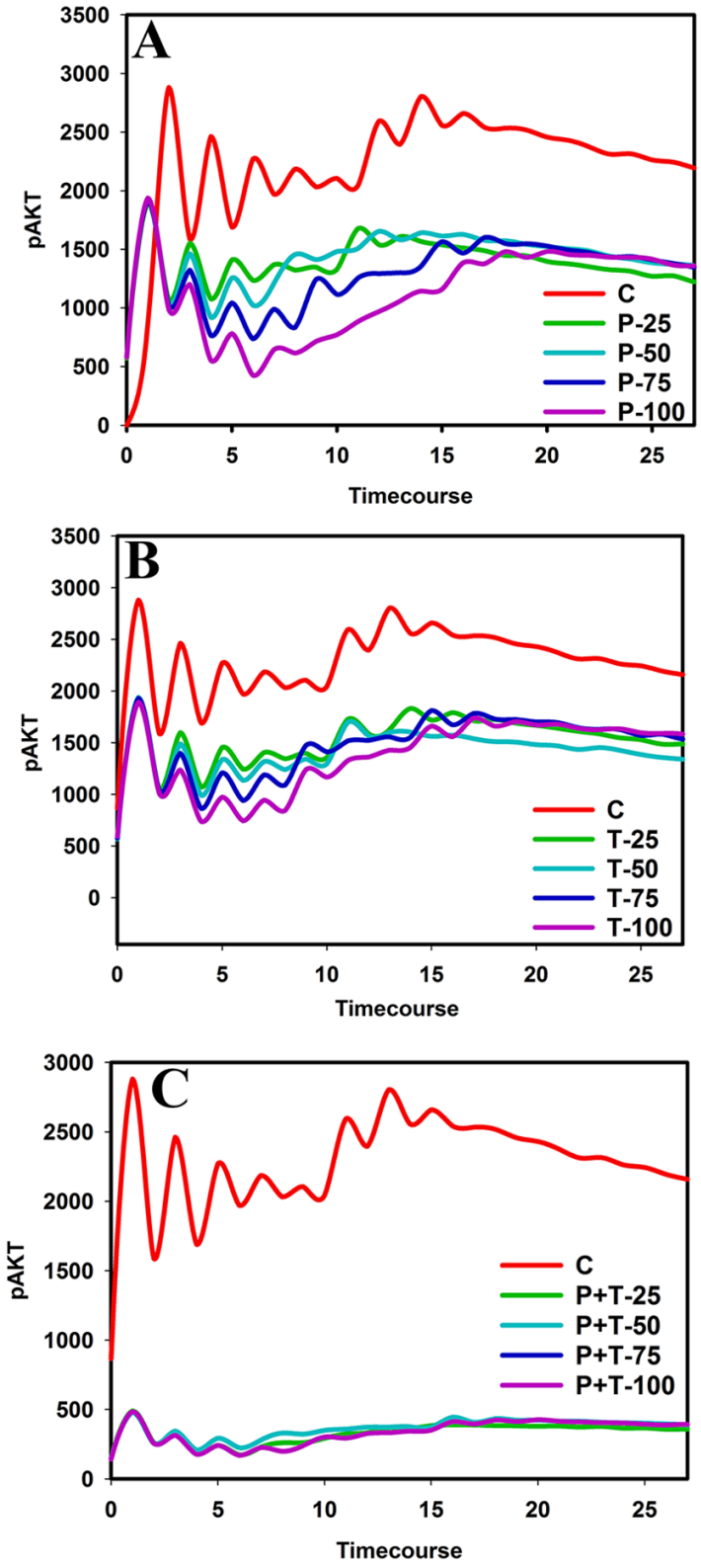


Figure 6.6 Activation of Akt for various doses of drug. A. A1 cell line treated with Pertuzumab at various doses (or number of agents used for simulation); 25000 (P-25), 50000 (P-50), 75000 (P-75), 100000 (P-100). B. A1 cell line treated with Trastuzumab at various doses (or number of agents used for simulation); 25000 (T-25), 50000 (T-50), 75000 (T-75), 100000 (T-100). C. A1 cell line treated with Pertuzumab and Trastuzumab at various doses (or number of agents used for simulation); 25000 (P+T-25), 50000 (P+T-50), 75000 (P+T-75), 100000 (P+T-100).

6.5.2 Simulation on D20 cell line

Given the importance of downstream signaling in tumorigenesis, the timecourse activation of Akt was examined in D20 cell line. Akt activation in D20 cell line also showed decreased activation in combination compared to monotherapy (Figure 6.7). A1 cell line is different from D20 in the expression of HER1 and HER3 resulting in the formation of more HER2-HER3 heterodimers that lead to a higher Akt activation level. This proves that receptor concentration and affinity plays a significant role in target therapies.

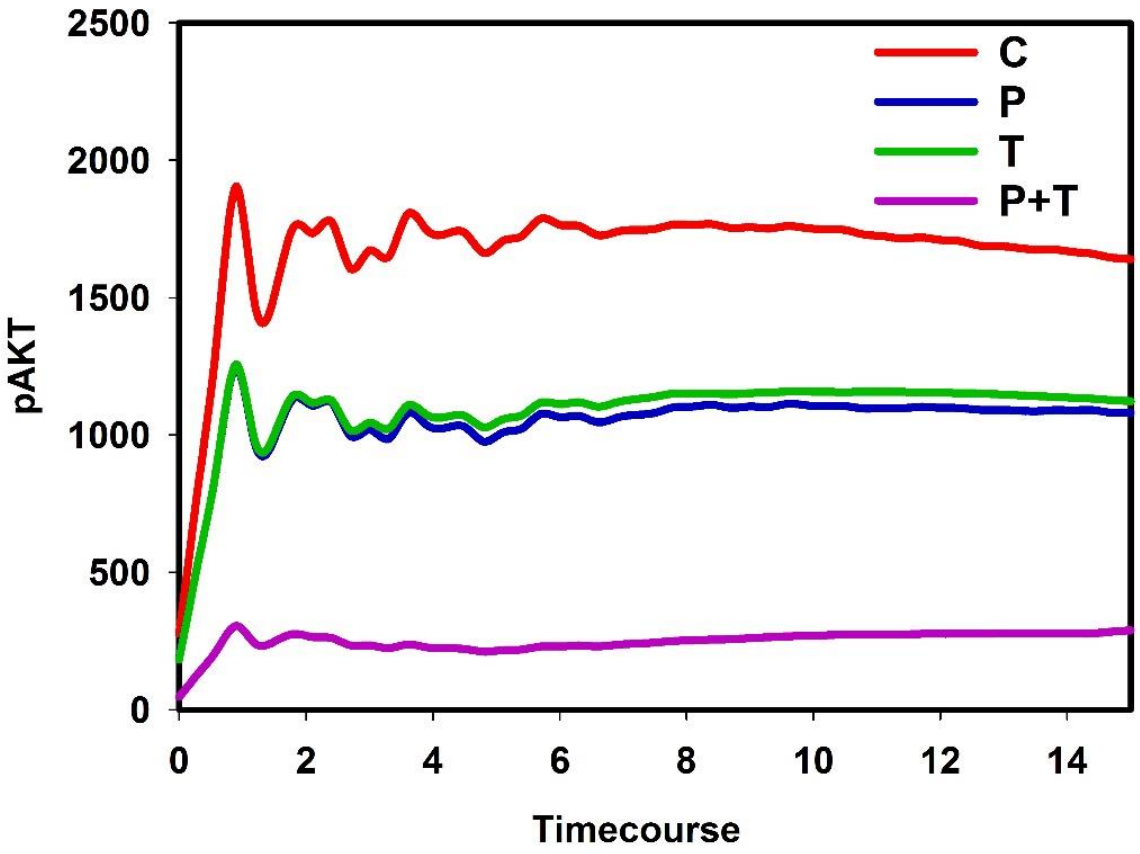


Figure 6.7 Akt activation plotted in D20 cell line on addition of Pertuzumab (P), Trastuzumab (T) and both (P+T) along with the Control(C).

ADCC was also assessed to determine the role of drugs in the inhibition of Akt signaling. ADCC in D20 cell lines treated with dual drug were strongly inhibited compared to Trastuzumab and Pertuzumab treatment alone (Figure 6.8).

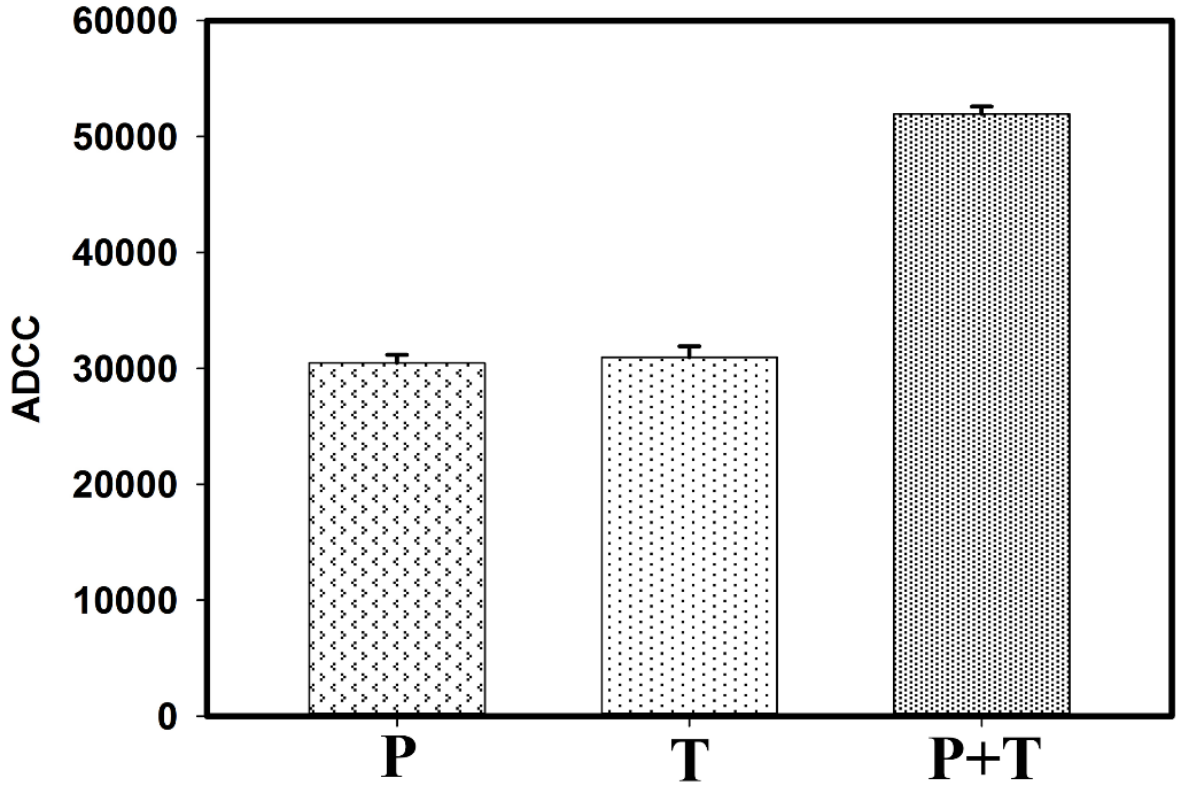


Figure 6.8 ADCC of D20 cells on treatment with Pertuzumab (P), Trastuzumab (T) and both (P+T). ADCC measured for P-50000, T-50000, P+T- 75000; Bars represent standard error for 20 simulations.

Tumor size of P+T decrease by ~16.7% from P and T alone (Figure 6.9). T and P shows a slight difference in their anti-tumor activity. Figure 6.10 shows the various doses of P, T and P+T used to determine the Akt activation. P+T showed tendency of saturation even at a lower dose of 50000 agents.

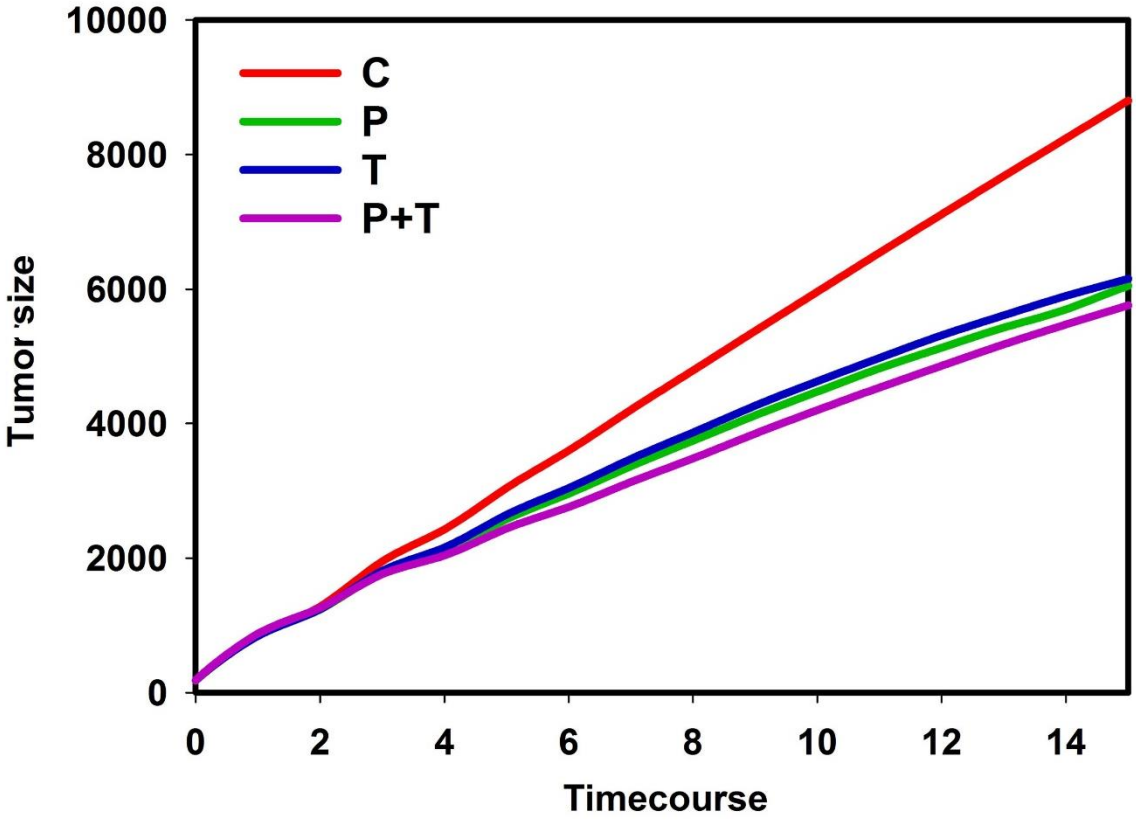


Figure 6.9 Tumor size in D20 cell line treated with P-50000, T-5000 and T+P-75000.

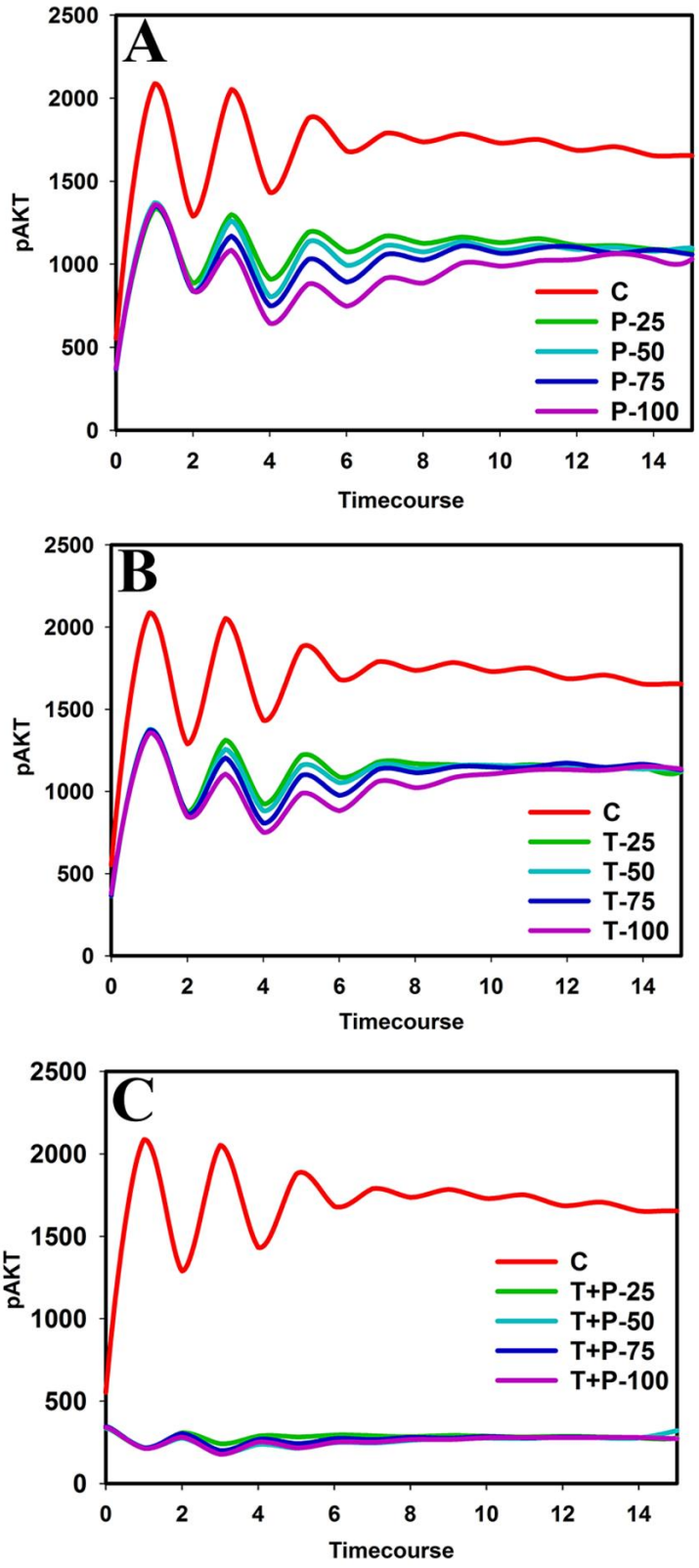


Figure 6.10 Activation of Akt for various doses of drug. A. D20 cell line treated with Pertuzumab at various doses (or number of agents used for simulation); 25000 (P-25), 50000 (P-50), 75000 (P-75), 100000 (P-100). B. D20 cell line treated with Trastuzumab at various doses (or number of agents used for simulation); 25000 (T-25), 50000 (T-50), 75000 (T-75), 100000 (T-100). C. D20 cell line treated with Pertuzumab and Trastuzumab at various doses (or number of agents used for simulation); 25000 (P+T-25), 50000 (P+T-50), 75000 (P+T-75), 100000 (P+T-100).

Even though, combination drugs show enhanced effects than each drug given alone, they do not necessarily indicate synergism (Chou, 2010). The determination of synergism requires quantitative analysis like isobolographic analysis, combination index (CI) (Tallarida, 2011) which is beyond the scope of this study. The aim of this simulation is to showcase how the agent-based model of the data analytics pipeline can be used for *in silico* drug trials (simulations) by defining the drug molecules as agents and following their action in the system.

6.6 Summary

Co-expression of the different receptors, the diversity in their ligand-independent and ligand-dependent activation, variation in their preference towards dimerization partners, and receptor-dependent specificity in cells play a major role in both redundancy in the HER network interaction and effective drug target identification (Goltsov, et al., 2014). We have demonstrated the effect of using drugs in combination than either of them alone, using the modeling pipeline architecture. The complementary action mechanisms of Trastuzumab and Pertuzumab provides complete blockade of HER2 compared to using them individually. Application of the modeling pipeline for combination studies reduces the number of costly experiments and test "what if" hypotheses that may not even be possible to achieve in the lab.

7

Conclusion

7.1 Summary

The remarkable achievements in molecular and cellular biology over the past two decades have resulted in the generation of huge amounts of experimental data. To understand precisely the nature of biological systems requires supplementing the empirical with the quantifiable, and hence, a synergistic collaboration of experimental and computational methods. Most of the successes in biological research have come from the reductionist agenda - by reverse engineering biological processes in terms of the smallest entities. Each such exercise is only a partial representation of the system. However, the agenda of integration of these parts has to be addressed, in order to complete the story. This is the main objective which has been achieved by the reverse engineering - re-engineering pipeline architecture.

The popular reductionist approach of reverse engineering with equation-based modeling has succeeded in throwing light on the process of signal transduction, in parts and patches, generating repositories of unconnected data. Re-engineering by analytics and learning models can probe into these databases to help piece together a system level understanding of signaling networks. The data generated by the top-down approach of mathematical modeling connects to the bottom up model of agents by encoding the mechanistic and parametric information into data-enriched rules.

'Act local to discover global' is the basis for the agent-based re-engineering model for ErbB signal transduction (Chapter 4). Particularly, in an agent-based representation, patterns, structures, and system behaviours emerge by self-organization of interacting heterogeneous agents. Integrating disparate sources of information and data on ErbB signaling by a data analytics pipeline, can provide new directions in translational research in ErbB network biology.

By establishing the link between receptor expression, dimer concentrations and their translation to signaling outcomes (Chapter 5), the agent-based re-engineering model can be used to conduct *in silico* experiments for validation, predictions and testing hypotheses. Translation of basic mechanistic knowledge into clinically effective therapeutics is another application of interest (Chapter 6). The HER-targeting drugs act in many ways - by inhibition of dimerization, prevention of receptor formation, blocking ligand binding or blocking dimer complexes. By introducing drugs as agents all the above action mechanisms were studied for drugs Pertuzumab and Trastuzumab both alone and in combination.

7.2 Future Perspectives

The time scale of events varies from minutes to days for *in silico* drug trials. While signaling events at the molecular level take place in minutes, the life of the drug in the system is of the order of days. The current implementation does not take the granularity into consideration. As the rules to decide the fate of molecular species, depends on a comparison of rates and not the rates themselves, a coarse-grained time interval is used for each iteration. A timestep (time of an iteration) is the interval for agent interactions leading to the occurrence of an event. A scheme for taking care of fine to coarse-grained time resolution will be implemented in the next version.

The ABM for ErbB signal transduction is built with parameter values taken from models in literature. The rates are formed from the dependency map in such a way that all major cellular events are represented. An elaborate sensitivity analysis needs to be carried out by perturbing the rates and by changing the choice of rules.

ABMs have been used to model biological complexity across multiple scales from molecules to cells to societies. However, mathematical and statistical tools for analysis are lacking and conclusions have to be derived from simulations alone. This is true of many of the open source tools also. Plugins for postprocessing of simulation data would go a long way in making ABMs more meaningful and attractive. The framework on which this work was carried out, takes input from data files and logs output data that has to be analyzed offline. Building tools for visualization, graphing, and analysis, embedded in a GUI would make it user-friendly.

Ideally, an agent model simulates a system by creating its digital replica. But, in practice, it is only possible to build a microcosm of the system. The size of the model, initial conditions and parameters must be chosen with discretion so that the effects of miniaturization are minimized. Nevertheless, models still get too large which calls for research on two mathematical aspects of ABM, namely, optimization and control.

7.3 Concluding Remarks

Modeling complex systems with millions of agents, needs high computational requirements. Problem decomposition is one way to get around this problem. In the context of signaling pathways, early signaling events and the downstream cascade can be separated and modeled separately. The output of the former can be passed to the latter. High performance computing, code parallelization and cloud migration are other options available for large scale systems.

The action of drugs on the ErbB target has been studied by *in vitro* and *in vivo* experiments on cells having a wide range of ErbB co-expression. *In silico* experiments can be conducted to investigate the action mechanisms of drugs on receptor-mediated signaling, dose-response modeling, efficacy of drugs and the problem of drug resistance. Apart from guiding the *in vitro* and *in vivo*, the data generated from the wet lab can dynamically modify the model in a feedback loop. As big data is generated by high-throughput bio-technologies and computing hardware and software become more powerful and cheaper, the trilogy of *in silico*, *in vitro* and *in vivo* can lighten the burden on animal studies and shorten the lead time for drug development.

Towards an integrated approach for a better understanding of a complex phenomena, one needs to begin at the level where there is maximum information and understanding, and move from the known to the unknown. In a more general sense, this work can be summarized as an effort to drive reductionism towards emergence using the power of data and deep data analytics. This is a philosophy that can have a far-reaching impact on the way data science works to elucidate the dynamics of complex systems.

REFERENCES

Ahmad, I.*, et al.* (2011) HER2 overcomes PTEN (loss)-induced senescence to cause aggressive prostate cancer, *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 16392-16397.

Al-Lazikani, B., Banerji, U. and Workman, P. (2012) Combinatorial drug therapy for cancer in the post-genomic era, *Nature Biotechnology*, **30**, 679-691.

An, G.*, et al.* (2009) Agent-based models in translational systems biology, *Wiley Interdisciplinary Reviews-Systems Biology and Medicine*, **1**, 159-171.

Araki, K.*, et al.* (2017) First report of eribulin in combination with pertuzumab and trastuzumab for advanced HER2-positive breast cancer, *The Breast*, **35**, 78-84.

Baselga, J.*, et al.* (2012) Pertuzumab plus Trastuzumab plus Docetaxel for Metastatic Breast Cancer, *New England Journal of Medicine*, **366**, 109-119.

Becker, K.*, et al.* (2013) Reverse-Engineering Post-Transcriptional Regulation of Gap Genes in Drosophila melanogaster, *Plos Computational Biology*, **9**.

Berryman, M.J. and Angus, S.D. (2010) Tutorials on Agent-based Modelling with NetLogo and Network Analysis with Pajek. In, *Complex Physical, Biophysical and Econophysical Systems*. WORLD SCIENTIFIC, pp. 351-375.

Birtwistle, M.R.*, et al.* (2007) Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses, *Molecular Systems Biology*, **3**.

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

Bonabeau, E. (2002) Agent-based modeling: Methods and techniques for simulating human systems, *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 7280-7287.

Bonneau, R.*, et al.* (2007) A predictive model for transcriptional control of physiology in a free living cell, *Cell*, **131**, 1354-1365.

Chakrabarti, A.*, et al.* (2012) Multiscale Models of Breast Cancer Progression, *Ann Biomed Eng*, **40**.

Chen, W.W.*, et al.* (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data, *Molecular Systems Biology*, **5**, 19.

Chou, T.C. (2010) Drug combination studies and their synergy quantification using the Chou-Talalay method, *Cancer Res*, **70**, 440-446.

Chung, I.*, et al.* (2010) Spatial control of EGF receptor activation by reversible dimerization on living cells, *Nature*, **464**, 783-U163.

Chung, J.K.*, et al.* (1997) STAT3 serine phosphorylation by ERK-dependent and -independent pathways negatively modulates its tyrosine phosphorylation, *Molecular and Cellular Biology*, **17**, 6508-6516.

Cosgrove, J.*, et al.* (2015) Agent-Based Modeling in Systems Pharmacology, *CPT: Pharmacometrics & Systems Pharmacology*, **4**, 615-629.

Creamer, M.S.*, et al.* (2012) Specification, annotation, visualization and simulation of a large rule-based model for ERBB receptor signaling, *Bmc Systems Biology*, **6**.

Crombach, A.*, et al.* (2012) Efficient Reverse-Engineering of a Developmental Gene Regulatory Network, *Plos Computational Biology*, **8**.

Dai, X.F.*, et al.* (2015) Breast cancer intrinsic subtype classification, clinical use and future trends, *American Journal of Cancer Research*, **5**, 2929-2943.

Emonet, T.*, et al.* (2005) AgentCell: a digital single-cell assay for bacterial chemotaxis, *Bioinformatics*, **21**, 2714-2721.

Epstein, J.M. and Axtell, R. (1996) *Growing Artificial Societies: Social Science from the Bottom Up*. The MIT Press.

Faeder, J.R., Blinov, M.L. and Hlavacek, W.S. (2009) Rule-based modeling of biochemical systems with BioNetGen, *Methods in molecular biology (Clifton, N.J.)*, **500**, 113-167.

Ferguson, K.M.*, et al.* (2003) EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization, *Molecular Cell*, **11**, 507-517.

Figueredo, G.P.*, et al.* (2014) Comparing Stochastic Differential Equations and Agent-Based Modelling and Simulation for Early-Stage Cancer, *Plos One*, **9**.

Flobak, A.*, et al.* (2015) Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling, *Plos Computational Biology*, **11**.

Folcik, V.A., An, G.C. and Orosz, C.G. (2007) The Basic Immune Simulator: An agent-based model to study the interactions between innate and adaptive immunity, *Theoretical Biology and Medical Modelling*, **4**.

Fuentes, G.*, et al.* (2011) Synergy between trastuzumab and pertuzumab for human epidermal growth factor 2 (Her2) from colocalization: an in silico based mechanism, *Breast Cancer Research*, **13**.

Gillespie, D.T. (1977) Exact Stochastic Simulation of Coupled Chemical Reactions, *Journal of Physical Chemistry*, **81**, 2340.

Goltsov, A., *et al.* (2014) Systems Analysis of Drug-Induced Receptor Tyrosine Kinase Reprogramming Following Targeted Mono- and Combination Anti-Cancer Therapy, *Cells*, **3**, 563-591.

Gong, C., *et al.* (2015) Integrated analysis reveals that STAT3 is central to the crosstalk between HER/ErbB receptor signaling pathways in human mammary epithelial cells, *Molecular bioSystems*, **11**, 146-158.

Gorochowski, T.E., *et al.* (2012) BSim: An Agent-Based Tool for Modeling Bacterial Populations in Systems and Synthetic Biology, *Plos One*, **7**.

Griffin, W.A. (2006), *Biological Theory*, **1**, 404–409.

Harbeck, N., *et al.* (2013) HER2 Dimerization Inhibitor Pertuzumab - Mode of Action and Clinical Data in Breast Cancer, *Breast Care*, **8**, 49-55.

Hasdemir, D., Hoefsloot, H.C.J. and Smilde, A.K. (2015) Validation and selection of ODE based systems biology models: how to arrive at more reliable decisions, *Bmc Systems Biology*, **9**.

Hendriks, B.S., *et al.* (2006) Computational modelling of ErbB family phosphorylation dynamics in response to transforming growth factor alpha and heregulin indicates spatial compartmentation of phosphatase activity, *Iee Proceedings Systems Biology*, **153**, 22-33.

Hendriks, B.S., *et al.* (2003) Quantitative analysis of HER2-mediated effects on HER2 and epidermal growth factor receptor endocytosis - Distribution of homo- and heterodimers depends on relative HER2 levels, *Journal of Biological Chemistry*, **278**, 23343-23351.

Hogg, J.S., *et al.* (2014) Exact Hybrid Particle/Population Simulation of Rule-Based Models of Biochemical Systems, *Plos Computational Biology*, **10**.

Hsieh, M.Y., *et al.* (2010) Spatio-temporal modeling of signaling protein recruitment to EGFR, *Bmc Systems Biology*, **4**.

Janes, K.A. and Lauffenburger, D.A. (2013) Models of signalling networks - what cell biologists can gain from them and give to them, *Journal of Cell Science*, **126**, 1913-1921.

Kaul, H. and Ventikos, Y. (2015) Investigating biocomplexity through the agent-based paradigm, *Briefings in Bioinformatics*, **16**, 137-152.

Kholodenko, B.N., *et al.* (1999) Quantification of short term signaling by the epidermal growth factor receptor, *Journal of Biological Chemistry*, **274**, 30169-30181.

Kim, M., *et al.* (2005) Re-engineering software architecture of home service robots: A case study. *27th International Conference on Software Engineering (ICSE 2005)*. St Louis, MO, pp. 505-513.

Koide, T., Pang, W.L. and Baliga, N.S. (2009) The role of predictive modelling in rationally re-engineering biological systems, *Nature Reviews Microbiology*, **7**, 297-305.

Kusiak, A., Larson, T.N. and Wang, J.R. (1994) REENGINEERING OF DESIGN AND MANUFACTURING PROCESSES, *Computers & Industrial Engineering*, **26**, 521-536.

Lee, B.Y.*, et al.* (2010) A Computer Simulation of Employee Vaccination to Mitigate an Influenza Epidemic, *American Journal of Preventive Medicine*, **38**, 247-257.

Li, Y.*, et al.* (2016) Agent-Based Modeling of Chronic Diseases: A Narrative Review and Future Research Directions, *Preventing Chronic Disease*, **13**.

Liu, Y.*, et al.* (2007) A Multiscale Computational Approach to Dissect Early Events in the Erb Family Receptor Mediated Activation, Differential Signaling, and Relevance to Oncogenic Transformations, *Annals of biomedical engineering*, **35**, 1012-1025.

Lobo, D., Hammelman, J. and Levin, M. (2016) MoCha: Molecular Characterization of Unknown Pathways, *Journal of Computational Biology*, **23**, 291-297.

Lopez, J.S. and Banerji, U. (2017) Combine and conquer: challenges for targeted therapy combinations in early phase trials, *Nature Reviews Clinical Oncology*, **14**, 57-66.

Lua, W.-H.*, et al.* (2015) A search for synergy in the binding kinetics of Trastuzumab and Pertuzumab whole and F(ab) to Her2, *npj Breast Cancer*, **1**, 15012.

Luke, S.*, et al.* (2005) MASON: A multiagent simulation environment, *Simulation-Transactions of the Society for Modeling and Simulation International*, **81**, 517-527.

Ma, J.*, et al.* (2014) Targeting of erbB3 receptor to overcome resistance in cancer treatment, *Molecular Cancer*, **13**.

Macal, C.M. and North, M.J. (2010) Tutorial on agent-based modelling and simulation, *Journal of Simulation*, **4**, 151-162.

Mayer, E.L. (2015) Targeting Breast Cancer with CDK Inhibitors, *Current Oncology Reports*, **17**.

Mujoo, K.*, et al.* (2014) Regulation of ERBB3/HER3 signaling in cancer, *Oncotarget*, **5**, 10222-10236.

North, M.J.*, et al.* (2013) Complex adaptive systems modeling with Repast Simphony, *Complex Adaptive Systems Modeling*, **1**, 3.

Oak, P.S.*, et al.* (2012) Combinatorial treatment of mammospheres with trastuzumab and salinomycin efficiently targets HER2-positive cancer cells and cancer stem cells, *International Journal of Cancer*, **131**, 2808-2819.

Orton, R.J., *et al.* (2008) Computational modelling reveals feedback redundancy within the epidermal growth factor receptor/extracellular-signal regulated kinase signalling pathway, *Iet Systems Biology*, **2**, 173-183.

Perez-Rodriguez, G., *et al.* (2015) Agent-Based Spatiotemporal Simulation of Biomolecular Systems within the Open Source MASON Framework, *Biomed Research International*.

Perez, L. and Dragicevic, S. (2009) An agent-based approach for modeling dynamics of contagious disease spread, *International journal of health geographics*, **8**, 50-50.

Politopoulos, I. (2007) Review and Analysis of Agent-based Models in Biology. University of Liverpool.

Quo, C.F., *et al.* (2012) Reverse engineering biomolecular systems using -omic data: challenges, progress and opportunities, *Briefings in Bioinformatics*, **13**, 430-445.

Resat, H., *et al.* (2003) An integrated model of epidermal growth factor receptor trafficking and signal transduction, *Biophysical Journal*, **85**, 730-743.

Roskoski, R. (2014) The ErbB/HER family of protein-tyrosine kinases and cancer, *Pharmacological Research*, **79**, 34-74.

Roskoski, R., Jr. (2014) The ErbB/HER family of protein-tyrosine kinases and cancer, *Pharmacological research*, **79**, 34-74.

Sahin, O., *et al.* (2009) Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance, *Bmc Systems Biology*, **3**, 20.

Schoeberl, B., *et al.* (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors, *Nature Biotechnology*, **20**, 370-375.

Scypinski, S., Sadlowski, T. and Baiano, J. (1996) Regrouping, reforming and re-engineering: Applying robotics to new challenges, *Journal of Automatic Chemistry*, **18**, 147-148.

Shah, D. and Osipo, C. (2016) Cancer stem cells and HER2 positive breast cancer: The story so far, *Genes & Diseases*, **3**, 114-123.

Shankaran, H., *et al.* (2008) Quantifying the effects of co-expressing EGFR and HER2 on HER activation and trafficking, *Biochemical and Biophysical Research Communications*, **371**, 220-224.

Shankaran, H., *et al.* (2013) Model-Based Analysis of HER Activation in Cells Co-Expressing EGFR, HER2 and HER3, *PLOS Computational Biology*, **9**, e1003201.

Shankaran, H., *et al.* (2013) Model-Based Analysis of HER Activation in Cells Co-Expressing EGFR, HER2 and HER3, *Plos Computational Biology*, **9**.

Sneddon, M.W., Faeder, J.R. and Emonet, T. (2011) Efficient modeling, simulation and coarse-graining of biological complexity with NFsim, *Nature Methods*, **8**, 177-U112.

Sorkin, A. and Goh, L.K. (2009) Endocytosis and intracellular trafficking of ErbBs, *Experimental Cell Research*, **315**, 683-696.

Stamos, J., Sliwkowski, M.X. and Eigenbrot, C. (2002) Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor, *Journal of Biological Chemistry*, **277**, 46265-46272.

Strandhagen, J.O. and Skarlo, T. (1995) A Manufacturing Business Process Reengineering Method: Design and Redesign of a Production Control Model. In Browne, J. and O'Sullivan, D. (eds), *Re-engineering the Enterprise: Proceedings of the IFIP TC5/WG5.7 Working Conference on Re-engineering the Enterprise, Galway, Ireland, 1995*. Springer US, Boston, MA, pp. 187-198.

Swain, S.M.*, et al.* (2015) Pertuzumab, Trastuzumab, and Docetaxel in HER2-Positive Metastatic Breast Cancer, *New England Journal of Medicine*, **372**, 724-734.

Tallarida, R.J. (2011) Quantitative Methods for Assessing Drug Synergism, *Genes & Cancer*, **2**, 1003-1008.

Thorpe, L.M., Yuzugullu, H. and Zhao, J.J. (2015) PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting, *Nature Reviews Cancer*, **15**, 7-24.

Troisi, A., Wong, V. and Ratner, M.A. (2005) An agent-based approach for modeling molecular self-organization, *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 255-260.

Van Regenmortel, M.H.V. (2004) Reductionism and complexity in molecular biology, *Embo Reports*, **5**, 1016-1020.

Viceconti, M., Henney, A., Morley-Fletcher, E. (2016) In silico clinical trials: how computer simulation will transform the biomedical industry, *International Journal of Clinical Trials*, **3**.

Villaverde, A.F. and Banga, J.R. (2014) Reverse engineering and identification in systems biology: strategies, perspectives and challenges, *Journal of the Royal Society Interface*, **11**.

von der Heyde, S.*, et al.* (2014) Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines, *Bmc Systems Biology*, **8**, 21.

Vu, T. and Claret, F.X. (2012) Trastuzumab: Updated Mechanisms of Action and Resistance in Breast Cancer, *Frontiers in Oncology*, **2**, 62.

Walker, D.C., Georgopoulos, N.T. and Southgate, J. (2008) From pathway to population - a multiscale model of juxtacrine EGFR-MAPK signalling, *Bmc Systems Biology*, **2**, 17.

Walker, D.C., *et al.* (2004) The epitheliome: agent-based modelling of the social behaviour of cells, *Biosystems*, **76**, 89-100.

Wang, J., *et al.* (2013) Multi-scale agent-based modeling on melanoma and its related angiogenesis analysis, *Theoretical Biology and Medical Modelling*, **10**.

Wang, Z., *et al.* (2015) Simulating cancer growth with multiscale agent-based modeling, *Seminars in Cancer Biology*, **30**, 70-78.

Way, J.C., *et al.* (2014) Integrating Biological Redesign: Where Synthetic Biology Came From and Where It Needs to Go, *Cell*, **157**, 151-161.

Wieduwilt, M.J. and Moasser, M.M. (2008) The epidermal growth factor receptor family: Biology driving targeted therapeutics, *Cellular and Molecular Life Sciences*, **65**, 1566-1584.

Wiley, H.S., Shvartsman, S.Y. and Lauffenburger, D.A. (2003) Computational modeling of the EGF-receptor system: a paradigm for systems biology, *Trends in Cell Biology*, **13**, 43-50.

Wood, E.R., *et al.* (2004) A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): Relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells, *Cancer Research*, **64**, 6652-6659.

Yan, L., Rosen, N. and Arteaga, C. (2011) Targeted cancer therapies, *Chinese Journal of Cancer*, **30**, 1-4.

Yarden, Y. and Shilo, B.Z. (2007) SnapShot: EGFR signaling pathway, *Cell*, **131**, 1018.

Yarden, Y. and Sliwkowski, M.X. (2001) Untangling the ErbB signalling network, *Nature Reviews Molecular Cell Biology*, **2**, 127-137.

Zhang, Y., *et al.* (2009) HER/ErbB receptor interactions and signaling patterns in human mammary epithelial cells, *Bmc Cell Biology*, **10**.

Zhang, Y., *et al.* (2008) System theoretical investigation of human epidermal growth factor receptor-mediated signalling, *Iet Systems Biology*, **2**, 273-284.

Zhong, Z., Wen, Z.L. and Darnell, J.E. (1994) STAT3 - A STAT FAMILY MEMBER ACTIVATED BY TYROSINE PHOSPHORYLATION IN RESPONSE TO EPIDERMAL GROWTH-FACTOR AND INTERLEUKIN-6, *Science*, **264**, 95-98.

APPENDIX I

TABLE OF ABM TERMINOLOGY

Entity	An independent element of the model, such as a cell or protein.
Agent	Autonomous set of entities comprising the system, their attributes and goals
Compartment/Environment	Computational representation of the physical space within which agents are contained
Neighbour	An agent that exists within the neighbourhood of another agent.
Model	A non-executable description of a system, which may be described in an abstract manner, or for a platform-specific implementation as a simulation.
Simulation	An executable implementation of a model specification
Step	An iteration in time
Multiscale	A model combining processes occurring at different orders of magnitude of time and length.

APPENDIX II

Input parameters for the simulation

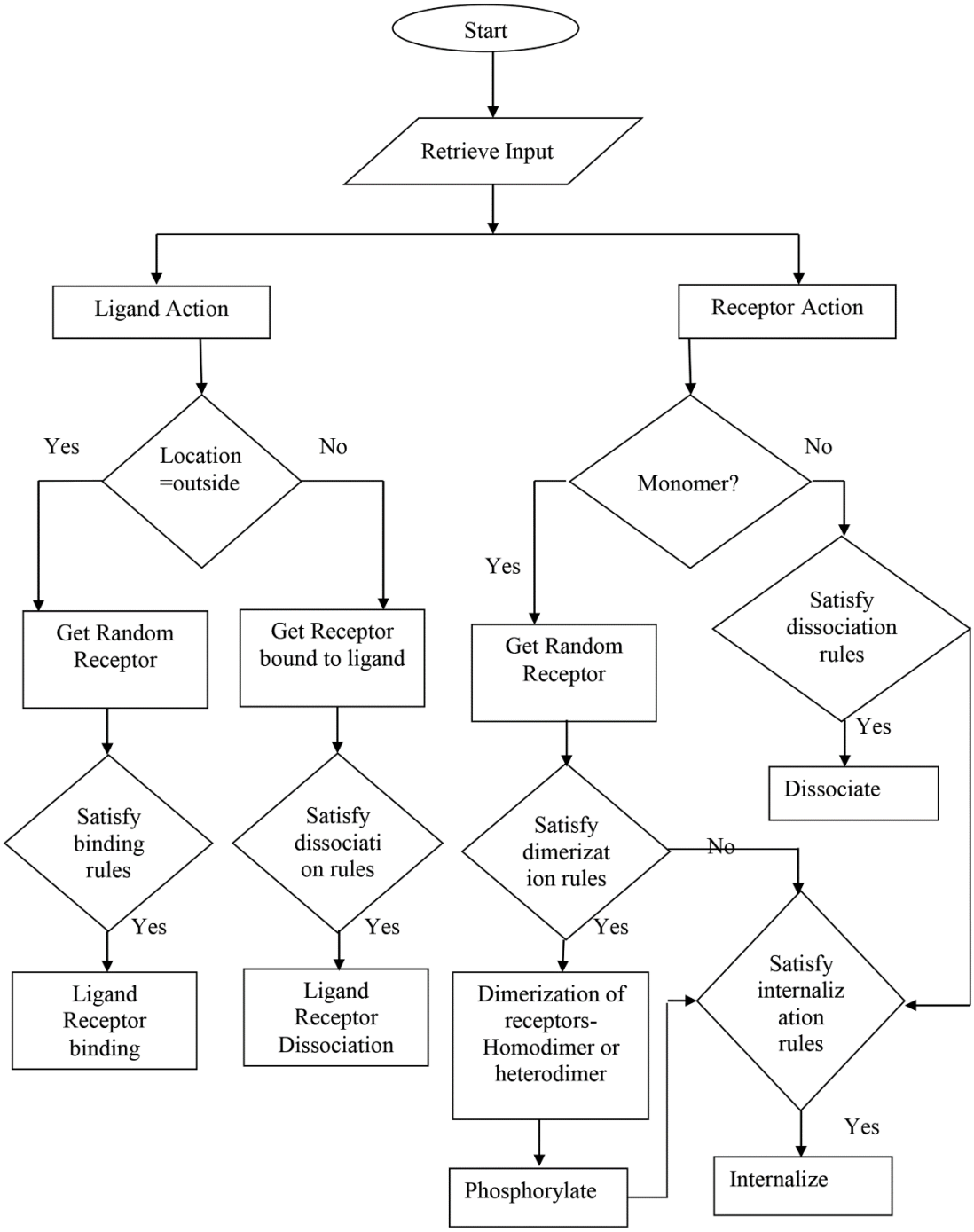
egf_num	Number of EGF	50000
hrg_num	Number of HRG	30000
Erbb1_num	Number of ErbB1	80000
Erbb2_num	Number of ErbB2	300000
Erbb3_num	Number of ErbB3	40000
erk_num	Number of Erk	25000
akt_num	Number of Akt	20000
Trast_num	Number of Trastuzumab	50000
Pert_num	Number of Pertuzumab	50000
MaxSize	Total number of agents	645000
egf_binding_rate	Rate at which EGF binds to the receptor	0.63
hrg_binding_rate	Rate at which HRG binds to the receptor	0.73
egf_disso_rate	Rate at which EGF dissociates the receptor	0.12
hrg_disso_rate	Rate at which HRG dissociates the receptor	0.34
trast_binding_rate	Rate at which Trastuzumab binds to the receptor	0.1
pert_binding_rate	Rate at which Pertuzumab binds to the receptor	0.1
trast_disso_rate	Rate at which Trastuzumab dissociates the receptor	0.12

pert_disso_rate	Rate at which Pertuzumab dissociates the receptor	0.34
dimerization_rate	Rate at which monomers form dimers	0.2
Threshold	Threshold for ADCC	5000
timestep	A period in which the agents can interact with each other and the drugs to carry out a unit action.	1

***Units for rates are in /min**

APPENDIX III

Flow chart for ligand and receptor action



APPENDIX IV

Pseudocode

Main Algorithm

Initialise the agents in the compartments PM and EE
Initialise rule vector
repeat for every time step=1minute until 100 minute
    repeat for every agent
        if agent has undergone a state change in this time step
        then go to the next agent
        else
            if agent is a ligand **then** Call function Ligand_Action
            if agent is a receptor **then**
                if agent is a monomer on PM **then** Call function Monomer_PM
                if agent is a monomer on EE **then** Call function Monomer_EE
                if agent is a dimer on PM **then** Call function Dimer_PM
                if agent is a dimer on EE **then** Call function Dimer_EE
            if agent is a Erk **then** Call function Erk_Action
            if agent is a Akt **then** Call function Akt_Action
            if agent is a Drug **then** Call function Drug_Action
            end if
    end for
    write output file
    calculate average rate values
    determine new rule vector
    initialise all local rates to zero
end for

Ligand_Action, Receptor_Action (which includes Monomer_PM, Monomer_EE, Dimer_PM, Dimer_EE), Erk_Action , Akt_Action, Drug_Action are functions of the decision engine.

Ligand_Action()

if ligand agent is located Outside

if random value is lesser than EGF/HRG binding rate

then bind to unbound EFGR/HER3 monomer agent

end if

if ligand agent is located in PM and bound to EGFR/HER3 monomer agent

if random value is lesser than EGF/HRG dissociation rate

then dissociate the ligand from receptor

end if

Receptor_Action()

Monomer_PM() // receptor agent is a monomer located in PM

Identify the rule subset pertaining to monomer in PM

if rate rules favour Internalisation **or** Dimerization

Change state/location of agent

else if rate rules favour Internalisation **and** Dimerization

Randomly select one event

Change state/location of agent

else

No change in state/location of agent

Monomer_EE() // receptor agent is a monomer located in EE

Identify the rule subset pertaining to monomer in EE

if rate rules favour Degradation **or** Dimerization

Change state/location of agent

else if rate rules favour Degradation **and** Dimerization

Randomly select one event

Change state/location of agent

else

No change in state/location

Dimer_PM()// receptor agent is a dimer located in PM

Identify the rule subset pertaining to dimer in PM

if rate rules favour Internalisation **or** Dissociation

Change state/location of agent

else if rate rules favour Internalisation **and** Dissociation

Randomly select one event

Change state/location of agent

else

No change in state/location

Dimer_EE()// receptor agent is a dimer located in EE

Identify the rule subset pertaining to dimer in EE

if rate rules favour Degradation **or** Dissociation

Change state/location of agent

else if rate rules favour Degradation **and** Dissociation

Randomly select one event

Change state/location of agent

else

No change in state/location

Erk_Action()

if Erk agent is located Cytoplasm and not phosphorylated

Randomly choose a receptor

if receptor is HER1-HER1/HER1-HER2 and random value is lesser than Erk binding rate

then phosphorylate the agent

end if

if Erk agent is phosphorylated

then dephosphorylate the agent

end if

Akt_Action()

if Akt agent is located Cytoplasm and not phosphorylated

Randomly choose a receptor

if receptor is HER2-HER3/HER1-HER3 and random value is lesser than Akt binding rate

then phosphorylate the agent

end if

if Akt agent is phosphorylated

then dephosphorylate the agent

end if

Drug_Action()

if drug agent is located Outside

if random value is lesser than Pertuzumab/Trastuzumab binding rate

then bind to unbound and non-mutated HER2 monomer agent

set domain II/IV to bound

end if

LIST OF PUBLICATIONS

Journals

1. Arya A. Das, T. Ajayakumar Darsana, Elizabeth Jacob; Agent-based re-engineering of ErbB signaling: a modeling pipeline for integrative systems biology, *Bioinformatics*, Volume 33, Issue 5, 1 March 2017, Pages 726–732, https://doi.org/10.1093/bioinformatics/btw709 **(SCI Impact factor: 7.307)**
2. Arya A. Das, Elizabeth Jacob; *In silico* prediction of ErbB signal activation from receptor expression profiles through a data analytics pipeline, *Journal of Biosciences*, Volume 43, Issue 2, 2018, doi:https://doi.org/10.1007/s12038-018-9747-4 **(SCI Impact factor: 1.422)**
3. *In silico* trials to study the effect of Combination drugs on HER2 targeted therapies for cancer. (Communicated)

Conference Papers/Presentations/Posters

1. **Paper Presentation:**

   Arya A. Das, Elizabeth Jacob. A data-derived modeling architecture for empowering virtual experiments: Simulating the ErbB signaling pathway. Paper presented in Dr. K V Rao Research Awards (Biology) –2017

   Arya A. Das and Elizabeth Jacob. Particle -based simulation of ERBB receptor activation and trafficking mechanisms-Virtual experiments in the cell. Paper presented in ICBB-2015-**Won best paper award.**

   Arya A. Das and Elizabeth Jacob. Agent-based Simulation of Receptor Trafficking in ErbB Signalling Network. Paper presented in Biospectrum-2013

2. **Poster Presentation:**

   Arya A. Das, Neenu E. S. and Elizabeth Jacob. Modeling Crosstalk between ErbB Signaling Pathways using a Data Analytics Pipeline. Poster presented in 29th Kerala Science Congress -2017

Arya A. Das and Elizabeth Jacob. A computational pipeline for Systems Biology data analytics: applied to ErbB receptor activation modelling. Poster presented in NextGen Genomics, Biology, Bioinformatics and Technologies (NGBT) Conference - 2016

Arya A. Das and Elizabeth Jacob. *In-Silico* Experiments on EGFR-HER2 Activation and Trafficking Using Particle-Based Simulations. Poster presented in 27th Kerala Science Congress, Allapuzha-2015

Arya A. Das and Elizabeth Jacob. Knowledge-based Computational Modelling of ErbB Receptor Signalling. Poster presented in International Conference on New Horizons in Biotechnology (NHBT-2015)

SUBJECT INDEX